UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE GEOCIÊNCIAS

VICTO JOSÉ DA SILVA NETO

THE CONSOLIDATION OF THE DIGITAL TECHNOLOGICAL SYSTEM: TRANSFORMATIONS IN SCIENCE AND OTHER DOMAINS

A CONSOLIDAÇÃO DO SISTEMA TECNOLÓGICO DIGITAL: TRANSFORMAÇÕES NA CIÊNCIA E EM OUTROS DOMÍNIOS

CAMPINAS
2022

VICTO JOSÉ DA SILVA NETO

THE CONSOLIDATION OF THE DIGITAL TECHNOLOGICAL SYSTEM: TRANSFORMATIONS IN SCIENCE AND OTHER DOMAINS

A CONSOLIDAÇÃO DO SISTEMA TECNOLÓGICO DIGITAL: TRANSFORMAÇÕES NA CIÊNCIA E EM OUTROS DOMÍNIOS

THESIS PRESENTED TO THE INSTITUTE OF GEOSCIENCES OF THE UNIVERSITY OF CAMPINAS TO OBTAIN THE DEGREE OF DOCTOR OF SCIENTIFIC

DISSERTAÇÃO/TESE APRESENTADA AO INSTITUTO DE GEOCIÊNCIAS DA UNIVERSIDADE ESTADUAL DE CAMPINAS PARA OBTENÇÃO DO TÍTULO DE MESTRE(A)/DOUTOR(A) EM POLÍTICA CIENTÍFICA E TECNOLÓGICA

ORIENTADORA: PROFª. DRª. MARIA BEATRIZ MACHADO BONACELLI

COORIENTADOR: PROF. DR. CARLOS AMÉRICO PACHECO

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO VICTO JOSÉ DA SILVA NETO E ORIENTADA PELA PROFª. DRª. MARIA BEATRIZ MACHADO BONACELLI E COORIENTADA PELO PROF. DR. CARLOS AMÉRICO PACHECO

CAMPINAS

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Geociências
Marta dos Santos - CRB 8/5892

Si38c

Silva Neto, Victo José da, 1987-
The consolidation of the digital technological system : transformations in science and other domains / Victo José da Silva Neto. – Campinas, SP : [s.n.], 2022.

Orientador: Maria Beatriz Machado Bonacelli.
Coorientador: Carlos Américo Pacheco.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Geociências.

1. Plataforma digital. 2. Inteligência artificial. 3. Transformação digital. 4. Inovações tecnológicas. 5. Aprendizado de máquina. I. Bonacelli, Maria Beatriz Machado, 1962-. II. Pacheco, Carlos Américo, 1957-. III. Universidade Estadual de Campinas. Instituto de Geociências. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: A consolidação do sistema tecnológico digital : transformações na ciência e em outros domínios
Palavras-chave em inglês:
Digital platform
Artificial intelligence
Digital transformation
Technological innovations
Machine learning
Área de concentração: Política Científica e Tecnológica
Titulação: Doutor em Política Científica e Tecnológica
Banca examinadora:
Maria Beatriz Machado Bonacelli [Orientador]
André Tosi Furtado
Fernanda de Negri
Marcelo Knobel
Anita Kon
Data de defesa: 22-07-2022
Programa de Pós-Graduação: Política Científica e Tecnológica

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-9009-1203
- Currículo Lattes do autor: http://lattes.cnpq.br/8293874077969904

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE GEOCIÊNCIAS

UNICAMP

AUTOR: Victo José da Silva Neto

THE CONSOLIDATION OF THE DIGITAL TECHNOLOGICAL SYSTEM: TRANSFORMATIONS IN SCIENCE AND OTHER DOMAINS

A CONSOLIDAÇÃO DO SISTEMA TECNOLÓGICO DIGITAL: TRANSFORMAÇÕES NA CIÊNCIA E EM OUTROS DOMÍNIOS

ORIENTADORA: Profª. Drª. Maria Beatriz Machado Bonacelli

COORIENTADOR: Prof. Dr. Carlos Américo Pacheco

Aprovado em: 22 / 07 / 2022

EXAMINADORES:

Profª. Drª. Maria Beatriz Machado Bonacelli - Presidenta

Prof. Dr. André Tosi Furtado

Prof. Dr. Marcelo Knobel

Profª. Drª. Anita Kon

Profª. Drª. Fernanda de Negri

A Ata de Defesa assinada pelos membros da Comissão Examinadora consta no processo de vida acadêmica do aluno.

Campinas, 22 de julho de 2022.

AGRADECIMENTOS

Esta tese foi construída entre 2018 e 2022. Quantas vidas cabem em quase quatro anos e meio, cinquenta e dois meses, ou mil e seiscentos dias? Quão diferentes éramos nós (eu e o mundo) em 2017, quando decidi começar essa jornada? Eu era, como Joaquim Nabuco conta em “Minha Formação”, “ávido de impressões novas, fazendo os meus primeiros conhecimentos com os grandes autores”. E como ele, “As minhas ideias eram, entretanto, uma mistura e uma confusão”. O doutorado, além de ser o período para desenvolver uma tese, serve para amadurecer o espírito do pesquisador. Não se trata de arrefecer sua curiosidade, que é o combustível primeiro e insubstituível da pesquisa, mas de permitir um contato íntimo com a ciência. Este contato permite ao pesquisador vislumbrar os caminhos da pesquisa, o valor da ciência, bem como seu próprio potencial.

Já o mundo em 2017, bom, este já estava confuso e estranho, mas confundiu-se e estranhou-se ainda mais desde então. Tendo escrito metade desta tese durante a pandemia, meu primeiro agradecimento vai para aqueles que enfrentaram com coragem esta situação dramática para o mundo e para o Brasil em particular. Cientistas, médicos, comunicadores, enfermeiras, professores, cidadãos que de alguma forma ajudaram quem precisava, fica aqui meu agradecimento a vocês. A pandemia e sua congênere, a infodemia, atestaram que meu tempo e minha energia estavam sendo investidos em algo realmente importante: a Ciência, tanto como profissão quanto como objeto de estudo.

Falar sobre ciência no Brasil de 2022 poderia parecer a alguns apologia ao supérfluo. Afinal, a combinação de uma pandemia global com um governo incapaz e despreparado geraram um resultado de terra arrasada. No entanto, repousa sobre a ciência, a arte e a cultura a possibilidade de recolocar o Brasil no seu devido lugar, recuperando-o material e moralmente. Tendo em vista este contexto desafiador, não surpreende que minha jornada tenha sido possível graças ao apoio de inúmeras pessoas. Estes muitos que conviveram comigo pessoal, digital e intelectualmente tiveram um papel fundamental na construção da minha visão do que é fazer pesquisa e no entendimento de que é preciso estar conectado ao mundo e às pessoas para que qualquer pesquisa tenha sentido.

Sem a menor dúvida, minha pesquisa não teria existido sem o apoio dos meus pais: Victo e Lidia. Este apoio incondicional que recebo desde sempre, no sentido de incentivar, celebrar junto nos momentos bons e apoiar nos momentos difíceis. Mais do que a universidade, mais do que o sistema nacional de pós-graduação, minha família é a grande responsável pela formação deste pesquisador que vos escreve. Obrigado por tudo!

Os meus amigos e colegas de DPCT foram também peça chave para minha formação e de minhas ideias. A mudança do prédio antigo para o prédio novo do Instituto de Geociências foi boa em muitos sentidos, mas em um foi prejudicial: a dinâmica de convívio diário entre o corpo discente do PPG se desfez quando os espaços comunitários foram reduzidos e, de alguma forma, esterelizados. Tive a sorte de carregar para este novo contexto as amizades feitas nos anos anteriores, durante o mestrado: Luis, Diego, Altair, Jeny, Edgar, Taty, Marina, Martha, Liz, Anna, Felipe, Jean, Dam, Paola, Rafa, Vini, Rodri e Luisa obrigado pelo seu tempo e pela sua companhia! Sempre que precisarem, contem comigo. Também conheci novos integrantes, que começam sua própria trajetória na pesquisa: Aldemir, Manu e Matheus, sigam em frente!

Os professores e funcionários do DPCT também foram muito importantes na minha trajetória. Sem dúvida alguma, minha orientadora, Bia, foi responsável por grande parte desta tese e pela minha formação. Construímos uma relação de confiança que foi muito importante para que eu pudesse desenvolver minha pesquisa, buscar autonomia e me inserir no mundo da pesquisa com confiança. Obrigado Bia! Devo agradecer também ao meu co-orientador, Pacheco, que sempre compartilhou seu conhecimento conosco. Ainda devo agrader aqueles com os quais aprendi muito: Leda, Janaína, Marko, Flávia, Fernando Mesquita e todos os professores do DPCT (acho que cursei quase todas as disciplinas). Ao professor Marko, obrigado mais uma vez, por todo o seu suporte no processo do meu período no exterior. Devo um agradecimento muito especial ao professor Bruno Fischer (FCA/UNICAMP), que me colocou em contato com os professores da Universidade de Utrecht. Obrigado Bruno, não apenas por isso, mas também pela sua postura com relação à pesquisa e seu convívio com os discentes, que são um exemplo para mim. Devo agradecer também aos membros da minha banca, prof. André Furtado, profª Anita Kon, prof. Marcelo Knobel e Fernanda de Negri.

Aos professores da Universidade de Utrecht, na Holanda, agradeço pela acolhida e por tudo que me ensinaram no período do meu doutorado sanduíche. Em especial meu tutor, professor Koen Frenken, uma pessoa brilhante e ao mesmo tempo simples e acessível; também agradeço a Iryna Susha, Wouter Boon, Jarno Hoekman e Thomas Bouwens pelas trocas e pelo trabalho em conjunto. Meus agradecimentos são endereçados também a todos os funcionários do Instituto de Geociências e *do Copernicus Institute of Sustainable Development* (*Utrecht University*), cujo trabalho viabiliza o trabalho de pesquisa em vários níveis. Ainda em Utrecht, a companhia de alguns foi muito importante. Bianca, David, Rodrigo e Ellen, vocês tornaram minha experiência muito mais rica e interessante.

A participação em projetos e eventos foi muito importante na minha formação e devo agradecer a algumas pessoas que conheci ou que me abriram as portas para fazer parte de projetos específicos. Tive a oportunidade de participar da Escola de História da Ciência na USP (FFLCH), onde conheci o grande professor Mauro Condé (UFMG). O InnSciD (IeA/USP), programa de diplomacia da inovação e da ciência, me permitiu conhecer o incansável Pedro Ivo (Itamaraty) que hoje lidera o movimento de diplomacia da inovação enquanto baseado na Alemanha; a Cátedra Oscar Sala (IeA/USP) criou um espaço privilegiado para trocas com muitos participantes brilhantes, especialmente a profª. Lucia Santaella e o prof. Eugênio Bucci; tive a sorte de conhecer e trabalhar com a Edilaine Camilo (UNICAMP) no projeto *Energy Big Push* (CEPAL/CGEE); outro projeto, o de *Avaliação da Estratégia de Governo Digital* (ENAP) me colocou no exercício prático de avaliador de uma política pública (e digital): obrigado Ana Maria Carneiro (NEPP), que me ensinou muito sobre avaliação, mas também sobre como gerir um projeto profissionalmente; obrigado também à Ana Carolina Spatti e Adriana Bin pela convivialidade e pelos ensinamentos. Por fim, meus agradecimentos à CAPES, pelo apoio financeiro que viabilizou esta tese e minha formação no Brasil e no exterior (PrInt).

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Tenho que destacar duas parcerias que surgiram no momento menos provável, quando o mundo se recolhia ao isolamento. Edu Spanó, que se desdobra em múltiplos e conhece tudo sobre inovação no setor público. No início da pandemia trabalhamos juntos no que foi um grande incentivo para não parar nem desanimar. Não tenho dúvidas de que seus projetos serão um sucesso, Edu! E Tulio Chiarini, esse carioca do sul de minas que me encontrou em um dos *Encontros Nacionais de Economia Industrial e Inovação* (ENEI) e dali surgiu não só uma dupla de trabalho, mas também uma amizade. Finalmente pudemos, em 2022, sair do convívio digital e dividir alguns dias na cidade maravilhosa. Pela sua parceria, paciêcia e por tudo que aprendi contigo, obrigado!

Concluo estes agradecimentos feliz, por poder olhar para trás e ver tantos momentos bons; momentos desafiadores também, mas eu nunca estive sozinho.

EPÍGRAFE

“mas eu, o que é que eu era? Eu ainda não era ainda. Se ia, se ia [...] Viajar! – mas de outras maneiras: transportar o sim desses horizontes!...”

J.G.R.

RESUMO

Esta tese é sobre a transformação digital da ciência. Muitos trabalhos investigam a transformação digital no setor privado, mas poucos olham para o mesmo fenômeno nas organizações e práticas científicas. Esta transformação envolve novas tecnologias, mas implica também na reestruturação organizacional do fazer científico e na reinvenção das organizações científicas. O objetivo da tese é oferecer ao formulador de políticas de ciência e tecnologia uma visão geral da interação entre a prática científica e as tecnologias digitais, por meio da contextualização histórica dessa interação e da proposição de novos conceitos relacionados à trajetória das tecnologias digitais na ciência. Este objetivo leva à questão central da tese: como a ciência interagiu com as tecnologias digitais e como ela se transformou nesse processo? A primeira parte, histórica e conceitual, identifica as principais tecnologias e trajetórias de inovação digital da última década e as conecta à revolução tecnológica em curso desde a década de 1970. A segunda parte trata das plataformas digitais e propõe uma tipologia de plataformas científicas: ressalta-se a co-existência de plataformas científicas desenvolvidas e operadas por diversos grupos sociais (*e.g.*, setor privado, comunidade científica) e observa-se como a ciência gerou inovações que se tornariam paradigmáticas décadas depois. O recorte da terceira parte recai sobre a inteligência artificial (IA) e sua influência sobre os caminhos da ciência. Ao propor um arcabouço analítico que permite sistematizar a influência de uma dada tecnologia sobre a ciência, documenta-se diversas formas pelas quais a IA, em especial o aprendizado de máquina, influenciou a ciência: a IA como fato a ser explicado epistemologicamente; como ferramenta em múltiplas áreas da pesquisa científica; como fonte de dados experimentais e como motivação econômica para conduzir pesquisa científica. Em síntese, os resultados indicam que a ciência sempre interagiu com as tecnologias digitais mediante um produtivo processo de co-evolução, estando sempre muito próxima dos desenvolvimentos tecnológicos de fronteira. As tecnologias digitais tiveram alguns de seus princípios gestados e testados nos espaços da ciência. A interação entre a comunidade científica e as tecnologias digitais ocorre muitas vezes via parcerias multissetoriais com o Estado ou o setor privado. Observa-se que atualmente a transformação da ciência não acompanha o ritmo da transformação de outras áreas: sistemas científicos legados demonstram dificuldades para absorver o potencial das novas tecnologias. Observa-se também uma resistência institucional, que detém transformações mais radicais. Tal fato não é necessariamente ruim: conforme detalhado, uma difusão descontrolada das tecnologias digitais no seio das práticas científicas (e.g., algoritmização) levanta inúmeras questões, como a opacidade algorítmica, a privatização das infraestruturas científicas e os limites da ética.

Palavras-chave: Plataforma digital; Inteligência artificial; Transformação digital; Inovações tecnológicas; Aprendizado de máquina.

ABSTRACT

This thesis is about the digital transformation of science. Many works investigate digital transformation in the private sector, but few look at the same phenomenon in scientific organizations and practices. This transformation involves new technologies, but it also implies the organizational restructuring of scientific work and the reinvention of scientific organizations. The objective of the thesis is to offer science and technology policymakers an overview of the interaction between scientific practice and digital technologies, through the historical contextualization of this interaction and the proposition of new concepts related to the trajectory of digital technologies in science. This objective leads to the central question of the thesis: how has science interacted with digital technologies and how has it been transformed in this process? The first part, historical and conceptual, identifies the main technologies and trajectories of digital innovation of the last decade and connects them to the technological revolution underway since the 1970s. The second part deals with digital platforms and proposes a typology of scientific platforms: it highlights the co-existence of scientific platforms developed and operated by different social groups (*e.g.*, the private sector, the scientific community) and it reveals how science generated innovations that would become paradigmatic decades later. The third part focuses on artificial intelligence (AI) and its influence on the paths of science. By proposing an analytical framework that makes it possible to systematize the influence of a given technology on science, several ways in which AI, especially machine learning, influenced science are documented: AI as a fact to be epistemologically explained; as a tool in multiple areas of scientific research; as a source of experimental data and as an economic motivation to conduct scientific research. In summary, the results indicate that science has always interacted with digital technologies through a productive process of co-evolution, always being very close to the frontier technological developments. Digital technologies had some of their principles gestated and tested in the spaces of science. The interaction between the scientific community and digital technologies often takes place via multisectoral partnerships with the State or the private sector. It is observed that currently the transformation of science does not keep pace with the transformation of other areas: legacy scientific systems show difficulties in absorbing the potential of new technologies. There is also an institutional resistance, which holds more radical transformations. This is not necessarily bad: as detailed, an uncontrolled diffusion of digital technologies within scientific practices (*e.g.*, algorithmization) raises numerous questions, such as algorithmic opacity, the privatization of scientific infrastructures and the limits of ethics.

Keywords: Digital platform; Artificial intelligence; Digital transformation; Technological innovations; Machine learning.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
AMT	Amazon Mechanical Turk
ANII	*Agencia Nacional de Investigación e Innovación*
ANPD	National Data Protection Authority
API	Application Program Interfaces
ASN	Academic Social Networks
BOAI	Budapest Open Access Initiative
CBPF	*Centro Brasileiro de Pesquisas Físicas*
CERN	Conseil Européen pour la Recherche Nucléaire
CGEE	*Centro de Gestão e Estudos Estratégicos*
CNPq	*Conselho Nacional de Desenvolvimento Científico e Tecnológico*
COMLATTES	*Comissão de Gestão do Lattes*
CONCYTEC	*Consejo Nacional de Ciencia, Tecnología e Innovación*
CV-Lattes	Lattes Curricula
DGP	*Diretório de Grupos de Pesquisa*
DIIP	*Diretório de Instituições e Infraestruturas de Pesquisa*
DL	Deep Learning
DTS	Digital Technological System
ECLAC	Economic Commission for Latin America and the Caribbean
EU	European Union
FINEP	*Financiadora de Estudos e Projetos*
FPGA	Field Programmable Gate Arrays
GPT	General-Purpose Technology
GPU	Graphics Processing Unit
IaaS	Infrastructure as A Service
ICT	Information and Communication Technologies
ICTs	Communication Technologies
IMI	*Invention in the Method tf Invention*
INPI	*Instituto Nacional de Propriedade Industrial*
IoT	Internet of Things
LGPD	*Lei Geral de Proteção de Dados*
ML	Machine Learning

NanoHUB	Nanotechnology Science Gateway
Nectar	National eResearch Collaboration Tools and Resources
NSF	National Science Foundation
O2O	Offline to Online
OAC	Office of Advanced Cyberinfrastructure
OAI	Open Archives Initiative
ONCTI	*Observatorio Nacional de Ciencia, Tecnología e Innovación*
OpenDOAR	Directory of Open Access Repositories
ORCID	Open Researcher and Contributor Identifier
OST	Office of Science and Technology
PaaS	Platform as a Service
PAP	Principal-Agent Problem
PMH	Protocol for Metadata Harvesting
PoA	Prolific Academic
PSC	Pittsburgh Supercomputer Center
S&T	Science and Technology
SaaS	Software as a Service
SCOT	Social Construction of Science and Technology
SENESCYT	Ecuador of *Secretaría de Educación Superior, Ciencia, Tecnología e Innovación*
SGCI	Science Gateway Community Institute
SICYTAR	*Sistema de Información de Ciencia y Tecnología Argentina*
SO1	Organizational Innovation
SO2	Scientific Organizations
SO3	Core Digital Technology
ST&I	Science, Technology and Innovation
TEP	Techno-Economic Paradigm
ToS	Terms of Service
TPU	tensor processing unit
UFPE	Federal University of Pernambuco
UFPel	*Universidade Federal de Pelotas*
UFSC	Federal University of Santa Catarina
UGC	User-Generated Content
UNCTAD	United Nations Conference on Trade and Development

UNISIST	International Scientific Information System
VC	Value Chain
VRE	Virtual Research Environments
WIPO	World Intellectual Property Organization
WWW	World Wide Web

SUMMARY

INTRODUCTION

The second and third decades of the 21st century have witnessed the first acts of a digital technological system that seems to be advancing in all directions. After decades of promises, artificial intelligence has been incorporated into various economic, scientific and cultural activities. New and traditional sectors are under intense transformations due to these new technologies and the organizational forms associated with them, such as digital platforms. In this context, even areas such as the State (KON, 2019) have undergone extensive changes.

There are many names for a new era such as platform capitalism (SRNICEK, 2017), digital economy (STURGEON, 2019; TEECE, 2018), platform society (VAN DIJCK; POELL; WAAL, 2018) and surveillance capitalism (ZUBOFF, 2019). A good deal of public, media and academic attention has been focused on the positives versus the negatives of this progressive digitization. However, this perspective is not the most fruitful: there are, as in every technological system, positive and negative uses of technologies. In addition, most of the attention is on the digital transformation of firms (FRANK; DALENOGARE; AYALA, 2019; NAMBISAN; WRIGHT; FELDMAN, 2019; ZHU *et al*., 2006). Still, there is great potential in the digital transformation of government organizations, cooperatives or associations. There is a gap, therefore, on the transformation of organizations beyond the market, seeking the best use of digital technologies.

The literature points to a temporal lag (up to two decades) in the adoption of new technologies by public organizations, due to the lack of market competition as a driver of transformations and innovations (PEREZ, 2009). Thus, organizational inertia would be greater in the public sector and in organizations isolated from market competition (such as associations, cooperatives, civil society organizations and scientific organizations). Would we observe this "lag" in scientific organizations? Science and science practitioners have also approached the latest digital technologies. We know that there is a growing use of artificial intelligence in laboratories and research (BIANCHINI; MORITZ; PELLETIER, 2022). We also know that, beyond the well know digital repositories of texts and data, there is some episodic innovative initiatives of digital platforms dedicated to scientific practice. An international example is Prolific[1], a platform that connects researchers with individuals willing to participate in surveys. A Brazilian example is SociaLab[2] (Figure 1), which connects

[1] https://www.prolific.co/
[2] https://www.socialab.com.br/

researchers and laboratories to share idle supplies and equipment. It is necessary, however, to go beyond the knowledge of some cases and unravel the pattern of interaction between science[3] and digital technology.

Figure 1 – Available research itens at SociaLab digital platform for sharing

Source: socialab.com.br; own translation: Available at this moment for you: 16292 reagents: thousands of solutions for your laboratory; 531 cells: viruses, bacteria, fungi, protozoa and cell lines; 439 laboratories: from different units in Brazil sharing resources; 502 thousand reais in donations: donations made by science partners.

The objective of the thesis is to offer the science and technology policy-maker an overview of the interaction between scientific practice and digital technologies, through the historical contextualization of this interaction and the proposition of new concepts related to the trajectory of digital technologies in science. This objective leads us to the central question of the thesis: **How has science interacted with the digital technologies and how did it transfomed itself in this process?**

To face the general objective of the thesis, we have broken it down into three specific objectives (SO1, SO2, SO3). The first is to establish a reference on relevant digital technologies with high transformative potential for organizations, through a long-term view of technological and organizational innovation (SO1) standards. The second objective seeks to elucidate the organizational transformations in science permeated by new digital technologies, by proposing a historically-informed taxonomy of these digitized scientific organizations (SO2). The third and final objective is to understand the nature of the relationship between science and digital technology in terms of its directionality, location and impact, through a

[3] Besides, science, as a sphere of social practices and values, is too broad a universe to be fully observed. Not only do we have to make choices about the scope of the object we are going to observe, but we also have to consider that there are interpretations of change in science that are more supported by history and philosophy than by the economics of innovation (SHAPIN, 1992). Although this thesis does not contribute to the literature of philosophy or history of science, in a way it dialogues with the perception that technique plays a central role in the evolution of science.

case study of a core digital technology (SO3). The specific objectives were articulated with the rest of the thesis structure in Table 1.

In light of the general objective and the general research question and the specific objectives presented in the paragraph above, we define some subsidiary research questions that will be addressed throughout the dissertation:

- What do we mean by digital technologies? Is there a specific set of technologies that are most representative? How do they articulate with each other?
- How was the historical process of interaction between science and these digital technologies, from the point of view of use (application) and generation (innovation)? Which were the main actors involved? (Chapter 2, 3)
- Has this interaction influenced the paths of scientific research? And if so, in what ways? (Chapter 4)

Table 1 – Analytic structure of the dissertation

Specific objective	**Chapter**	**Title**	**Subsidiary question(s)**	**Publication status**
SO1	1	The digital technological system: artificial intelligence, cloud computing and Big Data	What do we mean by digital technologies? Is there a specific set of technologies that are most representative? How do they articulate with each other?	Published (dec./2020) *Revista Brasileira de Inovação* (B2)
SO2	2	The platformization of science: towards a scientific digital platform taxonomy	How was the historical process of interaction between science and these digital technologies, from the point of view of use (application) and generation (innovation)? Which were the main actors involved?	Published (sep./2022) *Minerva: a review of science, learning and policy* (B1)
	3	The platformization of science: Lattes platform in a crossroads?		Published (may/2022) *Discussion paper/Institute for Applied Economic Research* (B3)
SO3	4	Framing the effects of machine learning on science	Has this interaction influenced the paths of scientific research? And if so, in what ways?	Published (jun./2022) *AI & Society* (B1)

Source: author's own

Targeting the first of the above subsidiary questions it is necessary to understand the phenomenon: why, at the same time, do different sectors seem to be invaded by attempts to digitize information or to algorithmize practices Are we going through a Fourth Industrial

Revolution or a special phase of a longer process? Understanding these issues requires a historical perspective on the formation of the current digital technology system. Second, it is necessary to observe the historical evolution of science pari passu the evolution of digital technologies. Have digital technologies caught up with science just now? Is science a *locus* for the application or development of these technologies (user or innovator)? Do the latest digital technologies only affect practices on the "factory floor" of science, or do they affect higher levels in terms of the principles and values of the scientific community?

It is easy to accept the discourse that we are in a new digital age and that everything is new. This discourse, however, carries some biases, such as the short-term view, the emphasis on technical change, the neglect of institutional changes and even a certain euphoria for change that does not distinguish between the frontier of technical possibilities and the collective values of society. This thesis seeks to circumvent these biases and view digital technologies as complex elements, historically and technically connected to other systems. In addition, we understand the current technological revolution, which is deepening with the latest digital technologies, as a process inherent to the capitalist system. It is interests, at the same time, technical and economic that "push" these new technologies to the social order. Despite this, there is no technological determinism. There is a process, in which many agents act, that ultimately defines how these technologies will be incorporated (or not) in our sociocultural practices.

What qualifies the current set of digital technologies as revolutionary is their capacity to transform economic and social sectors beyond their sector of origin, information and communication technologies (ICT). In other words, digital technologies have become a source of technical and organizational dynamism for the entire economy. This includes not just the private sector. In fact, the transposition of the principles associated with the system of digital technologies to other areas, such as the State and the academy, indicates that the process of techno-economic transformation is in a new phase.

In this new phase, the difficulties and discussions are no longer just technical. Technology, while it can always be improved, has already proven its ability to deliver efficiency, productivity, value. The difficulties and discussions turn to the institutions (NELSON, 1994, 2002; NELSON; SAMPAT, 2001). After all, technologies do not operate in a vacuum. They are mobilized by firms that follow certain routines; they serve consumers who have demarcated habits; assist states and governments to fulfill their mandates following established norms; affect society and challenge its pre-existing laws and regulations. In other words, the technological revolution, when it spills over beyond the sector that generated it, collides with existing institutions: “Any new technical base [...] cannot be implanted without pain” (FURTADO, 1986, p. 1). The evolutionary (or neo-Schumpeterian) economic literature provides us with a useful theoretical-conceptual matrix to frame our topic.

In the early 1980s, a Venezuelan researcher sought to understand how these technological revolutions relate to the cycles of growth and stagnation of capitalism. Carlota Perez (1939 –) integrated and refined the views of Joseph Schumpeter (1883 – 1950) and Simon Kuznets (1901 – 1985) (DRECHSLER; KATTEL; REINERT, 2009; REINERT; SAGALOVSKY, 2009), proposing an original perspective capable of integrating technology, economy and institutions in a long-term view of the capitalist system. Her view is, first of all, historical: which industries led economic growth at a given time? Based on which technological systems? Why does there seem to be the simultaneous emergence of “constellations of innovations”? When does the growth stagnation of this set of industries occur? Why, after stagnation, is there a time lapse for the emergence of a new set of industries based on new technologies? The author surveys these phenomena and finds patterns that are repeated throughout history: technological revolutions are born in one sector and progressively spread throughout society; financial capital, more flexible and less tied to past investments, sponsors these new technologies that offer high returns; the attractiveness of these new technologies increases as the profitability of established economic activities declines; each technological revolution presents key inputs (such as steel or microchips). She achieved this feat by reaching the conclusion that, in order to understand the long-term evolution of capitalism, “the proper object of study was not economic growth per se, but the process of technological diffusion and its various consequences.” (REINERT; SAGALOVSKY, 2009, p. 404).

Second, C. Perez’s view is institutionalist. The establishment of any industry, of any technological system, depends on its coupling to social institutions, broadly understood as norms, laws, regulations, customs, habits (PEREZ, 1983, 2002). These institutions are shared mental models that deliver predictability to the socioeconomic behavior of firms, consumers, regulators and governments. They include competition strategies vis-à-vis dominant technologies, personal and collective uses of technique embodied in products and services; the limits imposed on the frontier of technical possibilities that is based on the values of the collectivity; regulations that counterbalance tendencies of capital concentration and labor exploitation.

Such institutions provide a channel for predictability, legitimacy and reliability in technical systems. However, when profitability drops and the industry reaches saturation point, promising sectors, generally based on new technologies, start to be promoted: emerging technologies apply for the position of key technological system and the sector applies for the position of carrier branch (PEREZ, 2009) of an economic recovery. On this occasion, existing institutions may become anachronistic in the face of new technical possibilities. In other words, the shared heuristics that were part of the techno-economic system prior to social life become obstacles to the emergence of a new set of technologies. Along the same lines, new technologies may demand [new?] institutions to cover aspects that did not exist in the previous techno-economic and socio-cultural reality. A very clear and concrete example is the National Data Protection Authority (ANPD): there was no need for such an institution, prior to the ability of digital technological systems to economically exploit data at scale.

With renewed institutions that dialogue with new technologies (ordering them, limiting them when necessary, promoting them when convenient), a structural crisis marks the end of a techno-economic paradigm (TEP) and the establishment of a new one. For Freeman and Perez (1988, p. 8), TEP is the "combination of interrelated products and processes, technical, organizational and managerial innovations, incorporating a quantum leap in potential productivity for all or most of the economy". In her book, Perez (2002, p. 16) refers to the TEP as "a best-practice model made up of a set of all-pervasive generic technological and organizational principles, which represent the most effective way of applying a particular technological revolution and of using it for modernizing and rejuvenating the whole of the economy."

A TEP spans decades from its point of origin to its exhaustion. The current digital or informational TEP began in the 1970s with the microelectronics technological system and is still ongoing. It has already incorporated several technological systems, such as the PCs in the 1980s and 1990s and the current system of digital technologies (artificial intelligence, cloud computing and Big Data), which will be the subject of chapter 1. During this long process of establishment, the TEP goes through several phases. In the former, it coexists with the institutions of the TEP that precedes him. At a certain point, the crisis of the past TEP becomes more acute and the new TEP has a window of opportunity to emerge; this demands the institutional transformation we are going through today.

Our temporal and analytical perspective is not as broad as that of the informational TEP. We will not begin our investigation in the 1970s, in the microelectronic roots of the informational TEP (FURTADO, 1986; PEREZ, 1985). Nor will we discuss all phases of this TEP. The analytical focus of this thesis is of the digital technological system, a single system among several systems that make up the informational TEP (sometimes called the Third Industrial Revolution). Freeman and Perez (1988, p. 8) define the technological system as:

> Far-reaching technological changes, affecting many branches of the economy, as well as giving rise to entirely new sectors. They are based on a combination of radical and incremental innovations, along with organizational and managerial innovations that affect more than one or a few companies. [Relates to] the concept of 'constellations' of innovations, which were technically and economically interrelated.

Although we recognize that the digital technological system is part of a "system of systems" (PEREZ, 2009), we limit our analysis temporally to the 2010-2020 decade. The first part of the thesis, "The Digital Technological Revolution: Evolutionary and Institutional Economics" addresses this topic in Chapter 1. Entitled "The Digital Technological System: Artificial Intelligence, Cloud Computing and Big Data". This chapter was originally published as an article[4] in the *Revista Brasileira de Inovação*. The authors find key technologies within the digital technology suite and identify three core technologies that underpin most others: artificial intelligence, cloud computing, and big data. Furthermore, these technologies form a system, a set that is greater than just the sum of its parts, given their pattern of interrelation and complementarity.

The article recognizes that this system is part of a larger paradigm (the informational or digital TEP). At the same time, the system of digital technologies brings developments to numerous sectors, overflowing from its sectoral core of origin (ICT). In other words, with this last system, the TEP reaches a phase of consolidation. This consolidation is verified through three "areas of practice and perception" (PEREZ, 2009, p. 194) that spread horizontally: (i) in the dynamics of the relative cost structure of productive inputs; (ii) in the perception of spaces for innovation; and (iii) organizational principles and criteria. It is argued that, in these three axes, it is possible to observe the influence of the digital technological system: via digitization, altering the cost structure of productive inputs; via algorithmization opening up new space for innovation; and via platformization processes, offering a new organizational model. Finally, it is recognized that there is still institutional work to be done so that the full potential of the digital technological system (and the TEP that contains it) is released: there is a lag in the implementation of technologies due to the absence of idiosyncratic institutions that should regulate the use of new inputs, such as data.

---

[4] SILVA, V. J.; BONACELLI, M. B. M.; PACHECO, C. A. O sistema tecnológico digital: inteligência artificial, computação em nuvem e Big Data. **Revista Brasileira de Inovação**, Campinas, SP, v. 19, p. e0200024, 2020. DOI: 10.20396/rbi.v19i0.8658756.

The private sector and firms continue to be the preferred locus of innovation in the capitalist system. Our interest, however, lies in how other social spheres adopt digital technologies, are affected by them, transform, collapse or reinvent themselves. We are no longer at the moment when a sector emerges as an island of innovative dynamism. We have arrived at a time when "the common-sense principles of organizing for maximum efficiency and effectiveness embodied in the techno-economic paradigm gradually spread from the business world to government and other non-profit institutions" (PEREZ, 2009, p. 198). There is a transposition of the principles selected by the market as the most efficient in the use of new technologies for other areas. At this moment, what C. Perez sees as one of the pillars of a TEP is put to the test: "the isomorphism in the changes that occur in the most diverse institutions, starting with companies" (PEREZ, 2002, p. 16). Is it possible to see any form of isomorphism in science?

What the three lines of evolution identified in Chapter 1 do is to bring the discussion from a longer-term level (of the TEP) to a slightly more operational level in terms of science and technology policy and economic policy. Although these lines of evolution are clearer, it is necessary to understand how their dynamics occur in different contexts. The perezian approach has no tools to offer us when we seek to understand how the principles of a TEP propagate to a specific sphere, be it the state or science. It is necessary to resort to other concepts and theoretical tools. In our case, as we analytically focus on science, we find support in the work of Nathan Rosenberg (1927 – 2015). His view on the interaction between science and technology is mediated by the understanding that after the Second Industrial Revolution, but especially from the 20th century onwards, economic stimuli began to have more ascendancy over the evolution of science. Rosenberg's approach is predominantly historical and is adopted in the following chapters; as a consequence, this dissertation fits into the Economic History of Science and Technology. The next chapters seek to investigate the propagation of the digital TEP into the context of science.

The second part of the thesis is called "The platformization of science". It is the largest part of the thesis, consisting of two chapters. Out of the axis defined previously (digitization, algorithmization and platformization) this part focuses on platformization, *i.e.*, the establishment of digital platforms as an organizational model.

The second chapter, entitled "The platformization of science: Towards a scientific digital platform taxonomy", was originally published[5] by *Minerva: A Review of Science,*

[5] SILVA, V.; CHIARINI, T. The platformization of science: Towards a scientific digital platform taxonomy. **Minerva: A Review of Science, Learning and Policy**, 2022.

Learning and Policy. It seeks to historically characterize scientific digital platforms and aims to "trace a coherent and comprehensive interpretation of the emergence of scientific platforms". As an exploratory study, it proposes categories of scientific digital platforms: *e-portfolios, grid computing, science gateways, archives and repositories, citizen science platforms, academic social networks, publisher's platforms and crowdwork*. These platforms are characterized by being dedicated to at least one phase of the scientific process (*e.g.*, publication, experimentation). The chapter outlines a brief history of all the categories of scientific platforms, discusses how they are associated with different spheres (the State, the market and the academic community) and how much each of them stablished as a scientific infrastructure or a mere tool.

Although there are digital platforms developed and managed by the State, the market and the scientific community, they have different characteristics. Some are concentrated on specific phases of the scientific process. Private scientific digital platforms, for example, tend to target the final stages of publication and dissemination. They also differ in the degree of integration into the research lifecycle: some have infrastructure traits, such as grid computing platforms; others offer services that are complementary and have substitutes, such as the crowdwork platforms used to carry out surveys. The chapter illustrates the coexistence of digital platforms for science with different governance models, and that all of them can become infrastructures (with important consequences for the practice of scientific activities). Furthermore, it emphasizes how different scientific platforms innovated, anticipating trends that would become paradigmatic in digital innovation years or decades later. As examples, we can mention the distributed/open innovation that emerges with Archives, and organizational innovations to take advantage of idle assets, which guide the entire structuring of grid computing platforms.

The third chapter titled "The Platformization of Science: Lattes Platform at a Crossroads?" was originally published as an IPEA discussion paper[6]. This chapter is theoretically based on the discussion between digital platforms and infrastructure, given that the topic of infrastructure had only been touched upon in the previous chapter. Infrastructures is a classic area of social studies of science and technology (VAN DER VLEUTEN, 2004) and also a topic of science and technology policy (JUSTMAN; TEUBAL, 1995). Recently, digital platforms have proved capable of converting into infrastructures by exhibiting

---

[6] CHIARINI, T.; SILVA, V. **The platformization of science**: Lattes platform at a crossroads? Discussion paper / Institute for Applied Economic Research, 2022. Brasília: Rio de Janeiro: Ipea, 1990-. Available: https://portalantigo.ipea.gov.br/portal/images/stories/PDFs/TDs/ingles/dp_268.pdf.

characteristics such as criticality and invisibility. In addition, traditional infrastructures are fragmenting and platforming (PLANTIN *et al*., 2018). The case of the Lattes platform is paradigmatic as it is (i) a central element of the Brazil's scientific and technological infrastructure, (ii) a state platform and (iii) a pioneering case of a digital platform.

The case study of the Lattes platform demonstrates that it is not just the private sector that innovates. The State can innovate and use frontier technologies to do so. With more than two decades, the Lattes platform is a success story. However, as in the private sector, success can be an obstacle to overcome in the future. Trapped in the technical and organizational trajectory that ensured its success, the position of the Lattes platform is threatened by the emergence of new scientific digital platforms (such as ORCID and Academic Social Networks) and budget cuts. The new generation of digital platforms for science, mostly private and for-profit, has grown in size, influence and importance due to their network effects. The insinuation of private platforms over the scientific sphere emits an alert regarding the privatization of the new generation of scientific information infrastructure.

The third part of this dissertation is "The algorithmization of science". This part focus on the algorithmization trajectories identified in the first part/chapter, in the context of science. The fourth chapter is entitled "Framing the effects of machine learning on science". This chapter was originally published[7] in the journal *AI & SOCIETY: Journal of Knowledge, Culture and Communication*. The authors propose categories and sub-categories regarding the influence of technology on science based on the work of Nathan Rosenberg. With these categories structured as an analyticial framework, we investigate how artificial intelligence (its branches of machine learning and deep learning) has influenced the scientific sphere.

N. Rosenberg understood the avenues of influence of technology on science as mediating the economic endogeneity of science. Our approach goes further and also includes categories such as the intellectual effect: when a technology generates facts to be explained by science. Armed with these categories, it was possible to identify how the most recent advances in artificial intelligence have generated scientific responses in terms of the use of new tools, or the reinforcement of lines of research in AI. The article also relativizes the revolutionary power of AI over science, highlighting challenges such as the epistemological uncertainty associated with deep learning algorithms.

When reaching the end of the chapters, it is clear that science has always been very close to digital technologies. There is a two-way influence: science has generated new

[7] SILVA, V. J.; BONACELLI, M. B. M.; PACHECO, C. A. Framing the effects of machine learning on science. **AI & Soc.**, 2022.). DOI: https://doi.org/10.1007/s00146-022-01515-x.

digital technologies and scientific practices have even been at the forefront of the practical application of these new technologies; at the same time, the digital transforms the possibilities of scientific practice. Another point emerges from reading the chapters: although science has been at the forefront of technological and organizational innovation involving digital technology, in recent years there has been a decoupling, in the sense that scientific practices cannot (and perhaps should not?) keep the pace of transformation achieved by the system of digital technologies

Thus, if this thesis leaves any lesson, it is that it is necessary to rethink scientific institutions (KNOBEL; BERNASCONI, 2017) in order to better take advantage of the potential of digital. By clarifying how science can benefit from this rapprochement that already happens occasionally (BIANCHINI; MORITZ; PELLETIER, 2022; CHAN; KRISHNAMURTY; SADREDDIN, 2022; REICHENBACH; EBERL; LINDENMEIER, 2022), it will be possible to redesign the main scientific institutions for the digital 21st century. Hence, in the last part, we present a conclusion that offers a synthesis of the contributions of the thesis, the gaps in this research and possible paths for a research agenda.

<u>FIRST PART</u>: THE DIGITAL TECHNOLOGICAL REVOLUTION: EVOLUTIONARY AND INSTITUTIONAL ECONOMICS

CHAPTER 1 – THE DIGITAL TECHNOLOGY SYSTEM: ARTIFICIAL INTELLIGENCE, CLOUD COMPUTING AND BIG DATA[8]

Technologies makes worlds appear

Mercedes Bunz

1.1 Introduction

There is a movement of reinterpretation and labeling of the era in which we live and the candidates are many. Many of these interpretations share the view that digital technology in the current phase produces major changes in the socioeconomic structure. They differ in the emphasis they give to certain technologies: digital platforms would be responsible for the great wave of disruption, giving rise to a new era of platform capitalism (KENNEY; ZYSMAN, 2016; SRNICEK, 2017) or platform society (VAN DIJCK; POELL; WAAL, 2018). Others focus on the fantastical qualities of artificial intelligence (AI) and argue that, given its substitutive role for human labor, humanity would be heading towards a reissue of the Great Transformation (BALDWIN, 2019) or a Second Machine Age (BRYNJOLFSSON; McAFFE, 2014). Some emphasize general aspects of the digitization of business and economic activities and the genesis of digital capitalism (FUCHS; MOSCO, 2016).

Some authors argue that substantial socioeconomic changes depend not on just one, but on an interactive set of technologies. Perez and Freeman (1988) speak of constellations of innovations. Mokyr (1992) uses the concepts of microinventions and macroinventions. Nuvolari (2019) mentions blocks of development. Still in this sense, Perez (2009, p. 188) defines technological systems as follows: “the main innovations tend to induce new innovations; they require upstream and downstream complements and facilitate similar innovations, including competing alternatives [...] it is this kind of dynamic interrelation that is encompassed in the notion of a technological system”. Nuvolari (2019) states that current

[8] Originally published as: SILVA, Victo; BONACELLI, Maria Beatriz Machado; PACHECO, Carlos. O Sistema Tecnológico Digital: inteligência artificial, computação em nuvem e Big Data. **Revista Brasileira de Inovação**, n. 19, e0200024, p. 1-31, 2020. Available in: https://periodicos.sbu.unicamp.br/ojs/index.php/rbi/article/view/8658756

analyzes of digital technology are too narrow and too short. Narrow because, as mentioned, they do not consider the interrelationships that they may present and which may be the main cause of the generation of innovations and economic growth. Short, as they tend to assume too rapid cycles of innovation and diffusion for historical processes that take much longer.

Despite the recognition of the modularity and convergence of digital technologies, studies that address their interrelationships and the relationships between these technologies and the previous generation of information and communication technologies (ICT) are less frequent. From the point of view of temporality, there are authors who seek to extend the chronological cut. Among these studies, there are two fruitful approaches. One is based on major development spurs and refers to the current phase of updating the productive paradigm as the repetition of a common pattern in capitalism (PEREZ, 2002; 2009). Another is based on an even broader view of deep transitions (SCHOT; KANGER, 2018), whose reading of an industrial modernity (200-250 years) marked by discontinuities (*e.g.*: Perezian development spurs) and continuities (mechanization; fossil fuels) raises the possibility of overcoming it. Regarding the scope, Brynjolfsson and McAffe (2017) analyzed the interaction of artificial intelligence, platforms and distributed innovation (crowd). Its focus, however, is only on the new business models that emerge from this set. Mosco (2017) makes a good contribution by identifying the elements that, when interacting, produce what he calls the “next internet”: cloud computing, internet of things and Big Data. His interpretation is that these new technologies together would be subverting the democratic and participatory logic of the internet. His approach is systemic, but his focus is not the technology itself, but an analysis of political economy that seeks to understand which social actors exploit this process as it unfolds.

This article offers an appreciative approach to digital technologies that have matured (and emerged in public and academic debate) in the last decade. In this sense, we answer a call for “identifying and accessing the emergence and consolidation of autocatalytic connections between technology clusters is critical, but a topic still largely unexplored” (NUVOLARI, 2019, p. 38). This is done here in an exploratory way, using secondary data. In the next section, the interrelationships between the main digital technologies of the last decade (2010-2020) are analyzed in more detail. In the following section, the relationship of the so-called Digital Technological System (DTS) with preexisting trends in Information and Communication Technologies (ICTs) in the context of technical advances flows is discussed. Finally, in the third section after this introduction, the implementation lag that hinders the diffusion of these technologies, especially data rights, is discussed. Given the observed interrelationships, the technologies developed in the last decade are configured as a DTS. There is evidence of continuity linking this system with other systems of the ICT revolution. The process of commoditization of data, however, prompted by the maturation of the DTS, is evidence of a marked discontinuity that needs to be taken into account by the periodization and historiography of economics and technology.

1.2 The Digital Technological System (2010-2020)

The data economy (MAYER-SCHÖNBERGER; RAMGE, 2018; VELDKAMP; CHUNG, 2019) has been consolidated in the last decade (2010-2020), as a new technological system emerged. Which are the technologies that make up this system? Table 1.1 summarizes the vision of international organizations as to which digital technologies stand out in recent years as central to socioeconomic development. In line with Sturgeon (2019, p. 3), it is possible to aggregate the mentioned technologies into three large blocks: cloud computing, big data and artificial intelligence. They constitute the core of the digital technology system, as they are at the base or enable the others. This section briefly presents these three technology blocks and their relationships.

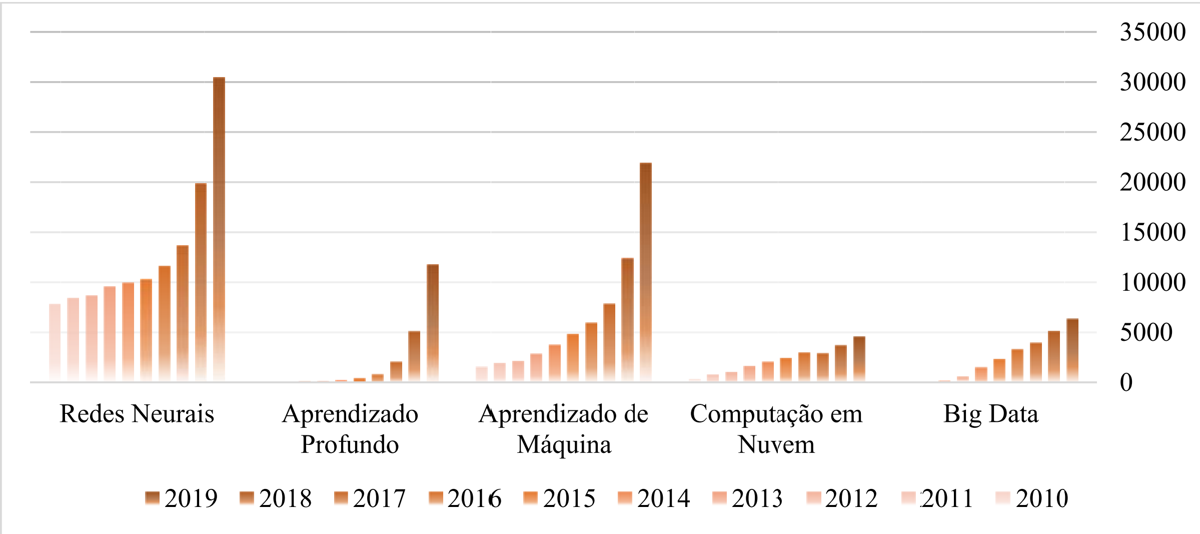
Table 1.1 – Digital technologies highlighted by international organizations

Organization	Technologies
United Nations Conference on Trade and Development (UNCTAD)	Blockchain, 3D printing, internet of things, 5G, cloud computing, automation and robotics and artificial intelligence/analytics
Economic Commission for Latin America and the Caribbean (ECLAC)	Internet of Things, blockchain, digital platforms, industrial internet of things, advanced manufacturing, artificial intelligence
European Union (EU)	Internet of Things, Big Data, Cloud Computing, Robotics, Artificial Intelligence, 3D Printing, Platforms

Source: UNCTAD (2019), CEPAL (2018), European Union (2016).

Figure 1.1 illustrates the growth of peer-reviewed articles in the Scopus database between 2010-2019 for three sets of technologies: AI (neural networks, deep learning and machine learning), cloud computing and big data. The sharp growth of publications in the area of AI in the last decade is noteworthy. Why did AI advance so much in the 2010s? Baldwin (2019) identifies two reasons for this: “much more computing power. It is Moore’s Law in action [...] and it is possible to collect, store and transmit large data sets” (BALDWIN, 2019, p. 110). Lee (2018) characterizes data and processing power as inputs for AI. In addition to these inputs, a breakthrough in the technique in the mid-2000s allowed for better models that resurrected the area of deep learning (AGRAWAL; GANS; GOLDFARB, 2018; LEE, 2018; NADELLA, 2018; VELDKAMP; CHUNG, 2019).

Figure 1.1 – Articles related to digital technologies selected in the Scopus database (2010-2019), in numbers



Source: *Scopus*, own elaboration

The fact that processors capable of supporting AI applications were developed would not be so decisive if there was not also the ability to distribute this computational power as, in the third technological revolution of the early 20th century, electrical energy was distributed. Hence the importance of cloud computing: "AI applications' needs in terms of computing power and access to large data sets make them predestined to run in the cloud" (COYLE; NGUYEN, 2019, p. 34).

Cloud computing is defined as "a system that moves data stored on individual computers and in institutions' IT departments to large distant datacenters operated by companies that charge for storage and usage" (MOSCO, 2017, p. 18). The idea of cloud computing is not something new. Centralized data servers were present at the beginning of the Internet. However, its current version is a little more sophisticated than that. Now, individual users can pay a fee to store data on these companies' servers, located in complex data centers, and the companies provide remote services for software and other applications. The possibility of outsourcing not just data storage or computing power, but also entire sections of a business, such as legal or the entire sales sector, is something new. This change and its importance for the democratization of AI are reflected in the vision that the Microsoft CEO has for the future of the company's cloud platform (Azure):

> We will offer cognitive services covering vision, speech, text, recommendations and facial, expression and emotion recognition. Developers will simply be able to use their APIs to extend user experiences, allowing solutions to see, hear, speak, and interpret the world around them. Our smart cloud will democratize these possibilities for startups and micro-enterprises as well as large organizations (NADELLA, 2018, p. 98).

Given the expansion of this elastic cloud servitization model, Reinsel, Gantz and Rydning (2018) predict that by 2025, 49% of the digital data stored in the world will be in public clouds to the detriment of traditional datacenters. The different types of services refer to three main levels, or stacks, of cloud computing services: infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The uses of cloud computing are manifold. UNCTAD (2019) highlights the usefulness of cloud computing for companies established in countries where the cost of software licensing is prohibitive. Also, "free" office apps are important for micro, small and medium businesses. The diffusion of cloud computing removes bottlenecks from the diffusion of digital technologies. Organizations willing to use artificial intelligence solutions do not need to invest capital in inflexible, dedicated systems of their own. They can consume these services elastically, on demand. This not only optimizes costs and eliminates waste in the economy as a whole, but also eliminates barriers to entry of "...advanced production techniques, such as artificial intelligence and robotic process automation" (COYLE; NGUYEN, 2019, p. 36-37). In turn, these digital technologies enable companies to put new data-driven business models into practice.

Processing power accessible via the cloud is the first enabler of the rise of machine learning. The second is data and its prominent role in society: "digital data is now a universal glue" (ATKINSON, 2013, p. 7). Sensors emerge (RFID), SMS, email messages, sustain social networks and geolocation, cartographic and satellite imagery, national security systems and also mundane transactions such as hitchhiking or making a reservation at a hotel or restaurant. Early reports on Big Data emphasized growth in volume, variety, and speed of data generation. UNCTAD (2019, p. 10) emphasizes the volume of data traffic: 100GB/s in 2002 jumped to 46,600GB/s in 2017. For 2022, the projected volume of traffic is around 150,000 GB/s.

Via Big Data Analysis, data has become an input to generate insights as well as valuable products and services (MAYER-SCHÖNBERGER; CUKIER, 2013). A five-step value chain allows one to generate value from data. The first three steps comprise the data management phase: (i) acquisition and storage; (ii) extraction and cleaning of the base; and (iii) integration and representation. The Big Data Analytics phase follows: (iv) modeling and analysis precede the (v) interpretation phase. Almost all of the data with the potential to generate value (95%) is not structured in spreadsheets, but semi-structured or unstructured, such as video, audio and unstructured text. There is a wide variety of techniques suitable for each of the five phases mentioned, summarized in Table 1.2 (GANDOMI; HAIDER, 2015).

Table 1.2 – Big Data Analytics techniques

Area	**Data Source**	**Techniques**
Text analysis/mining	Social networks, emails, blogs, online forums, questionnaires, reports, news, call center records	*Information extraction; text summarization; question answering; sentiment analysis*
Audio analysis	Call center data; Health area data	*Automatic-speech recognition; phonetic-indexing; search*
Video content analysis	Safety videos (internal circuits); decentralized video generation (YouTube)	*Server-based/edge-based architecture*
Social network analysis	Social networks, blogs, microblogs, media sharing, answer/question websites; wikis	*Content-based analytics; structure-based analytics (community detection; social influence analysis; link prediction)*

Source: own elaboration based on Gandomi e Haider (2015)

The result of these new techniques that make it possible to generate value from structured and unstructured databases is an even greater incentive to generate, store and flow data. Manyika *et al*. (2016) point to the stagnation of international flows of tangible goods, while the flow of data increased 45 times between 2005 and 2014. Big Data Analysis applications go through marketing, business intelligence, and decision-making process automation (VELDKAMP; CHUNG, 2019) and are associated with the emergence of modern AI algorithms.

Artificial intelligence (AI) is defined as “the theory and development of computer systems capable of performing tasks that normally require human intelligence” (AGRAWAL; GANS; GOLDFARB, 2019, p. 140-141). The history of modern AI dates back to the 1950s. The scientific AI community has divided into two major groups: the logical rules-based approach and the neural network-based approach. The first group, also called symbolists, defined *ex-ante* which rules the system should adopt to solve certain problems. The second group followed a learning approach, inspired by the architecture of the neuronal layers of the human brain: through examples, they created conditions for the processing in artificial neuronal layers to autonomously determine the final result of their operations (LEE, 2018). If in the first wave of AI (until the 1980s) there was great frustration with the learning approach, the second wave verified in the last decades happened thanks to it. In the last fifteen years, there has been a revolution in AI inputs (computer processing power and data availability) and a remarkable technical advance in the field of machine learning: the improvement of neural networks, now called deep learning systems (HINTON; OSINDERO; THE, 2006).

The Anglo-Canadian researcher Geoffrey Hinton (1947-) developed in the mid-2000s a new technique to improve the capacity of machine learning neural networks, discredited until that time. The scientific community did not pay much attention to his proposals. Until, in 2012, he and two other researchers at the University of Toronto developed a system (AlexNet) for autonomous classification in image recognition that “surprised the academic world” (ECLAC, 2018, p. 170) and gave new impetus to machine learning. Krizhevsky, Sutskever and Hinton (2012) register seventy thousand citations in Google Scholar and thirty thousand in Scopus. Its success promoted the dominance of machine learning techniques over its competitors (WIPO, 2019). In retrospect, it is quite possible that Hinton’s deep learning technique will be seen as a historic turning point, as important for the democratization of AI as the Bessemer process was for the democratization of steel production.

Its democratization allowed the emergence of new applications of AI as well as new areas of application in the economy. Machine translation is a good example of the emergence of new activities. Baldwin (2019, p. 128-129) narrates how, in 2016, Google applied machine learning to a massive database containing a corpus of languages available on the network and transformed the efficiency of automatic translations. The use of machine learning in medical diagnostics (BALDWIN, 2019, p. 176) is a good example of the transformation of a pre-existing area and, like machine translation, with enormous potential for generating social benefits. Broadly speaking, AI has lowered the cost of a crucial input to economic activity: prediction. Prediction is defined as “the process of filling in missing information. Prediction uses the information you have, often called ‘data’, to generate the information you don’t” (AGRAWAL; GANS; GOLDFARB, 2018, p. 24).

It is important to note that the successes of machine learning constitute advances in what is conventionally called Artificial Narrow Intelligence (or Narrow AI), as opposed to Artificial General Intelligence (or General AI). That is, they are highly efficient algorithms focused on singular tasks, unlike complex systems with flexible and multipurpose human-like intelligence. Lee (2018) defines the current field of machine learning as contextual to four elements: a huge dataset, a powerful algorithm, a limited scope and a clear objective. It is also important to point out that machine learning is just one area within the field of artificial intelligence. Although today there is a very clear technological trajectory (DOSI, 1982) to be followed for future innovations in machine learning, in the medium or long term, artificial intelligence can use other approaches (DOMINGOS, 2015).

The explosion in the volume of data enabled a seminal discovery in the field of artificial intelligence. Advances in machine learning have made data more valuable. The increase in the value of data demanded more transmission and storage networks, driving cloud computing. A technological interrelation is observed between Big Data, machine learning algorithms and cloud computing (LANDES, 1969, p. 2). As in past technological revolutions, the current one advances discontinuously, as complementary technologies (ROSENBERG, 1976) create mutual demands and resolve emerging bottlenecks. Nuvolari identifies this phenomenon as the autocatalytic nature of interrelated technologies: “innovations in one domain are, at the same time, dependent on innovations in another domain, as well as capable of enabling advances in related domains” (NUVOLARI, 2019, p. 38). Figure 1.2 summarizes the three components of DTS and their main features.

Figure 1.2 – Digital Technology System (2010-2020)

Ubiquitous data collection and big data analytics	Cloud computing	Artificial Intelligence (AI)
•Consumer and industry generated data streams from ubiquitous sensors and video monitoring, clickstreams, location data, "smart" products and machines, etc. •Data mining based on multivariate analysis techniques • Large sample sizes leading to more robust results, new insights and high fault tolerance	• Centralized storage and software as a service (SaaS) with on-demand and mobile access • Servers, databases, platforms (PaaS), analytics and intelligence (IaaS) are provided on demand • Data collection and processing of distributed applications (“edge”) • Constant and automated updating of software and systems	•From neural networks to machine learning, autonomy, prediction, replication and self-maintenance and increasing self-regulation •Mass customization for users •Better predictions and better business decision making •Perception of the physical world via pattern recognition (computer vision, speech-to-text technology)

Source: adapted from Sturgeon (2019); Lee (2018); Agrawal, Gans and Goldfarb (2018)

More studies are needed to uncover all the relationships between the technologies that make up the digital technological system (2010-2020). Despite this, evidence suggests that cloud computing, Big Data and AI make up a technological system that was able to overcome important bottlenecks in the diffusion and democratization of ICTs.

1.3 Streams of technical advances

In this section, we sought to put the DTS in motion, *i.e.*, to investigate its dynamics, which carry part of the history of these technologies (their links with previous systems). We resort to what Nuvolari (2019) calls “streams of technical advances”. He arrives at this term by analyzing three interrelated sets of technical advances in the work of David Landes, which were responsible for enabling the First Industrial Revolution: mechanization, ‘vaporization’ and new materials. If Landes speaks of this dynamic relationship in a historicized way and Nuvolari in general, Perez (2009, p. 194-195) indicates three main areas in which, analytically, it is possible to observe the streams of technical advances: (i) in the dynamics of the structure relative cost of productive inputs; (ii) in the perception of spaces for innovation; and (iii) organizational principles and criteria. It is argued that it is possible to identify trends in the three areas: digitization, algorithmization and platformization, which have been incorporated by companies, governments and other organizations.

The data stand out as the stream of technical advances with the greatest influence on the relative cost structure of productive inputs. Automation via machine learning depends on massive databases and reinforces the trend towards digitization, defined as: “[...] the work of transforming all kinds of information and media – text, sounds, photos, video, instrument and sensor data and so on – in the zeros and ones of the computer’s native language” (BRYNJOLFSSON; McAFFE, 2014, p. 61). Table 1.3 demonstrates the decreasing costs for the use and storage of data. Although economic theory has not yet found how to calculate the value of data (VELDKAMP; CHUNG, 2019), the reality is that for some pioneering firms, data is cheap. So cheap, that some researchers understand that there should be a reformulation of their market. Arrieta-Ibarra *et al.* (2018) argue that the scheme of exchanging “free” services for access to user data has generated an emphasis on the absorption of data that is not related to productivity. Productivity-related data is “stuck” with users – individuals, organizations and firms – not satisfied with the “free” bargain that is offered. Remunerating data holders, therefore, would be a way of accessing better quality and more relevant information for the purposes of productivity gains. Furthermore, the authors of these works argue that there would be an economic redistribution and a mitigation of the asymmetry of power between users and large digital conglomerates. The access to free data, therefore, seems to have been an asset of the pioneer companies, which harvested the "fruits in the lower branches". Reshaping this market would not make data expensive and inaccessible – it would make the market more efficient.

Table 1.3 – Decreasing costs in bandwidth, storage, and computing power (US$)

Service/Year	1992	2016
Processing	$ 222/ million of transistors	< $0,06/ million of transistors
Storage	$ 569/GB	< $0,01/GB
Transmition (banda larga)	$ 1.245/GB per second	< $10/GB per second

Source: Deloitte (2016)

In addition to being cheap, data is inexhaustible in the near future. As discussed in the section on Big Data, the growth trend in the volume of data generated and in the flow of data transacted continues. With the combination of new elements in the core DTS, such as the internet of things (IoT) and the industrial internet of things, the network of sensors and trackers (endpoints) generating more information tends to increase. Between 2015 and 2020, the accumulated data generated increased from approximately 20ZB to 40ZB. Between 2020 and 2025, the projected growth will be from 40ZB to 175ZB. This 100% advance over the last five years will be overshadowed by an increase of over 400% over the next five. Furthermore, by 2025, 75% of the global population (6 billion people) will interact with data on a daily basis. In this same year, a connected person will interact with data every 18 seconds on average – due to the billions of IoT devices connected to the network (REINSEL; GANTZ; RYDNING, 2018). It is important to note that these predictions are largely based on the diffusion of 5G technology, which will expand data transmission capacity. It is noteworthy here that digitization is anchored in a long-standing ICT trend: transmission capacity.

Data has applications in all human activities and, therefore, has been transforming markets (MAYER-SCHÖNBERGER; RAMGE, 2018). Data analysis has allowed efficiency and productivity gains in numerous sectors. Ferracane *et al*. (2018) warn that the removal of barriers to the free domestic and international circulation of data would raise the average productivity of firms, in addition to being responsible for the development of new markets and new service lines that did not exist before (OECD, 2015). This role of data has led experts to coin the term *data-driven innovation*: the use of data and data analysis to create and improve new products, processes and organizational models (OECD, 2015, p. 21).

The ubiquity of this new input directed the perception of spaces for innovation towards algorithmization. In this work, algorithmization is understood not only as the diffusion of the use of algorithms to deal with digitization, but as the emergence of algorithmic governance: "a form of social order based on the coordination between actors, based on rules, and that incorporates complex epistemic procedures based on computation" (KATZENBACH; ULBRICHT, 2019, p. 2). The authors emphasize algorithmic governance over algorithmic regulation to cover the different cases in which the coordination of social actors emanates from decentralized authorities, and not only from state regulation. They seek to understand how the use of algorithms enhanced coordination via governance, both producing beneficial effects (such as greater efficiency) and unwanted effects (such as selection bias). An algorithm-based decision system can be more or less transparent and guarantee more or less autonomy for users. They point to its applications and in several areas, *e.g.*: the efficient provision of public services and the surveillance of citizens (government use); the massification of individualized consumption, the filtering of content and the induction of user and consumer behavior (use by private digital platforms). "Algorithmic governance has many faces: it is seen as ordering, regulating and modifying behavior, as a form of management, optimization and participation" (KATZENBACH; ULBRICHT, 2019, p. 11). Kenney and Zysman (2016) argue that the algorithmic revolution is the technical basis of the platform economy. Computational power generates economic value because the algorithms act on databases. Furthermore, they see the software layer that covers not only services but also manufacturing (generating the so-called Industry 4.0) as an "algorithm framework". Algorithms are combined with digital data, platforms and intelligent tools to compose the technical foundation of the digital revolution (ZYSMAN; KENNEY, 2018). They enable the automation of services, bringing technological advances to an area previously considered far less prone to technological automation than manufacturing.

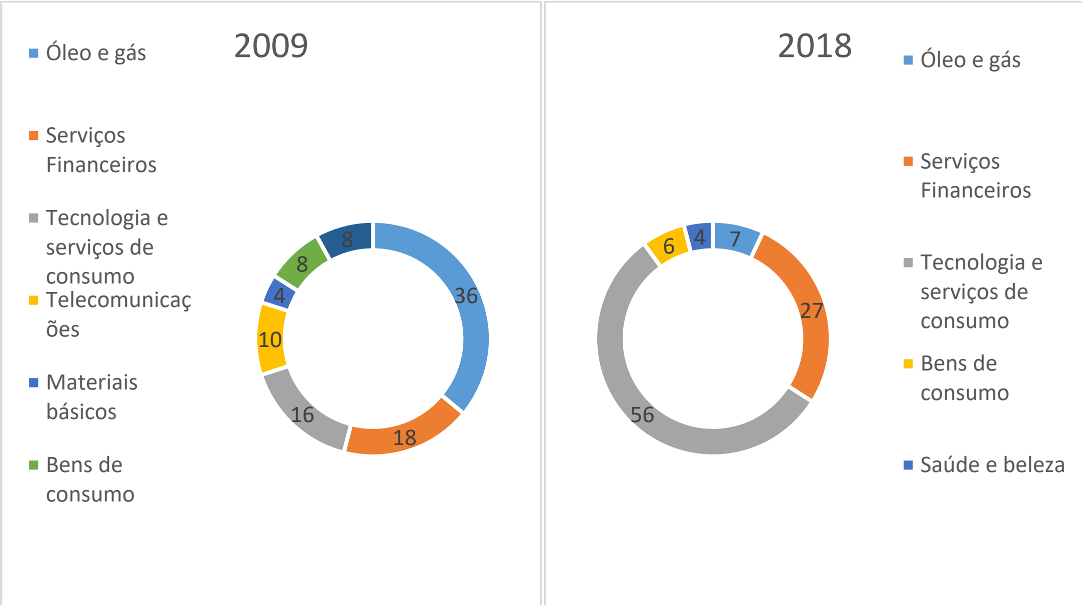
One of the processes by which this automation occurs is narrated by Mayer-Schönberger and Ramge (2018). The authors' thesis is that the Market has evolved from a one-dimensional transaction mechanism (price) to a multidimensional transaction mechanism (data). What they call data-rich markets is an evolution of the traditional market. With more data, you can generate more efficient transactions. However, it is necessary that: (i) the data are correctly labeled; (ii) a mechanism that facilitates the optimal combination between suppliers and demanders (matching) exists; and (iii) autonomous systems record user preferences. All these functions are performed by machine learning algorithms. These data-rich marketplaces have spawned almost all new categories of digital services: retail (Amazon, eBay, Buscapé), service delivery (AirBnb, 99, iFood), job brokerage (UpWork), platforms that make digital tools available online and support the creation of other platforms and marketplaces (GitHub). With the "mission of humanity to digitize the world" (REINSEL; GANTZ; RYDNING, 2018) in full swing, the perspective that the O2O (offline to online) movement will expand, allowing more and more applications of this combination of algorithms/big data.

It should be noted that algorithmization directs other innovation trajectories: those of specific processors. There is a great race for the development of AI chips. Each era of computing requires specific characteristics of chips. Those developed by Qualcomm and Intel fueled bygone eras. Machine learning requires "the rapid execution of complex mathematical formulas, something neither Intel's nor Qualcomm's chips were built for" (LEE, 2018, p. 119). According to Nadella (2018), temporary solutions so far have resorted to "creative accelerators such as graphics processing unit (GPU) farms, tensor processing unit (TPU) chips, and field programmable gate arrays (FPGA) in the cloud" (NADELLA, 2018, p. 168). Having great experience in GPUs, Nvidia has emerged as a major supplier of AI chips, but Google and Microsoft invest in developing specific chips (which they have historically avoided doing), while Intel and Qualcomm race to accelerate development of their own chips. There is also Chinese development, heavily subsidized by the State, to develop chips for AI (LEE, 2018).

With regard to organizational principles and criteria (PEREZ, 2009), digital platforms stand out. Technologically, "today's digital platforms consist of data processing software in the cloud" (ZYSMAN; KENNEY, 2018, p. 56), that is, they are the convergence of other streams of technical advances. However, the best way to understand what digital platforms are broadly is in Cohen (2019): platforms are friction points in networks. Its definition is simple, counter-intuitive and universal. Counterintuitive because economic studies on platforms reinforce their ability to generate savings in transaction costs (GOLDFARB; TUCKER, 2019). Nevertheless, Cohen understands that platforms were the way out that companies found to appropriate the value generated in a networked environment, in which horizontality prevailed over hierarchy and linked to the commercialization of non-rival goods, increasingly dependent on intellectual property. to earn income. Becoming a point of friction in the network, a new type of intermediary, guarantees companies the possibility of appropriating monetary and data rents. The definition is universal because it does not establish digital platforms in terms of market intermediaries. In reality, the platform is a specific configuration of networks, which can be replicated in numerous contexts, such as by governments seeking to upgrade to digital government, with different degrees of success (BLASIO; SELVA, 2019).

Despite the ability to transpose this organizational model, the platforms were consolidated by the market selection of best business practices. Although data has become the raw material of capitalism in the 21st century, “old business models were not particularly well designed to extract and use data” (SRNICEK, 2017, p. 30). Among the companies that show greater dynamism in the application of digital intelligence are digital platforms: “reprogrammable digital infrastructures that facilitate and shape personalized interactions between end users and complementers, organized through systematic collection, algorithmic processing, monetization and circulation of data” (POELL; NIEBORG; VAN DIJCK, 2019, p. 3). In some cases (generally, transactional platforms, as will be seen below) they act as market organizers (FRENKEN *et al*., 2018). Platforms take on the task of planning the market: structuring it, balancing supply and demand, coordinating agents interested in negotiating in this market, providing curation of inputs, in short, *market design* (ROTH, 2015). The platform architecture, despite admitting variations, is based on orchestrating the interactions between different economic agents, so that they can generate and exchange value. This organizational model, therefore, does not fit into the traditional categories of companies. Parker, Van Alstyne and Choudary (2016) cite three traditional categories – asset producers, service providers, technology manufacturers – and the newcomer: network orchestrators (PARKER; VAN ALSTYNE; CHOUDARY, 2016, p. 32). Based on ample computing power, these “network orchestrators” structured markets as other analog/traditional intermediaries could not. The platform’s own value proposition rests on the mechanism of positive network externalities (PARKER; VAN ALSTYNE; CHOUDARY, 2016). These mechanisms have provided extraordinary growth for companies based on platforms such as GAFAM (Google/Alphabet, Amazon, Facebook, Apple, Microsoft) in the US. Figure 1.3 illustrates how the technology sector has taken the top positions in the ranking of largest companies by market cap over the past decade.

Figure 1.3 – Top 20 companies by market capitalization, by sector, 2009 vs. 2018



Source: UNCTAD, 2019

Its technological structure allows platform controllers to establish ecosystems that involve a fixed center (owners/controllers) and a dynamic periphery (partners/users/complementers) (CUSUMANO; GAWER; YOFFIE, 2019; TIWANA, 2014; WAREHAN; FOX; GINER, 2014). Ecosystems are understood as organizational forms that, *ab initio*, involve more than one organization (GAWER, 2014; JACOBIDES; CENNAMO; GAWER, 2018). This inter-organizational model is, in a way, the digital counterpart of global value chains. In this new arrangement, value flows from the center to the periphery in the form of digital services and flows from the periphery to the center in the form of money, attention or data (VAN DIJCK; POELL; WAAL, 2018). The combination of digital platforms and DTS results in a techno-organizational structure that combines "decentralized data generation with centralized analytics" (VAN DOORN; BADGER, 2020, p. 3).

The new generation of organizations observed that a lot of value is created outside the company: in the chaotic network, in the first moment; in the structured ecosystem, in the second. The consolidation of digital platforms is the transition from the generic and horizontal network to the hierarchically structured ecosystem. They are able to "turn the company around," as Parker, van Alstyne, and Jiang claim:

> Using loosely affiliated ecosystems, firms are able to harness a global network of partners they don't even know beforehand who can con- nect through digital networks to innovate on top of a core set of resources thereby creating highly valuable products and services for their users (PARKER; VAN ALSTYNE; JIANG, 2016, p. 2).

This domain may be related to the trend pointed out by Mayer-Schönberger and Ramge (2018) that the explosion of data availability strengthens the Market while weakening the Firm. The platform emerges as a response from firms to capture the value being generated externally – controlling a captive market, what Srnicek (2017) calls *siloing*. The governance of a market, one of the main functions of the platform, would have rebalanced the game for the firms. Therefore, the new organizational unit, born digitally and capable of dealing with data, algorithms and digital technologies *latu sensu*, is the platform ecosystem. Its economic preponderance is attested by the figures in Table 1.4.

Table 1.4 – Average values for Forbes Global 2000, control sample and platforms, 1995-2015

Variable	Forbes Global 2000	Control sample	Innovation & transaction platforms	Innovation platforms	Transaction platforms
Number of firms	1939	100	43	18	25
Sales (US$ millions)	5,586	4,845	4,335	10,118	2,119
Employees	18.900	19.000	9.872	26.600	6.349
Operational profit	13%	12%	21%	21%	21%
Market value (US$ millions)	6,876	8,243	21,726	37,901	13,277
R&D/sales	4%	9%	13%	13%	11%
Market cap. growth	10%	8%	14%	12%	21%

Source: adapted from Cusumano, Gawer and Yoffie, 2019

It is noteworthy how innovative platforms surpass even transactional ones in terms of market value (Table 1.4). This is because their value proposition does not reside only in the coordination (or algorithmic governance) they exercise, but also in the co-generation of value with third parties. There is a “distinct nature of the logic of coordination and organization in ecosystems compared to other forms of organization (eg markets, hierarchies, alliances)” (CENNAMO; SANTALÓ, 2019, p. 3). This new logic of institutional organization (FRENKEN *et al*., 2020) develops as digital platforms advance over functions previously attributed to other institutions: coordination (market logic), pricing; critical selection and promotion of partners, content curation (firm logic), self-regulation (state logic). Poell, Nieborg and van Dijck (2019, p. 5-6) define platformization as “the penetration of infrastructure, economic processes and government frameworks based on platforms in different economic sectors and spheres of life, as well as the reorganization of cultural practices and imagination around these platforms." They are strong candidates for the fundamental technical-organizational innovation of DTS, and as developed here, they also demonstrate autocatalytic interrelationships with that one. set (NUVOLARI, 2019).

Just as there is a strong relationship between the technologies that make up the DTS, the flows of technical advances derived from it are closely related to each other and between previous technological systems. The classic trends of improvements in (i) storage, (ii) processing and (iii) data transmission are at the root of current advances flows, or are even more leveraged by them, in yet another demonstration of the autocatalytic nature of ICTs: platforms cloud solutions solve storage bottlenecks as well as on-demand availability (elasticity) of processing power; these innovations in processing power have benefited and benefited from advances in artificial intelligence; both trends reinforce the dynamo of digitization and push for advances in transmission capacity (5G). In order to appropriate monetary income and data income, digital platforms multiply, fragmenting the Internet in the process.

There seems to be enough evidence to characterize DTS as part of a larger block of information and communication technologies whose development is ongoing. Perez notes that technological revolutions can be defined as "a set of interrelated radical advances, forming a large constellation of interdependent technologies; a cluster of clusters or a system of systems" (PEREZ, 2009, p. 189). In a similar way, Nuvolari (2019) states that the Third Industrial Revolution depends on the advancement of a block of development that involves four large clusters: semiconductors, computers, software and network equipment. Perhaps the DTS is a strong candidate for a fifth element.

1.4 Implementation Lags

It does not escape economic observers that the euphoria over new technologies has so far not been reflected in productivity statistics. One of the explanations for this productivity paradox is implementation lags: "the deeper and far-reaching the restructuring potential [of the new technology], the longer the time lag between the initial invention of the technology and its full impact on the economy and in society" (BRYNJOLFSSON; ROCK; SYVERSON, 2017, p. 10).

There are two main reasons for this. The first is that it takes time to build up the stock of new technologies, enough to be able to measure their impact in aggregate statistics. The second reason for implementation lags is the fact that "additional investments are required to obtain the full benefits of the new technology, and it takes time to discover and develop these add-ons and implement them" (BRYNJOLFSSON; ROCK; SYVERSON, 2017, p. 10). The authors explore how the pervasiveness of AI spurs the development of complementary innovations, *e.g.*: self-driving cars, one of its most popular applications. Bresnahan (2019, p. 347) states that the technical adaptations necessary to import DTS from B2C sectors to B2B sectors (application sectors) are in the focus of industrialists and executives.

These complementary investments, however, are not just technical. Acemoglu and Restrepo (2020, p. 32), also commenting on the development of AI, state that "Educational applications of AI would need new, more flexible skills from teachers". By mentioning human resource competencies, they enter another sphere of complementary innovations. In a similar vein, it is argued in this work that the DTS requires the updating/creation of important institutions for its full development. In addition to AI, the importance of "discovering and developing" institutions is emphasized so that society can deal with the new role of data.

First, data increases in value by being complementary goods to other DTS technologies. As AI advances and the cost of prediction decreases, the value of the data generally grows (AGRAWAL; GANS; GOLDFARB, 2018, p. 43). Second, unlike other DTS technologies, data provision is decentralized in nature. Meanwhile, cloud computing (and AI/Big Data Analytics applications linked to it) is dominated by an oligopoly that generated $111 billion in revenue between 2019-2020, as shown in Table 1.5. In light of this, data has redistributive potential. Finally, the formation of data assets is a *sui generis* commodification process (ZUBOFF, 2019; VAN DOORN; BADGER, 2020).

Table 1.5 – Global Cloud Computing Oligopoly, 2020

Firm/Cloud solution	Cloud computing market share (%)
Amazon/Amazon Web Service (AWS)	33
Microsoft/Microsoft Azure	18
Alphabet/Google Cloud	9
Alibaba/Alibaba Cloud	6
IBM/IBM Cloud	6
Tencent/Tencent Cloud	2
Oracle/Oracle Cloud	2

Source: Richter (2020)

UNCTAD (2019, p. 131) points out three central themes: the definition of data ownership and control, consumer privacy and trust, and regulation of international data flows. Experts propose individual data remuneration, insisting that the personal data market will be a more efficient way to link content generators with stakeholders (ARRIETA-IBARRA *et al*., 2018). Alternatives have been discussed: custody of data by a centralized entity, which could increase the bargaining power of data generators; and public data funds based on collective ownership, which could combine individualized control over the type of data shared with adequate remuneration from companies interested in the dataset. The funds raised would be invested in public goods. Finally, there is the proposal of data commons, sets of data that would be provided free of charge to non-profit enterprises (UNCTAD, 2019).

Despite these proposals, it still seems that it will take some time for authorities at different levels (national/international) to build a normative framework for data rights (VAN DOORN; BADGER, 2020). This task involves not just a new product or a new market emerging from new technologies, but a new asset class. The commoditization process is “marked by taking things that live outside the sphere of the market and declaring their new life as market assets” (ZUBOFF, 2019). This perception may be related to the Polanyian renaissance (POLANYI, 1944) of historical interpretation: the view that a fictitious new commodity is in the process of formation has led several analysts to make comparisons between the present moment and past industrial revolutions, when land, capital and labor went through similar processes (ATHIQUE, 2019; CHEN *et al*., 2020; GRABHER; KÖNIG, 2020).

The slow progress in this process, according to the present work, is a plausible explanation of the delay in the potential of DTS to be reflected in productivity statistics. And not only in productivity statistics, but in general in their more complete diffusion through the economic fabric, regardless of whether or not the total productivity of factors may be raised. Claims that a supposed fourth industrial revolution would be “rewriting the rules of manufacturing” (LEE; MALERBA; PRIMI, 2020) must be put into perspective from this point of view. There is evidence that data governance, more than the current state of technologies, has hindered the diffusion of DTS. Culot *et al*. (2020) conduct a delphi to obtain the perceptions of industry managers on the risks and benefits of the diffusion of STD to the industry. They find widespread concern about the changing hands of value chain (VC) control points, due to data flow/sharing:

> further analyses are required as the ways in which Industry 4.0 is changing manufacturing VCs' "control points" – i.e., which activities along the VC hold the greater value or power – within increasingly complex networks of business partners and competitors. Data ownership, control over sales channels, standardization of IoT product-service platforms emerged from our study as increasingly relevant elements, and still occupy a contested territory between manufacturing incumbents and born-digital companies (CULOT *et al*., 2020, p. 27).

Data governance and ownership may explain the difference in DTS diffusion between consumer-oriented ICT-based companies and the industrial sector. While consumer-oriented companies have restructured in terms of platforms and use of DTS, manufacturing is still moving in this direction (BRESNAHAN, 2019). Consumer platforms do not face the specificity of assets present in the industry, nor the heterogeneous demands found there, so they are easier to grow and leverage their network effects (STURGEON, 2019). Furthermore, the average consumer is not concerned (perhaps incorrectly) about their data sharing or the possibility of becoming dependent on a platform for a specific service, whereas in the industry these considerations are of strategic importance. Hence the lower penetration of B2B platforms compared to B2C (KENNEY *et al*., 2019; STURGEON, 2019). As seen, the relationship between digital technologies and platforms also has autocatalytic traits (NUVOLARI, 2019). In the absence of these relationships in the industry, the streams of complementary technical advances are potentially impeded.

1.5 Conclusions

In the present study, the digital technology system (DTS) (2010-2020) was analyzed from a historical and evolutionary perspective and it was also argued that it has artificial intelligence, cloud computing and Big Data as its core technologies. The relationships between the three core technologies demonstrate important feedback mechanisms or autocatalytic relationships. In a second moment, the flows of technical advances driven by DTS were analyzed. Digitization, algorithmization and platformization constitute broad trajectories of technical and organizational innovation with an active interface with previous ICT systems (digitization involves storage/transmission; algorithmization depends on processing power; platformization is a special configuration of network logic). Finally, the importance of complementary technical and institutional innovations that allow the full diffusion of DTS throughout the socioeconomic fabric was addressed. In particular, the role of institutions that will provide the governance of a new asset class: data was emphasized.

Due to the fact that the DTS presents strong traces of interrelation between its components, there is a great possibility that government incentives for research in one of the components should take into account its relationship in the DTS; in addition to research incentives, regulation is another sensitive point that must be thought of in a systemic way. Industrial policies and data policies must consider the interrelated aspect of the technological dimensions addressed in the article. Furthermore, by explaining the autocatalytic relationships between the components of the DTS, blocking or misdirecting one of them can lead to socially undesirable results in other technological trajectories as well. Acemoglu and Restrepo's (2020) warning, of an AI trajectory that prioritizes replacement over complementing human work (good AI vs. bad AI), is an example and may reflect the impulses received from the evolution of the other technologies that make up the DTS. It is also possible to reason that the use of data by platforms for triangulation with targeted advertising (ZUBOFF, 2019) can affect autocatalytic interactions of algorithmic improvement related to the services provided to users, thus hindering the technological advancement of the system as a whole.

Also from the fact that the DTS demonstrates strong interrelationships with the preceding technological systems, there is historical evidence of continuity: the DTS is configured as an advanced phase of the ICTs that emerged in the 1970s. was observing the "era of implementation" (LEE, 2018) of what began decades ago. In this case, the Perezian interpretation gains strength, that is, what is happening is an advanced phase of a great spurt of development based on ICTs that began decades ago. This Perezian great surge of development is also called the Third Industrial Revolution and

> when one considers that the life cycles of the technological systems of the first and second industrial revolution were long drawn-out processes, the current technological trends can be possibly more insightfully characterized as a further advancement of the ICT revolution (NUVOLARI, 2019, p. 42).

From the fact that the DTS involves and drives a *sui generis* commodification process, a strong argument for discontinuity is obtained. The Polanyian process of the formation of fictitious commodities took place between the 18th and 19th centuries, when land, labor and capital became market assets, not without broad social resistance and institutional mismatches. The conversion of data (referring to all kinds of objects, including human experiences) into market assets is, analytically, a similar process. The reading of the deep transition (SCHOT; KANGER, 2018), in this case, gains strength. It is noteworthy that, if this proposition is correct and digital technologies constitute the basis not only of a new development spurt (measured in decades), but of a new profound transition (measured in centuries), it becomes more crucial. Also the development of technological and institutional competences for national incursions in this stage of human development.

Perhaps it is possible to reconcile the two interpretations if one considers a transitional development spurt. The development spurt associated with ICT (Third Industrial Revolution) would be configured as the last of the first deep transition. Exhibiting socio-economic processes of expansion of the market logic to other social spheres, it would take longer than usual to enter its stages of maturity. In fact, the difficulty in renewing institutions to deal with a new and sui generis asset such as data seems to be at the root of the slowness in the diffusion of DTS among the industrial sector, exactly the sector that could spread the benefits on an even greater scale. arising from the adoption of these technologies.

1.6 References

ACEMOGLU, D.; RESTREPO, P. The wrong kind of AI? Artificial intelligence and the future of labour demand. **Cambridge Journal of Regions, Economy and Society**, v. 13, p. 25-35, 2020. DOI: https://doi.org/10.1093/cjres/rsz022.

AGRAWAL, A.; GANS, J.; GOLDFARB, A. Economic Policy for Artificial Intelligence. **Innovation Policy and the Economy**, v. 19, n. 1, p. 139-159, 2019.

AGRAWAL, A.; GANS, J.; GOLDFARB, A. **Máquinas Preditivas**: a simples economia da inteligência artificial. Rio de Janeiro: Alta Books, 2018.

ARRIETA-IBARRA, I. *et al*. Should we treat data as labor? Moving beyond "free". **American Economic Association Papers and Proceedings**, v. 108, p. 38-42, 2018.

ATHIQUE, A. Integrated commodities in the digital economy. **Media, Culture & Society**, v. 42, n. 4, p. 554-570, 2019. DOI: https://doi.org/10.1177/0163443719861815.

ATKINSON, M. *et al.* (ed.). **The Data Bonanza**: Improving Knowledge Discovery in Science, Engineering, and Business. Hoboken, New Jersey: Wiley, 2013.

BALDWIN, R. **The globotics upheaval**: globalization, robotics and the future of work. New York, NY: Oxford University Press, 2019.

BLASIO, E.; SELVA, D. Implementing open government: a qualitative comparative analysis of digital platforms in France, Italy and United Kingdom. Quality & Quantity, v. 53, p. 871-896, 2019. DOI: https://doi.org/10.1007/s11135-018-0793-7.

BRESNAHAN, T. Technological change in ICT in light of ideas first learned about the machine tool industry. **Industrial and Corporate Change**, v. 28, n. 2, p. 331-349, 2019. DOI: https://doi.org/10.1093/icc/dty076.

BRYNJOLFSSON, E.; McAFFE, A. **Machine, platform, crowd**: harnessing our digital future. New York: W. W. Norton & Company, 2017.

BRYNJOLFSSON, E.; McAFFE, A. **The Second Machine Age**: Work, Progress, and Prosperity in a Time of Brilliant Technologies. New York: W. W. Norton & Company, 2014.

BRYNJOLFSSON, E.; ROCK, D.; SYVERSON, C. **Artificial Intelligence and the Modern Productivity Paradox**: a clash of expectations and statistics, 2017. (NBER Working Paper Series, n. 24001). Disponível em: http://www.nber.org/papers/w24001. Acesso em: 28 ago. 2020.

BUNZ, M. **The silent revolution**: how digitalization transforms knowledge, work, journalism and politics without too much noise. Palgrave MacMillan, 2014.

CENNAMO, C.; SANTALÓ, J. Generativity Tension and Value Creation in Platform Ecosystems. **Organization Science**, v. 30, n. 6, p. 447-646, 2019. DOI: https://doi.org/10.1287/orsc.2018.1270.

CEPAL. **Datos, algoritmos y políticas**: la redefinición del mundo digital. Santiago, Chile: United Nations, 2018.

CEPAL. **Science, technology and innovation in the digital economy**: the state of the art in Latin America and the Caribbean. Santiago, Chile: United Nations, 2016.

CHEN, B. *et al*. The disembedded digital economy: Social protection for new economy employment in China. **Soc Policy Adm**, v. 54, n. 7, p. 1246-1260, 2020.

COHEN, J. **Between truth and power**: the legal constructions of informational capitalism. New York, NY: Oxford University Press, 2019.

COYLE, D.; NGUYEN, D. Cloud computing, cross-border data flows and new challenges for measurement in economics. **National Institute Economic Review**, n. 249, p. 30-38, 2019.

CULOT, G. *et al*. The future of manufacturing: A Delphi-based scenario analysis on Industry 4.0. **Technological Forecasting & Social Change**, v. 157, 2020. DOI: https://doi.org/10.1016/j.techfore.2020.120092.

CUSUMANO, M.; GAWER, A.; YOFFIE, D. **The business of platforms**: strategy in the age of digital competition, innovation and power. New York: Harper Business, 2019.

DELOITTE. **Exponential technologies in manufacturing**: transforming the future of manufacturing through technology, talent, and the innovation ecosystem, 2018.

DELOITTE. **The rise of the digital supply network**: Industry 4.0 enables the digital transformation of supply chains, 2016.

DIEGUES, A. C.; ROSELINO, J. E. **Indústria 4.0 e as redes globais de produção e inovação em serviços intensivos tecnologia**: uma tipologia e apontamentos de política industrial e tecnológica. Campinas, Unicamp. IE, 2019. (Texto para Discussão, n. 356).

DOMINGOS, P. **The Master Algorithm**: how the quest for the ultimate learning machine will remake our world. New York: Basic Books, 2015.

DOSI, G. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. **Research Policy**, v.11, n. 3, p. 147-162, 1982.

DOSI, G.; VIRGILLITO, M. Whither the evolution of the contemporary social fabric? New technologies and old socio-economic trends. **International Labour Review**, v. 158, n. 4, p. 593-625, 2019.

EUROPEAN UNION. **Digitising European Industry**: reaping the full benefits of a Digital Single Market. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. European Commission, 2016.

EVANGELISTA, R.; GUERRIERI, P.; MELICIANI, V. The economic impact of digital technologies in Europe. **Economics of Innovation and New Technology**, v. 23, n. 8, p. 802-824, 2014. DOI: https://doi.org/10.1080/10438599.2014.918438.

FERRACANE, M.; LEE-MAKIYAMA, H.; VAN DER MAREL, E. **Digital Trade Restrictiveness Index**. Brussels: European Center for International Political Economy – ECIPE, 2018.

FREEMAN, C., PEREZ, C. Structural crises of adjustment: business cycles and investment behavior. *In:* DOSI, G. *et al.* (ed.). **Technical change and economic theory**. London, New York: Printer publishers, 1988. p. 38-66.

FRENKEN, K. *et al*. An Institutional Logics Perspective on the Gig Economy. *In*: MAURER, I.; MAIR, J.; OBERG, A. (ed.). **Theorizing the Sharing Economy**: Variety and Trajectories of New Forms of Organizing. Bingley, UK: Emerald Publishing Limited, 2020.

FUCHS, C.; MOSCO, V. (ed.). **Marx in the age of digital capitalism**. Leiden, UK, Boston, US: Brill, 2016.

GANDOMI, A.; HAIDER, M. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, v. 35, n. 2, p. 137-144, 2015.

GAWER, A. Bridging Differing Perspectives on Technological Platforms: Toward an Integrative Framework. **Research Policy**, v. 43, n. 7, p. 1239-1249, 2014.

GOLDFARB, A.; TUCKER, C. Digital economics. **Journal of Economic Literature**, v. 57, n. 1, p. 3-43, 2019. DOI: https://doi.org/10.1257/jel.20171452.

GRABHER, G.; KÖNIG, J. Disruption, embedded. A Polanyian framing of the platform economy. **Sociologica**, v. 14, n. 1, 2020.

HINTON G.; OSINDERO, S.; THE, Y. A fast learning algorithm for deep belief nets. **Neural Comput**, v. 18, n. 7, p. 1527-1554, 2006.

JACOBIDES, M.; CENNAMO, C.; GAWER, A. Towards a theory of ecosystems. **Strategic Management Journal**, v. 39, n. 8, p. 2255-2276, 2018.

KATZENBACH, C.; ULBRICHT, L. Algorithmic Governance. **Internet Policy Review**, v. 8, n. 4, p. 1-18, 2019. DOI: https://doi.org/10.14763/2019.4.1424.

KENNEY, M. *et al*. Platforms and industrial change. **Industry and Innovation**, v. 26, n. 8, p. 871-879, 2019. DOI: https://doi.org/10.1080/13662716.2019.1602514.

KENNEY, M.; ZYSMAN, J. The rise of the platform economy. **Issues in Science and Technology**, v. 32, n. 3, p. 61-69, 2016.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. ImageNet classification with deep convolutional neural networks. **Advances in neural information processing systems**, p.1097-1105, 2012.

LANDES, D. **The Unbound Prometheus**. Cambridge, England: Cambridge University Press, 1969.

LEE, K. **AI Super-powers**: China, Silicon Valley and the New World Order. Boston, New York: Houghton Mifflin Harcourt, 2018.

LEE, K.; MALERBA, F.; PRIMI, A. The fourth industrial revolution, changing global value chains and industrial upgrading in emerging economies. **Journal of Economic Policy Reform**, v. 23, n. 4, p. 359-370, 2020. DOI: https://doi.org/10.1080/17487870.2020.1735386.

LIU, J. *et al*. Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors. **Technological Forecasting & Social Change**, v. 158, 2020. DOI: https://doi.org/10.1016/j.techfore.2020.120142.

MANYIKA, J. *et al*. **Digital globalization: the new era of global flows**. New York: McKinsey Global Institute, 2016.

MAYER-SCHÖNBERGER, V.; CUKIER, K. **Big Data**: a revolution that will transform how we live, work and think. Boston, New York: Eamon Dolan Book, Houghton Mifflin Harcourt, 2013.

MAYER-SCHÖNBERGER, V.; RAMGE, T. **Reinventing capitalism in the age of Big Data**. New York: Basic Books, 2018.

MOKYR, J. **The Lever of Riches**: Technological Creativity and Economic Progress. Oxford, UK: Oxford University Press, 1992.

MOSCO, V. **Becoming digital**: toward a post-internet society. UK: Emerald Publishing: Bingley, 2017.

NADELLA, S. **Aperte o F5**: a transformação da Microsoft e a busca de um futuro melhor para todos. São Paulo: Benvirá, 2018.

NELSON, R. The co-evolution of technology, industrial structure, and supporting institutions. **Industrial and Corporate Change**, v. 3, n. 1, p. 47-63, 1994.

NUVOLARI, A. Understanding successive industrial revolutions: A "development block" approach. **Environmental Innovation and Societal Transitions**, v. 32, p. 33-44, 2019. DOI: https://doi.org/10.1016/j.eist.2018.11.002.

OECD. **Data-Driven Innovation**: Big Data for Growth and Well-Being. Paris: OECD Publishing, 2015. DOI: https://doi.org/10.1787/9789264229358-en.

OWEN, T. **The case for platform governance**. Waterloo, Canada: Center for International Governance Innovation, 2019.

PARKER, G.; VAN ALSTYNE, M.; CHOUDARY, S. **Platform revolution**: how networked markets are transforming the economy and how to make them work for you. New York, London: W.W. Norton & Company, 2016.

PARKER, G.; VAN ALSTYNE, M.; JIANG, X. Platform ecosystems: How developers invert the firm. **MIS Quarterly**: Management Information Systems, v. 41, n. 1, p. 255-266, 2016.

PEREZ, C. Structural change and the assimilation of new technologies in the economic and social systems. **Futures**, v. 15, n. 5, p. 357-375, 1983.

PEREZ, C. **Technological revolutions and financial capital**: the dynamics of bubbles and golden ages. Cheltenham: Edward Elgar, 2002.

PEREZ, C. Technological revolutions and techno-economic paradigms. **Cambridge Journal of Economics**, v. 34, n. 1, p. 185-202, 2009.

PLANTIN, J. *et al*. Infrastructure studies meet platform studies in the age of Google and Facebook. **New Media & Society**, v. 20, n. 1, p. 293-310, 2018.

POELL, T.; NIEBORG, D.; VAN DIJCK, J. Platformisation. **Internet Policy Review**, v. 8, n. 4, 2019. DOI: https://doi.org/10.14763/2019.4.1425.

POLANYI, K. **The Great Transformation**: The Political and Economic Origins of Our Time. Beacon Press, Boston, MA, 1944. Reimpressão em 2001.

REINSEL, D.; GANTZ, J.; RYDNING, J. **The digitization of the World**: from Edge to Core. IDC White Paper, 2018.

RICHTER, F. Amazon Leads $100 Billion Cloud Market. **Statista**, 18 Aug. 2020. Disponível em: https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/. Acesso em: 25 ago. 2020.

ROSENBERG, N. **Perspectives on technology**. Cambridge: Cambridge University Press, 1976.

ROTH, A. **Como funcionam os mercados**. São Paulo: Portfólio Penguin, 2015.

SCHOT, J., KANGER, L. Deep transitions: emergence, acceleration, stabilization and directionality. **Research Policy**, v. 47, n. 6, p. 1045-1059, 2018.

SRNICEK, N. **Platform capitalism**. Cambridge, UK: Polity Press, 2017.

STURGEON, T. Upgrading strategies for the digital economy. **Global Strategy Journal**, v. 11, n. 1, p 1-24, 2019. DOI: https://doi.org/10.1002/gsj.1364.

TIWANA, A. **Platform ecosystems**: aligning architecture, governance, and strategy. Walthan, USA: Morgan Kauffman, 2014.

UNCTAD. **Digital Economy Report 2019**. Value creation and capture: implications for developing countries. Geneva: United Nations, 2019.

UNCTAD. **World Investment Report 2017**. Investment and the Digital Economy. Geneva: United Nations, 2017.

VAN DIJCK, J.; NIEBORG, D.; POELL, T. Reframing Platform Power. **Internet Policy Review**, v. 8, n. 2, p. 1-18, 2019. DOI: https://doi.org/10.14763/2019.2.1414.

VAN DIJCK, J.; POELL, T.; WAAL, M. de. **Platform society**: public values in a connective world. New York: Oxford University Press, 2018.

VAN DOORN, N.; BADGER, A. Platform capitalism's hidden abode: producing data assets in the gig economy. **Antipode**, v. 52, n. 5, p. 1475-1495, 2020. DOI: https://doi.org/10.1111/anti.12641.

VELDKAMP, L.; CHUNG, C. Data and the aggregate economy. **Journal of Economic Literature**, 2019. (Working paper em preparação).

WAREHAN, J.; FOX, P.; GINER, J. Technology Ecosystem Governance. **Organization Science**, v. 25, n. 4, p. 1195-1215, 2014. DOI: https://doi.org/10.1287/orsc.2014.0895.

WIPO. **WIPO Technology Trends 2019**: Artificial Intelligence. Geneva: World Intellectual Property Organization, 2019.

ZUBOFF, S. **The Age of Surveillance Capitalism**: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs, 2019.

ZYSMAN, J.; KENNEY, M. The next phase in the digital revolution: intelligent tools, platforms, growth, and employment. **Communications of the Association of Computing Machinery**, v. 61, n. 2, p. 54-63, 2018.

SECOND PART: THE PLATFORMIZATION OF SCIENCE

CHAPTER 2 – THE PLATFORMIZATION OF SCIENCE: TOWARDS A SCIENTIFIC DIGITAL PLATFORM TAXONOMY[9]

"Everywhere there can be a platform, there will be a platform"
MIT Platform Strategy Summit 2018

2.1 Introduction

With the World Wide Web (WWW) emergence in the mid and late 1980s, a fragmentation process of traditional academic infrastructures began. This process is twofold: while on the one hand, we can witness the obsolescence of traditional structures (e.g., printed journals); on the other, new online entities – boutique digital libraries, institutional repositories, content management platforms, open protocols, metadata aggregation (PLANTIN *et al*., 2018) – emerge, fighting for their own space. Despite the existence of studies addressing the historical development of digital platforms (ACS *et al*., 2021; LANGLOIS, 2012; HELMOND, 2015; TABARÉS, 2021; HELMOND; NIEBORG; VAN DER VLIST, 2019), none of them has yet drawn a coherent and comprehensive interpretation of the emergence of scientific platforms.

Plantin *et al*. (2018) account for the multiplicity of societal groups that seek to re-integrate some aspects of the fragmented scientific infrastructure caused by the emergence of the web. Mirowski (2018) offered a survey on science-focused private platforms and concluded that "Science 2.0" is a neoliberal project that seeks to profit from phases of the scientific process leading to decreasing autonomy of researchers. He also provided a "landscape of science platforms", crossing phases of scientific activities (getting started; preparatory; research protocols; writeup; publication; and post-publication) with users (normal scientist; funders; competing scientist; spectator scientist; outsider citizens; and, kibitzer), illustrating with some representative platforms. Both Mirowski (2018) and Plantin *et al*. (2018) did not present a chronological narrative or a relational ordering of their historical trajectories. In that regard, Jordan (2019) outlined the historical expansion of science platforms, but only for a specific class: academic social networks (ASN). She

[9] Originally published as: SILVA, V.; CHIARINI, T. The platformization of science: towards a scientific digital platform taxonomy.*Minerva: a Review of science, learning and policy*, 2022.

presented a timeline from 2002 to 2017, in which it is possible to observe the evolution of ASN and general-purpose social networks. Mirowski (2018) also did not separate in his analysis general-purpose social networks (such as Twitter and LinkedIn) from exclusive/dedicated science platforms.

Those previous studies on platforming approach the development of digital platforms process from a relatively short-term perspective, narrowly emphasizing the role of the Market. They neither propose a taxonomy for scientific digital platforms nor provide a definition for scientific digital platforms. Advancing on that ground, we propose a longer-term view (from 1990 onwards), which allows us to identify the participation of other actors in different subsystems in the process of science platformization, with possible impacts on the scientific infrastructure. This way, dialoguing with the most up-to-date literature, we contribute to broadening the understanding of the ongoing process of platformization of research life cycle, answering the following research question: what are the ways in which the scientific phases have been platformed?

We organize the article into five sections, including this introduction. In the next section, we lay out our conceptual background and the method used. We follow Yin's (2003) guidance and structure our research strategy to develop an exploratory study. In section 3, we start our exploratory case study exposing emblematic scientific digital platforms developed within three "great subsystems of modernity": State, Market and Science (NOWOTNY; SCOTT; GIBBONS, 2003). In this context, three large groups of social actors from each one of the subsystems appear crucial to us: public administration, private sector, and, research communities respectively. In section 4, we propose a discussion on our findings and suggest systematizing figures of the exploratory study presented. The evidence throughout the paper unveils important conclusions: (i) the changes (caused by platformization) in each of the phases of the research cycle are not at all linear and are not happening simultaneously; (ii) actors from different subsystem played important roles in the platformization of science; and, (iii) a large number of platforms that become infrastructures and infrastructures that have become platformed, although this does not invariably occur Finally, in the last section, we raise pertinent questions for future inquiries.

2.2 Conceptual background and method

Our exploratory case study – whose goal is "to develop pertinent hypotheses and propositions for further inquiry" (YIN, 2003, p. 17) – departs from four conceptual dimensions, which appear particularly relevant in the context of our study.

The *first is* that the research process "includes the key steps in the research life cycle" (DAI; SHIN; SMITH, 2018, p. 8). The European Comission (2016) proposition of a-five-phase process – i.e., (i) conceptualization; (ii) data-gathering; (iii) analysis; (iv) publication; and, (v) review – was advanced by Dai, Shin, and Smith (2018) who included both "planning" and "funding" activities in the conceptualization phase. They also included "processes of outreach" and "impact assessment" as a sixth phase of the research process.

The *second* dimension recognizes that platformization accelerates. Digital platforms have been penetrating in different economic sectors and spheres of life (POELL; NIEBORG; VAN DIJCK, 2019; VAN DIJCK; POELL; WAAL, 2018). They are defined as organizational models (GAWER, 2021; KENNEY; ZYSMAN, 2020) or (virtual) spaces "where social and economic interactions are mediated online, often by apps" (ACS *et al*., 2021, p. 1635). Digital platforms are "governing systems" (SCHWARZ, 2017) that exhibit regulatory power over the markets/social arenas they mediate.

Our *third* conceptual dimension states that platformization is a multiagent phenomenon. The subsystems that lead the platformization processes are diverse, not restricted to the Market (MANSELL; STEINMUELLER, 2020). Corporations have developed and marketed digital platforms. These proprietary platforms exhibit institutional logics that combine previously separate logics (such as the logic of markets and the logic of the regulatory state) (FRENKEN; FUENFSCHILLING, 2020). Governments and their agencies have developed platforms for the delivery of public services (MUKHOPADHYAY; BOUWMAN; JAISWAL, 2019; THOMPSON; VENTERS, 2021). International associations (OTTO; JARKE, 2019) and cooperatives also have digital platforms, which exemplifies how different actors can develop and operate digital platforms.

Finally, the *fourth* conceptual dimension highlights that digital platforms and socio-economic infrastructures are merging. Once platforms achieve properties such as ubiquity, invisibility and essentiality, they generate what has been called "infrastructuralization of platforms" (PLANTIN *et al*., 2018). *Pari passu*, motivated by the control of data flows, intermediary platforms seek to expand upstream and downstream, occupying spaces of often public infrastructure platforms, causing what can be called the "platformization of infrastructures" (PLANTIN *et al*,. 2018; VAN DIJCK, 2020).

Considering the conceptual background above, the phases of the scientific process may go through platforming processes led by different social actors, with possible impacts on the scientific infrastructure. From the above, follows our research question: what are the ways in which the scientific phases have been "platformed"? To guide our investigation, we have broken down the main question into subsidiary questions: what is a scientific platform? What are the main types of scientific digital platforms? What are the degrees of integration of the main types of scientific digital platforms into the research cycle? To answer the previous questions, we first define scientific digital platforms as the digital governing systems of virtual spaces that leverage network effects towards one (or more) phase(s) of the scientific research process[10]. However, there are differences among this large population set and we propose a taxonomy of scientific platforms summarized in Table 2.1. As per Hodgson (2019, p. 4), taxonomic definitions "concern populations of social phenomena that exhibit some degree of commonality and some degree of diversity" and they "identify the minimum number of properties that are sufficient to demarcate one group of entities from all other entities".

We characterize each scientific platform category resorting to emblematic scientific platforms (Table 2.1). Beforehand it is necessary to make a caveat: we do not aim to make an inventory of all scientific digital platforms developed from 1990 to 2020; despite the irruption (and "death") of many scientific digital platforms in this 30-year life span, making a census survey of them all is out of the scope of this paper. Instead, according to specific criteria we select digital platforms that reveal typifying features. The cases were selected through a purposeful sampling guided by four criteria: (i) historical relevance, (ii) emergence within the WWW context from 1990 on; (iii) availability of public information for analysis; and (iv) diversity of platform groups in relation to specific social subsystems. The next section discusses each of the types in detail.

It is worth noting how each of the groups of scientific platforms is associated with one of the three "great subsystems of modernity" (NOWOTNY; SCOTT; GIBBONS, 2003): State, Market and Science. The associations are not exclusive, but they reflect historical relevance. For example, there are archives developed by Market and State actors, but the role

[10] We thus disregard general purposed platforms that may also be used at some stage of the scientific process, *e.g.*, LinkedIn, Twitter, Amazon Mechanical Turk.

of the scientific community (Science subsystem) was and continues to be the most relevant in the development and operation of this group of scientific platforms.

Table 2.1 – Taxonomy of scientific platform' groups and representative science platforms

Great subsystems of modernity	Scientific Platform group	Emblematic Scientific Platforms	Platform sides
Science	Archives and repositories	ArXiv	Authors and readers of articles and preprints
	Citizen Science Platform	Zooniverse	Project managers and volunteers
State	National e-portfolios	Lattes Platform	Researchers and research institutions
	Grid	TeraGrid/XSEDE	Holders and demanders of idle computing capacity
	Science gateways	HUBzero	Application developers and users
Market	Publishers	Elsevier' Journals	Article authors and readers
	Academic Social Networks (ASN)	Academia.edu; ResearchGate; Mendely	Researchers; supply and demand for job positions
	Crowdwork	Prolific Academic	Survey researchers and participants

Source: authors' own

As an exploratory study, we resorted to documents from companies such as corporation brochures; public platform documents, such as government reports, white papers and awards; news and letters published in specialized journals for the communication of the scientific community (*e.g.*, Nature). For comprehensibility reasons, Annex 1 presents a complete list of documents accessed.

2.3 The platformization of research phases: an exploratory analysis

Two steps were essential to enable the emergence of digital platforms: the creation of the physical structure of the network of networks, also known as the Internet; and the definition of a standard for navigation and guidance on the internet via common protocols, the WWW (MANSELL; STEINMUELLER, 2020). In its origin, before all commercial restrictions on its use were lifted in 1995, the Internet was used mainly by academic communities (VICKERY, 1999). Notwithstanding that, other actors, including funding bodies and profit-seeking companies, were also responsible for developing new technological trajectories. Our analysis revolves around three social subsystems for analytical simplicity:

Science (*e.g.*, research community), State (*e.g.*, public administration), and Market (*e.g.*, private sector).

2.3.1 Science subsystem

2.3.1.1 Archives and repositories: quick and dirty access

When Tim Berners-Lee developed the WWW in the early 1990s, within the *Conseil Européen pour la Recherche Nucléaire* (CERN), the aim was to facilitate scientific communication and the dissemination of scientific research (TIM BERNERS-LEE; FISCHETTI, 2000). It quickly disseminated and allowed the upsurge of new artifacts. Therefore, it has dramatically reshaped the way scientific production could be stored and shared, allowing research communities to create their web platforms. That was the case of Paul Ginsparg, who wanted to develop an online academic repository. He envisioned "an expedient hack, a quick-and-dirty email-transponder written in csh to provide short-term access to electronic versions of preprints" (GINSPARG, 2016, p. 2620) and then came up with ArXiv[11] in 1991. Ginsparg's "expedient hack" became, in 2020, a central platform for scientific dissemination of results, registering 4.2 million active users, storing 1.8 million articles, and managing 1.89 billion downloads (ARXIV, 2020).

ArXiv's success is due to a set of characteristics. Perhaps the main one is the balance between the speed of dissemination and the maintenance of the quality of the material admitted to the platform, which are essential shared values for the scientific communities. The voluntary curatorship of scientists who assess whether the submitted material reaches a minimum degree of validity and interest to the community achieves this balance. Given the information overload (ArXiv received 16,000 articles per month in 2020), the platform complements the curatorial work with machine learning systems that warn of possible items that deserve special attention (GINSPARG, 2016, p. 2623).

In addition to the balance mentioned, the platform's governance also stands out: it is maintained by contributions from user institutions, and Cornell University Library conducts its maintenance and operation. Important decisions are made in committees that involve both the maintainers and the user community, following the participatory ethos of the community that developed and used the platform (GINSPARG, 2016). ArXiv has always valued openness and transparency and provides application program interfaces (API) so that third parties can

[11] https://arxiv.org, accessed on 04/20/2022.

use their data responsibly. In addition to the open data policy, the platform promotes ArXivLabs, a window through which the community proposes innovations as additional features for the platform.

Since its inception, the platform has paved the way for innovations attributed to the users themselves, such as decentralized data collection. This pioneering type of crowdsourcing "foreshadowed the 'interactive web', insofar as it provided a rudimentary framework for users to deposit content" (GINSPARG, 2011, p. 146).

ArXiv was disruptive, and it became the model for many followers. For instance, in 1993, NetEc – a website to improve the communication of research in Economics – was created, and in 1997, it became RePEc[12], a decentralized repository of scientific articles in that field. Another example is the "e-biomed" initiative, which looked for ways to update the dissemination of results in life sciences. The results of "e-biomed" are now PubMedCentral and PLoS, essential centers of open science (GINSPARG, 2016). Other initiatives were flourishing worldwide, as the development of the network for geological and environment data (Pangea) in 1993 in Germany, conceived to be a data archive and a scientific tool (DIEPENBROEK *et al*., 2002).

During the 1990s, the scientific community had used the WWW to bypass traditional means of publication, using platforms such as ArXiv. In the first decade of the 2000s, part of the scientific community pressured the private institutions that dominated the cycle of publication and distribution of scientific literature: the publishers. They did so by promoting open science, defined as "a scientific culture that is characterized by its openness. Scientists share results almost immediately and with a vast audience" (BARTLING; FRIESIKE, 2014, p. 10). Traditional journals' practice of charging for access to scientific results became a natural enemy of open science.

The 2000s testified the growth of this movement[13]. In 2002 came the Budapest Open Access Initiative (BOAI), and the declarations of Berlin (2003) and Bethesda (2003) soon followed. According to Suber (2012), this "BBB [Budapest, Berlin and Bethesda] definition of open access" was devoted to removing price tags and permission barriers (copyrights) from peer-reviewed research. These and other initiatives focused on the wide dissemination of scientific results constitute what Fecher and Friesike (2014) call the "democratic school" of open science. This stream within open science is concerned with

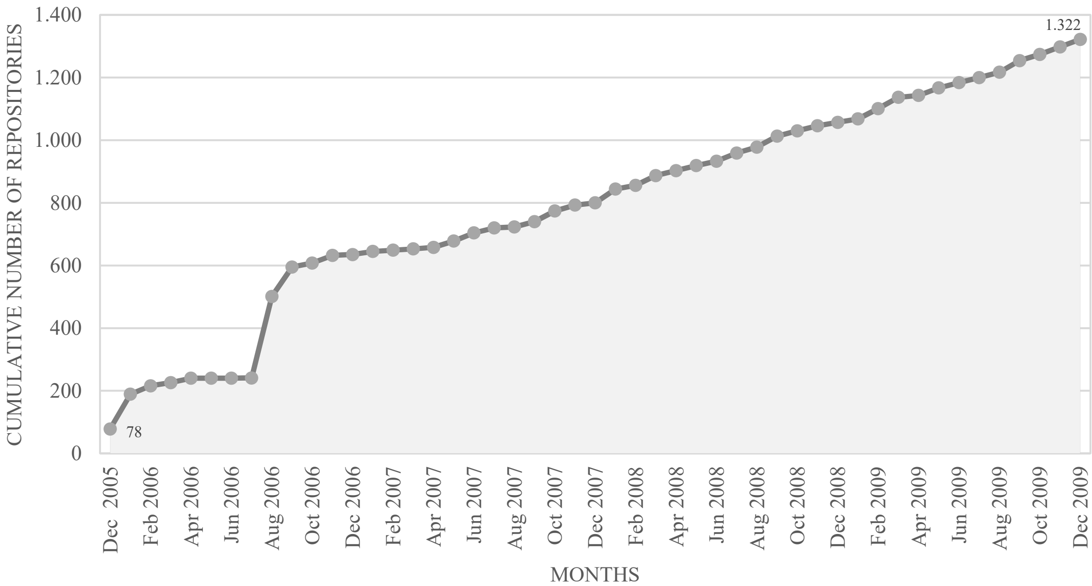
[12] http://repec.org, accessed on 04/20/2022.

[13] Suber's (2012) extensive notes allow us to observe how open science initiatives existed before the 2000s. At the same time, his records demonstrate how the movement took off from this decade on. View the timeline at https://dash.harvard.edu/bitstream/handle/1/4724185/suber_timeline.htm, accessed on 04/20/2022.

access to knowledge and advocates that "any research product should be freely available […] everyone should have the equal right to access knowledge, especially when it is state-funded." (FECHER; FRIESIKE, 2014, p. 25).

The technical possibility of opening science through the web 2.0 is responsible for the growing momentum of open science in the 2000s[14]. The BOAI highlighted how the Internet would allow the free distribution of scientific information. Bartling and Friesike (2014) find that any open science approach is closely linked to the technological developments of the Internet. These developments make up a system of intertwined components (*e.g.*, unique researchers' IDs, platforms, altmetrics, social networking) that make open science viable when established. From the global Directory of Open Access Repositories (OpenDOAR), founded in 2005, it is possible to see the increase from just 78 directories/archives in December 2005 to 1,322 in December 2009 (Figure 2.1).

Figure 2.1 – Growth of open access repositories, Dec/2005–Dec/2009



Source: Authors' own. Data sourced from OpenDOAR

On the agenda, the democratic open science school promotes both open data and open access. Open data has a more researcher-centered connotation, as it intends to enable reproducibility, leverage the confidence of scientific results, and optimize research efforts. On

[14] An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the worldwide electronic distribution of the peer-reviewed journal literature and unrestricted access to it by all scientists, scholars, teachers, students, and other curious minds.

the other hand, open access is less researcher-centric and results in the view of scientific knowledge as a universal right (FECHER; FRIESIKE, 2014). The 2000–2010 decade saw a multiplication of digital platforms to enable open data and open access.

Archives and repositories' platforms connect knowledge producers with knowledge consumers. Like mundane sharing economy platforms, they thrive on network effects: platforms with great articles will be more attractive to the knowledge seeker. However, due to their non-commercial nature, these repositories often follow the guidance of the Open Archives Initiative (OAI) Protocol for Metadata Harvesting (PMH), which makes them interoperable. By including this protocol in the architecture and governance of the platform, repositories allow users to use a single search tool to find the information they are looking for (SUBER, 2012, p. 56). This is an essential difference between digital platforms emerging from the research community and commercial digital platforms: while the formers seek to reduce the balkanization of knowledge and reduce search frictions, the latter only seek to do this within their closed gardens, where monetization strategies can be applied. It follows from there that, in a commercial-dominated platform universe, reducing the fragmentation of scientific knowledge can only occur if there is only one dominant platform, *i.e.*, a monopoly. Interoperability is antagonistic to monetization.

2.3.1.2 Citizen science platforms: extending the republic of science

Digitally enabled citizen science is defined as "the emerging practice of using digital technologies to crowdsource information about natural phenomena." (WYNN, 2017, p. 2). This definition implicitly emphasizes the productivity-enhancing role of scientific activity based on the engagement of ordinary citizens. However, according to Sauermann *et al.* (2020), the "productivity view" is only half the history of citizen science. Other authors emphasize the "democratization view" of citizen empowerment through participation in generating scientific knowledge. This second view, particularly, leads Fecher and Friesike (2014) to place citizen science within the larger trend of the open science's *public school*, in which "Web 2.0 technologies allow scientists, on the one hand, to open up the research process and, on the other, to prepare the product of their research for interested non-experts".

Citizen science recent developments are due to the new technical possibilities of web 2.0, the use of smartphones, and the diffusion of digital platforms (KULLENBERG; KASPEROWSKI, 2016; LEMMENS *et al.*, 2021; SAUERMANN *et al.*, 2020). Kullenberg and Kasperowski (2016) emphasize that, although the voluntary participation of lay people in science is not new, it is no longer invisible due to the digital records afforded by digital platforms. They identify a growing trend in the number of scientific articles on citizen science around 2010, which coincides with an increase in the number of citizen science projects[15] hosted by dedicated digital platforms. Scientific articles that resulted from projects hosted on digital platforms also grow in number: "citizens can be involved in new instances of the scientific process, and in much larger numbers due to the logistical affordances of digital platforms." (KULLENBERG; KASPEROWSKI, 2016, p. 13). Citizen science digital platforms exhibit five characteristics according to Liu *et al*. (2021, p. 440):

> i. present active citizen science projects and activities;
> ii. display citizen science data and information;
> iii. provide overall guidelines and tools that can be used to support citizen science projects and activities in general;
> iv. present good practice examples and lessons learned; and,
> v. offer relevant scientific outcomes for people who are involved or interested in citizen science.

The first digital platform for citizen science (and the biggest one, according to SIMPSON; PAGE; DE ROURE, 2014 and LIU *et al*., 2021) is Zooniverse (formerly Galaxy Zoo), launched in 2007 Kullenberg and Kasperowski (2016). Zooniverse, hosted and managed by the University of Oxford, the Adler Planetarium, and the University of Minnesota, evolved from Galaxy Zoo, a project to crowdsource the classification of galaxies images. At the time of writing, the platform boasts[16] six hundred twenty-two million classifications, filled by approximately 2.3 million volunteers. The projects include space to discuss and socialize, and some volunteers are granted specific roles and privileges (TINATI *et al*., 2015) whose purpose is to encourage participants further to fulfill their tasks (KRAUT; RESNICK, 2011).

Zooniverse allows scientists and citizen scientists to propose projects to the crowd. Then it displays projects and lets volunteers engage in the classification efforts. It also connects volunteers with project managers. In a sense, it matches science project needs with crowd volunteers. Among the results, Zooniverse projects already classified "more than a million galaxies, the discovery of nearly a hundred exoplanet candidates, the recovery of lost fragments of ancient poetry, and the classification of more than 18,000 thousand wildebeest in

[15] Citizen science projects vary in terms of the depth of citizen participation. While all projects involve data collection and recording, not all allow laypersons to analyze results or set goals. Sauermann *et al*. (2020) propose four categories: from more restricted participation to data collection (collaborative projects) to more decisive participation, in which there are no professional scientists in the project (autonomous projects).

[16] https://www.zooniverse.org, accessed on 04/20/2022.

images from motion sensitive cameras in the Serengeti" (SIMPSON; PAGE; DE ROURE, 2014, p. 1049).

Citizen science platforms are thus important ways to democratize science and publicly promote scientific data and information and facilitate knowledge transfer to wider audiences (WAGENKNECHT *et al*., 2021), which are typical values of open scientific communities. Moreover, they may "facilitate mutual learning and multi-stakeholder collaboration, get inspiration, integrate existing citizen science activities, develop new citizen science initiatives and standards, and create social impact in science and society" (LIU *et al*., 2021, p. 441).

2.3.2 State subsystem

2.3.2.1 National e-portfolios: ordering the ivory tower

Government platformization is directly linked to the development of systems for scientific and technical information. This field of information science which intersects with science and technology (S&T) policy, has evolved with UNESCO's post-war efforts to create an International Scientific Information System (called UNISIST) (UNESCO, 1971). It was a worldwide bibliographic system that, for various reasons, has not materialized (COBLANS, 1970). Notwithstanding that failure, UNESCO successfully promoted scientific and technical information systems worldwide (SARACEVID, 1980). In the 1970s, for example, several developing countries created their own S&T information offices, presuming a positive relationship between S&T with economic development (AVGEROU, 1993).

While most developing countries designed national or sectorial scientific and technical information systems, developed countries had organic specific-subject ones (e.g., chemistry systems, physics systems). Thus, different actors were involved in the process: in developing countries, the State was the main responsible for developing and operating the systems; in developed countries, they were run mainly run by research communities (SARACEVID, 1980, p. 224).

Thus, in the 1990s, governments in developing (and in a few developed) countries took advantage of new ICT-based technologies to improve their scientific and technical information systems. The euphoria was such that "[in many less developed countries], 'Computer' seems to be in many writings a magic word, the expectations somewhat naive and unrealistic" (SARACEVID, 1980, p. 241). State platforms were generally e-portfolios that standardized researchers' outputs. By centralizing and standardizing the scientific production of researchers, e-portfolios systematized the inputs that informed funding allocation decisions. Some have also become, both de facto and de jure, the standard for disseminating scientific production. According to Saracevid (1980), scientific and technical information systems had different users: scientists and engineers; business managers and industrialists; administrators and policymakers; extension workers; semi-educated persons and illiterates.

Lattes Platform[17] is an example resulting from scientific and technical information system developments in Brazil. It was developed in a fruitful collaborative network formed by government agencies, public universities, and the private sector. The main goal with Lattes Platform was the standardization of Brazilian researchers' curricula and the provision of unique research IDs (before ORCID) This standardization aimed at constructing a database, making it possible to find specialists and provide statistics on the distribution of scientific research countrywide (LANE, 2010). Since it was launched, Lattes Platform has been increasing its scope and its base of users. Brazilian government implemented features to spur users' connectivity based on the stimuli to join all sub platforms. For instance, having updated information in Lattes Curricula is a precondition for accessing public funding and scientific research. Lattes Curricula homepage provides an outlet that allows users to monitor other scholars' activities (main researches, publications, filiation) (CHIARINI; SILVA, 2022). Network effects are present in the Lattes Platform. There are one-sided effects: the more researchers use the platform, the greater the value perceived by an individual researcher of participating in the network. There are also cross-network effects: the more researchers make their information available on the Lattes platform, the more institutions require it as a document proving scientific production. The more institutions that use the Lattes curriculum, the greater the incentive for a researcher to update theirs (CHIARINI; SILVA, 2022).

2.3.2.2 Cyberinfrastructure networks: power to the grid

In the early 2000s, developed countries invested in large digital infrastructure projects for science. In Europe, the U.K. took the lead: John Taylor, then director-general of Research Councils at the Office of Science and Technology (OST), led a five-year, £ 250 million e-Science program. In the U.S., the Atkins Report (2003) became the landmark of the cyber-infrastructure promoted by the U.S. government through funding from the National

[17] https://lattes.cnpq.br/, accessed on 05/03/2022.

Science Foundation (NSF), leading to the establishment of the Office of Advanced Cyberinfrastructure (OAC)[18] (JANKOWSKI, 2007, p. 551). As OAC's name reveals, this initiative (as well as its British counterpart) was focused on infrastructure for science in the digital era, as Branscomb (1992) had advertised a decade before: investment in public goods as infrastructure would be an S&T policy aimed at "enabling innovation" instead of "picking winners".

The British program supported "the e-Science pilot projects of the different Research Councils and work[ed] with industry in developing robust, "industrial strength" generic Grid middleware." (HEY; TREFETHEN, 2002, p. 1019). This short quotation reveals that the program was (i) infrastructure-focused, (ii) experimental, (iii) interdisciplinarity (iv) interactive with the private sector.

Grid computing had its name inspired by the traditional infrastructure of electric energy: computational services offered on-demand, as well as electric energy. To achieve this vision of e-utility, it was necessary to put in place the Grid. In its genesis, there was the idea of using idle cycles of distributed computing, *i.e.*, computers with processing power would connect to a network and operate on demand[19]. Matching users with idle resources seem to be a direct ancestor of what became the sharing economy platform's principles (FRENKEN; SCHOR, 2017). However, e-Science scope was much broader than the pooling of resources via aggregation systems, encompassing: "international collaboration among researchers; increasing use of high-speed interconnected computers, applying Grid architecture; visualization of data; development of Internet-based tools and procedures; construction of virtual organizational structures for conducting research; electronic distribution and publication of findings." (JANKOWSKI, 2007, p. 552). Platforms were among the middleware layer, but the main goal was more ambitious: "Our emphasis is thus on Grid middleware that enables dynamic interoperability and virtualization of IT systems, rather than Grid middleware to connect high-performance computing systems, to exploit idle computing cycles or to do 'big science' applications" (HEY; TREFETHEN, 2003). As we will see, this goal would have to wait for the next generation of State science platforms.

It is worth noting a representative case briefly. The NSF launched a call in 2000 for what would be the seed of the TeraGrid project. The USD 51 million awards directed to the MPC Corporation aimed to develop the Terascale Computing System: "The Pittsburgh Supercomputer Center (PSC), a jointly supported venture of The University of Pittsburgh and

[18] https://www.nsf.gov/div/index.jsp?div=OAC, accessed on 04/20/2022.
[19] Such as in the SETI@home project (HEY; TREFETHEN, 2003).

Carnegie Mellon University acting through the MPC Corporation, in collaboration with Westinghouse Electric Corporation, will put in place at 6 peak teraflop computing system for use by U.S. researchers in all science and engineering disciplines”[20]. In 2001, the project gained scale: approximately USD 80 million were allocated to it, through two awards, to develop what was called TeraGrid: centered at the University of Illinois, in partnership with four more national research centers and with IBM, Intel, Myrinet, Qwest, Oracle and SUN, the NSF expected to see a distributed terascale facility “with an aggregate of 11.6 TF of computing capability”[21] at the end of the project. The NSF granted seven more awards until 2005, in a new contribution of USD 80 million. In 2011, the project was replaced by XSEDE: eXtreme Science and Engineering Discovery Environment. XSEDE’s first award[22], worth USD 121 million, provided for the development and operation of the grid between 2011 and 2016. The second[23], at the cost of another USD 110 million, financed the structure for another five years, until 2021. This case demonstrates the government’s leadership, represented by the NSF, in setting the goal and funding via awards to develop a broad grid computing infrastructure.

The focus on virtualization of IT systems directed the e-Science project to the utility model or servitization of computing. Servitization was also a central goal of the companies usually involved in funded projects: IBM, Microsoft, Sun (HEY; TREFETHEN, 2002). It was not by chance that the leader of the e-Science project, Tony Hey, was hired by Microsoft Research after leaving command and that the head of the U.K. OST John Taylor had worked at Hewlett-Packard. This “revolving door” movement facilitated cooperation between the government and the private sector. The millions of pounds and dollars invested on e-Science and Cyberinfrastructure projects acted as “public procurement,” in which private companies responsible for the development of grid systems evolved into servitization years later (PaaS, SaaS, IaaS). It is not surprising, therefore, that “cloud computing can be seen as a natural next step from the grid or utility model.” (ALLAN, 2009, p. 135). Protected from the risk by public money, Big Techs had “intensive training” on providing IT infrastructure during the 2000s.

---

[20] https://www.nsf.gov/awardsearch/showAward?AWD_ID=0085206, accessed on 04/20/2022.
[21] https://www.nsf.gov/awardsearch/showAward?AWD_ID=0122296, accessed on 04/20/2022.
[22] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1053575, accessed on 04/20/2022.
[23] https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562, accessed on 04/20/2022.

2.3.2.3 Science gateways: the science app store

As e-science projects advanced, it became more apparent that grid middleware alone would not be enough to allow scientific research to enter the digital era. There was too much focus on middleware, and there was a lack of focus on end applications for users (scientists, researchers, scholars) who could not always customize digital tools[24]. Around 2010, virtual research environments (VRE) projects began to multiply to fill this gap, since they could "lower barriers by hiding the complexity of the underlying digital research infrastructure and simplifying access to best-practice tools, data and resources, thereby democratizing their usage" (BARKER *et al*., 2019, p. 243). They were called science gateways in the USA.

Science gateways are "digital platforms that facilitate the use of complex research and computing resources" (PARSONS *et al*., 2020, p. 491). They focus "on end-to-end solutions for a single domain" (MADDURI *et al*., 2015, p. 2), so even if there are multidisciplinary gateways, the majority are discipline-based. Specifying a gateway concerning the needs of a specific research community (*e.g*., GHub[25], the science gateway for Glaciology) "also differentiates between science gateways and the generic cyberinfrastructure on which they build" (BARKER *et al*., 2019, p. 241). As of this writing, the Science Gateway Catalog[26] of the Science Gateway Community Institute (SGCI) (funded by the National Science Foundation) registers 531 gateways.

Over the past decade, many governments have created programs to fund the development of science gateways. Allan (2009) presents a list of projects that, in the late 2000s, moved from e-science to science gateways and VRE. Barker *et al*. (2019) mention CANARIE, Canada's government-funded initiative; SGCI, founded in 2016 by NSF with initial funding of US$15 million; SCI-BUS, financed by the Horizon 2020 program of the European Union; and National eResearch Collaboration Tools and Resources (Nectar), funded by the Australian Government (2011–2017).

---

[24] One clue to the need to focus on usability is found in the award for the second iteration of XSEDE, the grid computing network mentioned in the previous session: XSEDE 2 will also respond to the evolving needs and opportunities of science and technology. [...] The project will continue to innovate the use of "e-science portals" (also known as Science Gateways). Science gateways provide interfaces and services that are customized to a domain science and have an increasing role with facilities and research centers, collaborating on large research undertakings (e.g., Advanced LIGO, Polar Geospatial Center). This approach facilitates broad community access to advanced computers and data resources. Science gateways are now serving more than 50% of the user community." (https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562, accessed on 04/20/2022).

[25] https://vhub.org/groups/ghub, accessed on 04/20/2022.

[26] https://catalog.sciencegateways.org/#/home, accessed on 04/20/2022.

Although we relate science gateways with the State subsystem, it is essential to emphasize that the scientific community has a central role in the subsequent development of these platforms: in the design of the platform, providing input on its idiosyncratic needs; in the continuous construction of new platform features when it is open source and therefore extensible (ALLAN, 2009, p. 12); in the maintenance of the platform itself, which demands new jobs and career-paths for facilitators, research software engineers, and science gateway creators (PARSONS *et al*., 2020), generation of content that is shared by it, among other functions. It is critical to highlight the government projects' initiative, funding, and institutionality given to science gateways.

The contribution of the NSF was fundamental for the development of HUBzero, "an open-source software platform for building powerful websites that host analytical tools, publish data, share resources, collaborate and build communities in a single web-based ecosystem"[27]. In other words, HUBzero is a customizable platform for developing other platforms. Its intention to facilitate "Gateway as a Service" was responsible for bringing the digital scientific workflow to the masses (MCLENNAN *et al*., 2015). Funded by the NSF to create the Nanotechnology Science Gateway (NanoHUB), the project successfully created a customizable framework that houses more than 25 other active science gateways. The science gateways built on the HUBzero framework register a flow of more than 2 million users – serving over 280 thousand users annually distributed over 172 countries (MADHAVAN; ZENTNER; KLIMECK, 2013) – and include features ranging from the conceptualization of research projects to the publication of data sets, in addition to, of course, accessing high-performance computational resources on the grid. In addition to being free to use, the platform is open and allows for community additions, with codes made available via GitHub.

2.3.3 Market subsystem

2.3.3.1 Publishers: Knowledge oligopolies?

Academic publishers own the most traditional scientific platform. Controllers and disseminators of scientific journals were established about 350 years ago as bridges between writers and readers. In the late 1990s, they moved into a dual process of market consolidation and digitization. Larivière, Haustein, and Mongeon (2015) demonstrated that from 2000

---

[27] https://hubzero.org/about, accessed on 04/20/2022.

onwards, journal acquisitions increased, and five commercial houses – Reed-Elsevier, Springer, Wiley-Blackwell, SAGE Publications, and Taylor & Francis – became an oligopoly (LARIVIÈRE; HAUSTEIN; MONGEON, 2015).

In some scientific disciplines, such as psychology, the Big 5 controls more than 70% of the journals present in WoS. Their profit margin exceeded that of industries such as pharmaceuticals, and operating profits grew nearly sixfold (for Scientific, Technical & Medical division of Reed-Elsevier) between 1991-2013 (LARIVIÈRE; HAUSTEIN; MONGEON, 2015). It should be noted how, in the 2000s, publishers integrated other platforms with journals. Hence the launch of Scopus by Elsevier in 2004, competing with the WoS index. Today, Reed-Elsevier.

The advancement of the digital age has led publishers to reposition themselves. Elsevier's corporate brochure[28] states that "while the proliferation of information brings opportunity, it also brings challenges." For this reason, the company calls itself "a global leader in information and analytics." This broadening of scope relates to the expansion of scientific knowledge flows in the digital age. As pointed out by Delfanti (2021, p. 8), "the boundaries between 'gray' and 'formal' scholarly objects blur" due to "the technological affordances of digital media, which can publish any number of objects and are not limited by the constraints that limit journals, with their periodical schedules and cumbersome peer review processes."

Since academic publishers could not contain the leakage of scientific knowledge flows, the solution for their business model is to control as many platforms (and as many scientific processes phases) as possible that position themselves as curators and distributors of these flows:

> In 2016, the owner of Web of Science spun off that unit to purchase by a private equity firm, where it was renamed 'Clarivate Analytics.' Then, in 2017, Clarivate bought Publons, with the justification that it would now be able to sell science funders and publishers 'new ways of locating peer reviewers, finding, screening and contacting them' […] Elsevier first purchased Mendeley (a Facebook-style sharing platform) in 2016, then followed that by swallowing the Social Science Research Network, a pre-print service with strong representation in the social sciences (Pike, 2016). In 2017 it purchased Berkeley Economic Press, as well as Hivebench and Pure (MIROWSKI, 2018, p. 197).

That occurred without the decentralization of the traditional peer-reviewed journals market. In 2020, "Elsevier's article output accounts for about 18% of global research

[28] https://www.elsevier.com/__data/assets/pdf_file/0010/1143001/Elsevier-corporate-brochure-2021.pdf, accessed on 04/20/2022.

output while garnering approximately 27% share of citations"[29]. The expansion to new digital platforms reflects the emergence of a new competitor in scientific platforms: academic social networks.

2.3.3.2 Academic Social Networks: you have a new follower!

Despite HASTAC[30] being considered the world's first ASN developed within the research community, many for-profit venture capital-funded technology startup companies created virtual *loci* for personal communication and information change, creating specific online communities for researchers (JORDAN, 2019). Those for-profit platforms started to pop up in the late 2000s – following the footsteps of conventional and generic social networks such as MySpace, Facebook, and LinkedIn – and their growth from 2010 started to call attention to their benefits but also about how they could disrupt the research landscape by capturing public content (VAN NOORDEN, 2014).

Table 2.2 – Main information about selected ASN

	Academia.edu	**ResearchGate**	**Mendeley**
General details			
Founded year	2007	2008	2008
Location	San Francisco	Berlin	London
Legal name	Academia Inc.	ResearchGate GmbH	Mendeley Ltd.
Company type	For-profit	For-profit	For-profit
Number of employees	101-250	251-500	51-100
Web monthly visits (million)*	64.31	165.89	6.68
Funding round details			
Number of funding rounds	6	4	2
Total funding amount (USD million)	33.80	87.60	2.10
Funding raise period	2007-2019	2010-2017	2009-2012
Number of investors	11	22	6
Technological details			
Number of patents at USPTO	1	19	0
Number of active technologies used for the website**	13	44	73

Source: authors' own. Data sourced from Crunchbase and USPTO. Note: (*) Total (non-unique) visits to site for April/2021; includes desktop and mobile web. (**) These include Viewport Meta, IPhone/Mobile Compatible, and SPF.

The three giant ASN platforms are today Academia.edu, ResearchGate, and Mendeley (acquired by Elsevier in 2013) (Table 2.2). However, their historical development is not similar, and it is possible to divide them into two main categories. First, there are

[29] Ibidem.
[30] https://www.hastac.org/, accessed on 04/20/2022.

platforms whose aim was to facilitate profile creation and connection (as Academia.edu and ResearchGate); secondly, there are those whose primary aim was to store and share academic-related content, which later combined social network capabilities (as Mendeley) (JORDAN, 2019). Despite the two categories, ASN positions itself in competition with academic publishers rather than social media. (JORDAN, 2019).

As of ResearchGate, for example, it was built referencing prior technologies such as Google's system and method for searching and recommending objects from a categorically organized information repository (PITKOW; SCHUETZE, 2006) and LinkedIn's method and system for reputation evaluation of online users in a social networking scheme (WORK; BLUE; HOFFMAN, 2005). It was then in its original idea the development of a system for sharing academic content for academic users (HOFMAYER *et al*., 2015).

Jordan (2019) provides a comprehensive review of the empirical literature related to ASNs, and throughout assessing over 60 publications, she concluded that despite open access dissemination of academic publications being the most used benefit related to the role of ASNs, they also benefit in terms of speed in comparison to academic repositories and enhance reach and citations. Notwithstanding that ASNs are more than publishing and repository digital platforms, they also allow scholars to interact virtually – but it is also true that interaction is undertaken by a minority of users so far. As for all sorts of technologies, not only does ASN provide benefits, but it may also open unprecedented risks for the structure and evolution of science.

The rapidly growing academic inputs into ASNs generate scholarly big data that can be used by research communities to understand scientific development and academic interactions and for policymakers to solve resource allocation issues (KONG *et al*., 2019). However, scholarly big data gathered by ASNs are kept privately. ASN are also expanding to other scientific phases; for instance, Academia.edu has developed its "Academia Letters," which "aim to rapidly publish short-form articles such as brief reports, case studies, 'orphaned' findings, and ideas dropped from previously-published work."[31] Publishers have been threatening to remove millions of papers from ASN (VAN NOORDEN, 2017) and have sued ResearchGate over copyright infringement (CHAWLA, 2017; ELSE, 2018) which had to remove thousands of publications from its website: the dispute is far from being solved.

[31] https://www.academia.edu/letters/about, accessed on 04/20/2022.

2.3.3.3 Crowdwork: over the shoulders of crowds

Outsourcing work is not a new trend; however, digital platforms have opened up new horizons. "Crowdwork functions as a marketplace for the mediation of both physical as well as digital services and tasks" (HOWCROFT; BERGVALL-KÅREBORN, 2019, p. 23). This marketplace has five main constitutive elements: the digital platform, the labor pool, employment contracts, algorithmic control and, digital trust. Among several types, online task crowdwork platforms (HOWCROFT; BERGVALL-KÅREBORN, 2019) are the most active in the context of scientific research phases.

The online task crowdwork platforms "offers paid work (sometimes subject to requester satisfaction) for specified tasks and the initiating actor is the requester. The tasks are modular, ranging from microtasks to more complex projects, with the potential for further Taylorisation" (HOWCROFT; BERGVALL-KÅREBORN, 2019, p. 26). Researchers have used these platforms to carry out surveys quickly and cheaply, when compared to traditional ways. The main platform[32] dedicated to this type of mediation between researchers and respondents is Prolific Academic (PoA).

PoA was founded in 2014 and received its first round of venture capital (USD 1.2 million) in 2019. The platform matches over 130,000 respondents with 25,000 researchers from over 3,000 institutions around the world[33]. Researchers can select specific profiles of respondents: "Our participant pool is profiled, high quality and fast. The average study is completed in under 2 hours. Filter participants using 250+ screeners (*e.g.*, sex, age, nationality, first language), create demographic custom screeners, or generate a UK/US representative sample."[34]

The increased use of this solution in behavioral research has led to questions about the quality of these data. Eyal *et al*. (2021) conducted a study comparing the quality of data on three crowdwork platforms (Amazon Mechanical Turk – AMT, CloudResearch and PoA) and two panels (Qualtrics and Dynata) used in behavioral research. They tested the quality of the data in terms of respondent's attention, comprehension, reliability, and dishonesty. Their results point to PoA as a platform with better data quality and AMT as a platform with worse data quality. Although the authors see an improvement in data mediation

---

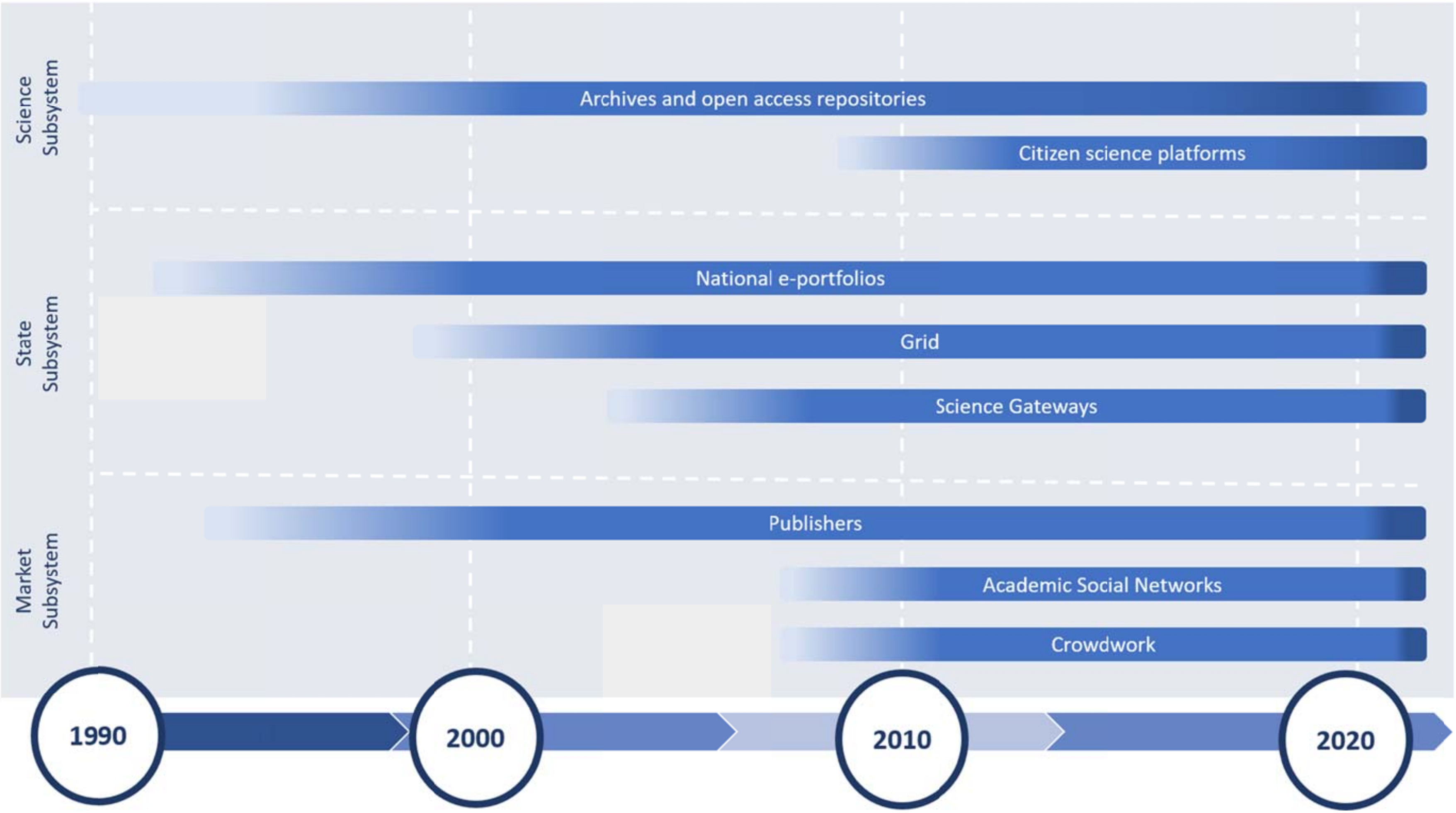
[32] Amazon Mechanical Turk is used extensively for the same purpose, but it is not dedicated to the mediation of a scientific activity (General-purpose crowd work platform, outside the scope of analysis of this article)
[33] https://www.prolific.co/about, accessed on 05/03/2022.
[34] Accessed on 05/03/2022.

patterns, they point out that “data quality remains a concern that researchers must deal with before deciding where to conduct their online research” (EYAL *et al*., 2021).

2.4 Discussion

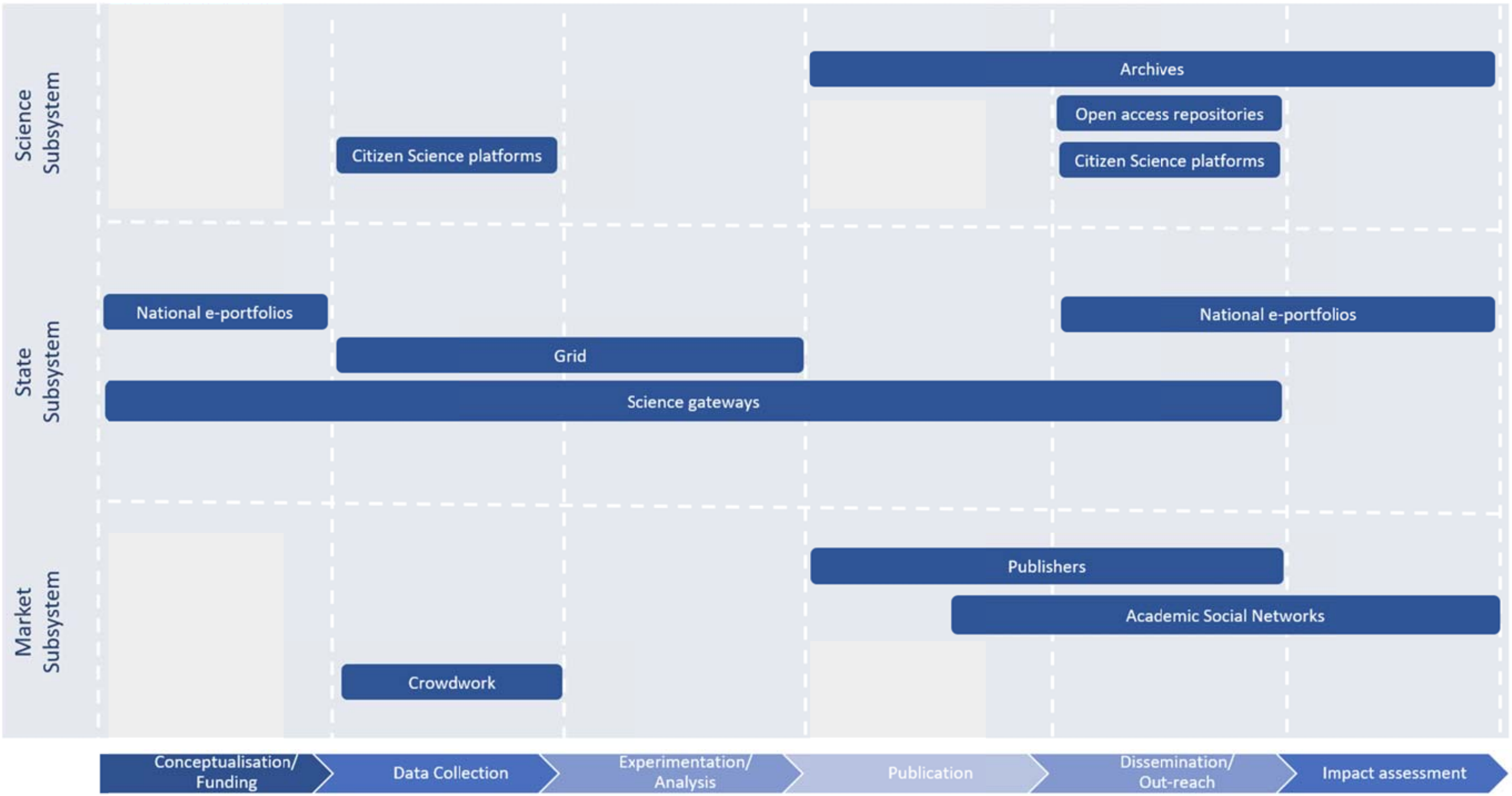
The science platform ecosystem is complex and there is a constellation of scientific platforms from varying historical moments coexisting (Figure 2.2). The types of platforms analyzed in the previous section have in common the fact they are governing systems of virtual spaces that leverage network effects towards one (or more) phase(s) of the scientific research process (Figure 2.3). Notwithstanding that, they also differ in terms of the social subsystem (Science, Market and State, as presented previously), and at least, in two other aspects: (i) the degree of integration into the research life cycle; and, (ii) the research phase they target.

Figure 2.2 – Science digital platform ecosystem, 1990-2020



Source: Authors' own

Figure 2.3 – Science digital platform ecosystem by science phases



Source: Authors' own

2.4.1 Degree of integration into the research life cycle

It is possible to identify scientific digital platforms according to their degree of integration into the research life cycle. When scientific digital platforms are so embedded into the scientific practice, being more than just technical assemblage of things, they can be considered part of the research infrastructure, which can be understood "as deeply relational and adaptive systems where the material and social aspects are in permanent interplay. They are embedded in the social practice of research and influenced by environmental factors". (FECHER *et al.*, 2021, p. 500). In other words, when platforms are part of the actual scientific practice, they can be considered part of the infrastructure, otherwise, they are purely a service.

For example, while scientific journals going online shows the platformization of an infrastructure, national e-portfolios becoming indispensable (Lattes Platform is an example) is example of infrastructuralization of a platform. In 2021, Lattes Platform suffered a breakdown and was unavailable for more than a month, clarifying its infrastructural character: its criticality was revealed when it failed and lost its "invisibility" (CHIARINI; SILVA, 2022). Other platforms, on their turn, are still not indispensable or ubiquitous, *e.g.*, crowdwork is restricted to a specific type of research – surveys – and there are traditional substitutes. Still others, such as ASNs seem to be aiming at dominance through infrastruturalization, but they seem to be in a process of embedding into scientific practices (FECHER *et al.*, 2021). This process that naturally takes some time due to the need to "conquer" users, may take a little longer in this case. As Table 2.3 demonstrates, ASNs and publishers overlap in terms of the phase of the scientific process; this puts them on a collision course, as already discussed.

Table 2.3 – Scientific digital platforms types by social subsystem and by degree of integration into the research life cycle

	Degree of integration into the research life cycle		
Social Subsystems	**Service**	**"In transition"**	**Research Infrastructure**
Science	-	Citizen science	Archives, Repositories
State	-	-	Cyberinfrastructure, Science gateways, e-portfolios
Market	Crowdwork	ASNs	Publisher's scientific journals

Source: Authors' own

2.4.2 Science phases: from data collection to dissemination

2.4.2.1 Data collection

Digitization has impacted one of the most fundamental and traditional phases of the scientific research process: data collection. With digitization, trends such as open data, open government data and citizen science have created new ways of obtaining data for scientific purposes (DAI; SHIN; SMITH, 2018).

We presented that State actors sponsored the development of two types of digital platforms with impacts on the data collection phase. Both cyberinfrastructure and science portals offer tools to expand the potential for capturing and manipulating data. This is due not only to access to computational power and new customized software, but also to the network of researchers formed around science gateways. Both types are infrastructures, but science gateways are fragmented and differ in terms of research areas. Although the funding is from the State, the management of these platforms is carried out closely within the Science subsystem. Therefore, these platforms present participatory governance (which aggregates actors from more than one social subsystem).

Citizen science platforms enable distributed capture and centralized analysis of data. There are many challenges related to obtaining this type of data: public mobilization, correctly selected audience, organizational obstacles, technical difficulties and questions about the validity of the data (DAI; SHIN; SMITH, 2018). It should be noted that citizen science platforms are voluntary initiatives led by the research community. Their governance is participatory and their scope is limited to some research areas, positioning them as a service and not as an infrastructure. Crowdwork platforms offer a cheap and efficient way to collect data for surveys. Because they are proprietary, they do not face some of the difficulties that citizen science platforms do, such as encouraging participation (which occurs via remuneration). However, there are questions about the quality of these data. Crowdwork platforms offer a service to a very specific niche of the Science subsystem: researchers whose method is based on surveys. Even if this makes the use more concentrated in the areas of psychology and medicine, for example, this service may, at some point in the future, become a proprietary scientific infrastructure.

2.4.2.2 Experimentation/analysis

Experimentation is the *leitmotif* of grid computing enabled by cyberinfrastructure initiatives. As we have seen, without the friendly and customizable complement of science gateways, their effectiveness was constrained. Together, they engender the “platformization of [experimentation] infrastructure” (PLANTIN *et al*., 2018). As we have seen, its governance is participatory, which does not exempt it from problems, but it seems to guarantee that major conflicts are resolved through consultation within the platforms themselves. According to Dai, Shin, and Smith (2018), citizen science platforms are in the process of extending their capabilities. They would become “a powerful mechanism for co-designing experiments and co-producing scientific knowledge with large communities of interested actors” (DAI; SHIN; SMITH, 2018, p. 15).

2.4.2.3 Publication

The publishing phase witnesses friction between platform types in a quest for space. It is marked by the platforming of an infrastructure almost as old as science: scientific journals are owned by half a dozen publishers and have been digitized over the last two decades. A few years ago, publishers were so well positioned that some researchers saw no solution other than to rebel against the transfer of resources (mostly public) from universities to the cashier of an oligopoly of scientific publishing. Today, ASNs are fighting a battle over the interface (SRNICEK, 2017): by making articles available on their platforms, ASNs seek to neutralize the utility of publishers and redirect the flow of attention to their interface. As publishers activated legal mechanisms to reverse this strategy, ASNs now seem to have adopted two complementary strategies: approaching the low-end of the market, i.e., small publishing houses; and offer a new publishing medium embedded in their own platforms (*e.g*., Academia Letters). Today ASNs mainly offer two services: alternative publishing to traditional publishers and dissemination. Their strategy is clear: move from a service provider to a provider of proprietary scientific infrastructure (FECHER *et al*., 2021). For this reason, they are “transitioning” from service to infrastructure (Table 2.3).

2.4.2.4 Dissemination/Outreach

Mostly all groups of digital platforms contribute to the dissemination phase of the research process. ArXiv is dedicated to “the rapid dissemination of scholarly scientific research”[35] and Zooniverse curates “datasets useful to the wider research community, and many publications”[36]. Within State subsystem, national e-portfolios such as Lattes Platform can assist in the preservation and dissemination of scientific production[37] and science gateways as HUBzero “publish data, [and] share resources”[38] and, although it serves different communities, “the hubs all support (…) dissemination of scientific models” (MCLENNAN; KENNELL, 2010, p. 48). Finally, considering the Market subsystem, publishers as Elsevier has the “job to disseminate research and improve understanding of science”[39] and ASNs as ResearchGate “offer[s] a home for you—a place to share your work and connect with peers around the globe”[40] and Academia.edu aims to “broad dissemination”[41].

2.5 Conclusions

We defined scientific digital platforms and outlined a typology based on three criteria: association with a specific social sub-system, degree of integration into the research process and phase of the scientific research process. We observe that there is a large number of platforms that become infrastructures and infrastructures that have become platformed, although this does not invariably occur. Our long-term perspective and broad scope allow us to frame the platformization of science as a process that already has a broad legacy constituted from the development of platforms by actors from different social subsystems (Market, State and Science). Research communities and the public administration were the first actors to be engaged in this process in the first half of the 1990s, while private companies engaged in this process in the late 1990s. We now observe a very dynamic commitment of the Market to be part of the science platformization process.

The limitations of the article stem from our scope and methodological choices. When trying to understand the big picture, there is no space for detailing each of the

[35] https://arxiv.org/help/policies/code_of_conduct, accessed on 05/03/2022.
[36] https://www.zooniverse.org/about, accessed on 05/03/2022.
[37] https://lattes.cnpq.br/, accessed on 05/03/2022.
[38] https://hubzero.org/, accessed on 05/03/2022.
[39] https://www.elsevier.com/open-science/science-and-society, accessed on 05/03/2022.
[40] https://www.researchgate.net/about, accessed on 05/03/2022.
[41] https://www.academia.edu/journals/1/about, accessed on 05/03/2022.

taxonomy presented, and the mini-cases described offer only a glimpse of their main features. Furthermore, by choosing to structure our paper around the notion of large groups of platforms, we can bypass platforms that exhibit mixed characteristics, which would place them in more than one group in our taxonomy. Our clustering proposal is limited in the sense that digital platforms are highly flexible and dynamic (VAN DIJCK, 2020). New configurations and functionalities are added and deleted overnight, super platforms emerge and coalesce once separate categories. Still, to understand the historical movement in the last thirty years, the groupings come close to a rough taxonomy.

As a result of our exploratory study, we conclude with three issues that deserve further attention. First, the relationship between scientific platforms and platform economics in general deservers further investigation. We found evidences that platforms for science exhibited traces of avant-garde platform economics. Platforms from the Market subsystem were not always the "innovation locomotive" as one may expect. Many traits of private scientific platforms, and even platform economy general features, can be traced back to State or Science social subsystems initiatives. In order to understand the genesis of the platform economy, this topic comes up with great importance.

Although it was not the scope of this paper, we also believe that the science platformization process might have considerable impacts on science production as a social process, and it may alter established principles and values of the scientific communities. "Research communities have always been vitual communities that cross national and cultural borders. (…)", however, these interactions were "limited by constraints, both physical and technical; now, as a result of advances in ICT, interaction is unconstrained, and instantaneous" (NOWOTNY; SCOTT; GIBBONS, 2003, p. 187).

There are already studies on how ICT-based technologies, such as ASNs transform the ways scholars conduct their work (BORGMAN, 2007; VELETSIANOS; KIMMONS, 2012; VELETSIANOS, 2016; VELETSIANOS; JOHNSON; BELIKOV, 2019; WELLER, 2011), for example, by shaping e-publishing (TAHA *et al*., 2017). We wonder whether science digital platforms could be considered a Grilichesian "invention of a method of inventing" (IMI) (GRILICHES, 1958), with a much larger impact on science process than simply considering the "general purpose technology" (GPT) character of digital platforms.

Finally, it is necessary to critically analyze digital platforms as scientific infrastructures. In particular, consideration should be given to the extent to which private scientific infrastructures serve the public interest well. Coming up with the classes and classifying them according to the framework provides only a very superficial idea of how these structures that co-opt scientific flows affect society beyond the scientific sphere. An in-depth investigation could consider how each platform group is positioned, considering whether there are subordinate relationships between them (VAN DIJCK, 2020) and how these relations affect social-relevant values and outputs.

2.6 References

ACS, Z. J. et al. **The Evolution of the Global Digital Platform Economy: 1971-2021**. SSRN Electronic Journal, 2021. DOI: https://doi.org/10.2139/ssrn.3785411.

ALLAN, R. **Virtual Research Environments**: From Portals to Science Gateways. Cambridge: Chandos Publishing, 2009.

ArXiv. **ArXiv Annual Report 2020**. New York: Cornell Tech, 2020. Available in: https://indd.adobe.com/view/fdd63397-e4b0-41af-b479-4845dc1ef48e. Access at: 22 mar. 2022.

AVGEROU, C. Information Systems for Development Planning. **International Journal of Information Management**, v. 13, p. 260-73, 1993.

BARKER, M. *et al*. The Global Impact of Science Gateways, Virtual Research Environments and Virtual Laboratories. **Future Generation Computer Systems**, v. 95, p. 240-48, 2019. DOI: https://doi.org/10.1016/j.future.2018.12.026.

BARTLING, S.; SASCHA F. Towards Another Scientific Revolution. *In*: BARTLING, S.; FRIESIKE, S. (eds.). **Opening Science**: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing. New York: Springer, 2014. p. 3-15.

BORGMAN, C. L. **Scholarship in the Digital Age Information, Infrastructure, and the Internet**. Cambridge (USA): MIT University Press, 2007.

BRANSCOMB, L. M. U.S. Scientific and Technical Information Policy in the Context of a Diffusion-Oriented National Technology Policy. **Government Publications Review**, v. 19, p. 469-482, 1992.

CHAWLA, D. Publishers Take ResearchGate to Court, Alleging Massive Copyright Infringement. **Science**, Oct., 2017. DOI: https://doi.org/10.1126/science.aaq1560.

CHIARINI, T.; SILVA, V. **The Platformization of Science: Lattes Platform in a Crossroad?** DP 02?? Discussion Paper. Brasilia: IPEA, 2022.

COBLANS, H. Control and Use of Scientific Information. **Nature**, v. 226, p. 319-321, 1970.

DAI, Q.; SHIN, E.; SMITH, C. **Open and Inclusive Collaboration in Science**: A Framework. 2018/07. OECD Science, Technology and Industry Working Papers. Paris: OECD, 2018.

DELFANTI, A. The Financial Market of Ideas: A Theory of Academic Social Media. **Social Studies of Science**, v. 51, n. 2, p. 259-276, 2021. DOI: https://doi.org/10.1177/0306312720966649.

DIEPENBROEK, M. *et al*. PANGAEA—an Information System for Environmental Sciences. **Computers & Geosciences**, v. 28, n. 10, p. 1201-1210, 2002. DOI: https://doi.org/10.1016/S0098-3004(02)00039-0.

DIJCK, J. V.; POELL, T.; WAAL, M. **The Platform Society**: Public Values in a Connective World. Oxford: Oxford University Press, 2018.

DIJCK, J. V. Seeing the Forest for the Trees: Visualizing Platformization and Its Governance. **New Media & Society**, July, 1-19, 2020. DOI: https://doi.org/10.1177/1461444820940293.

EC. **Open Innovation, Open Science, Open to the World**. Brussels: Directorate-General for Research and Innovation: European Commission, 2016.

ELSE, H. Major Publishers Sue ResearchGate over Copyright Infringement. **Nature**, Oct., 2018. DOI: https://doi.org/10.1038/d41586-018-06945-6.

EYAL, P. *et al*. Data Quality of Platforms and Panels for Online Behavioral Research. **Behavior Research Methods**, Sep., 1-20, 2021. DOI: https://doi.org/10.3758/s13428-021-01694-3.

FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. *In*: BARTLING, S.; FRIESIKE, S. (eds.). **Opening Science**: The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing. New York: Springer, 2014. p. 17-47.

FECHER, B. *et al*. Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure. **Science and Public Policy**, v. 48, n. 4, p. 499-507, 2021. DOI: https://doi.org/10.1093/scipol/scab026.

FRENKEN, K.; SCHOR, J. Putting the Sharing Economy into Perspective. **Environmental Innovation and Societal Transitions**, v. 23, p. 3-10, 2017. DOI: https://doi.org/10.1016/j.eist.2017.01.003.

FRENKEN, K.; FUENFSCHILLING, L. The Rise of Online Platforms and the Triumph of the Corporation. **International Journal for Sociological Debate**, v. 14, n. 3, p. 101-13, 2020. DOI: https://doi.org/https://doi.org/10.6092/issn.1971-8853/11715.

GAWER, A. Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. Innovation, v. 24, n. 1, p. 110-124, 2021. DOI: https://doi.org/10.1080/14479338.2021.1965888.

GINSPARG, P. ArXiv at 20. **Nature**, v. 476, n. 7359, p. 145-47, 2011. DOI: https://doi.org/10.1038/476145a.

GINSPARG, P. Preprint Déjà Vu. **The EMBO Journal**, v. 35, n. 24, p. 2620-25, 2016. DOI: https://doi.org/10.15252/embj.201695531.

GRILICHES, Z. Research Costs and Social Returns: Hybrid Corn and Related Innovations. **Journal of Political Economy**, v. 66, n. 5, p. 419-31, 1958. DOI: https://doi.org/10.1086/258077.

HELMOND, A.; NIEBORG, D. B.; VAN DER VLIST, F. N. Facebook's Evolution: Development of a Platform-as-Infrastructure. **Internet Histories**, v. 3, n. 2, p. 123-46, 2019. DOI: https://doi.org/10.1080/24701475.2019.1593667.

HELMOND, A. The Platformization of the Web: Making Web Data Platform Ready. **Social Media + Society 1**, n. 2, p. 1-11, 2015. DOI: https://doi.org/10.1177/2056305115603080.

HEY, T.; TREFETHEN, A. E. The UK E-Science Core Programme and the Grid. **Future Generation Computer Systems**, v. 18, p. 1017-1031, 2002. DOI: https://doi.org/S0167-739X(02)00082-1.

HEY, T.; TREFETHEN, A. E. E-Science and Its Implications. **Philosophical Transactions of the Royal Society A**: Mathematical, Physical and Engineering Sciences, v. 361, p. 1809-1825, 2003. DOI: https://doi.org/rsta.2003.1224.

HODGSON, G. M. Taxonomic Definitions in Social Science, with Firms, Markets and Institutions as Case Studies. **Journal of Institutional Economics**, v. 15, n. 2, p. 207-333, 2019. DOI: https://doi.org/10.1017/S1744137418000334.

HOFMAYER, S. *et al*. **System, computer program product and computer-implemented method for sharing academic user profiles and ranking academic users**. US 20130346497 A1, issued 2015.

HOWCROFT, D.; BERGVALL-KÅREBORN, B. A Typology of Crowdwork Platforms. **Work, Employment and Society**, v. 33, n.1, p. 21-38, 2019. DOI: https://doi.org/10.1177/0950017018760136.

JANKOWSKI, N. W. Exploring E-Science: An Introduction. **Journal of Computer-Mediated Communication**, v. 12, n. 2, p. 549-62, 2007. DOI: https://doi.org/10.1111/j.1083-6101.2007.00337.x.

JORDAN, K. From Social Networks to Publishing Platforms: A Review of the History and Scholarship of Academic Social Network Sites. **Frontiers in Digital Humanities**, v. 6, n. 5, p. 1-53, 2019. DOI: https://doi.org/10.3389/fdigh.2019.00005.

KENNEY, M.; ZYSMAN, J. The Platform Economy: Restructuring the Space of Capitalist Accumulation. **Cambridge Journal of Regions, Economy and Society**, v. 13, n. 1, p. 55-76, 2020. DOI: https://doi.org/10.1093/cjres/rsaa001.

KONG, X *et al*. 2019. Academic Social Networks: Modeling, Analysis, Mining and Applications. **Journal of Network and Computer Applications**, v. 132, n. April, p. 86-103, 2020. DOI: https://doi.org/10.1016/j.jnca.2019.01.029.

KRAUT, R. E.; RESNICK, P. Encouraging Contribution to Online Communities. *In*: KRAUT, R. E.; RESNICK, P. **Building Successful Online Communities**: Evidence-Based Social Design. Cambridge (MA): Massachusetts Institute of Technology and Center for International Development, Harvard University: The MIT Press, 2011. P. 21-76.

KULLENBERG, C.; KASPEROWSKI, D. What Is Citizen Science? - A Scientometric Meta-Analysis. Edited by Pablo Dorta-González. **PLOS ONE**, v. 11, n. 1, p. e0147152, 2016. DOI: https://doi.org/10.1371/journal.pone.0147152.

LANE, J. Let's Make Science Metrics More Scientific. **Nature**, v. 464, p. 488-89, 2010.

LANGLOIS, R. N. Design, Institutions, and the Evolution of Platforms. **Journal of Law, Economics & Policy**, v. 9, n. 1, p. 1-13, 2012.

LARIVIÈRE, VINCENT, STEFANIE HAUSTEIN, AND PHILIPPE MONGEON. The Oligopoly of Academic Publishers in the Digital Era. Edited by Wolfgang Glanzel. **PLOS ONE**, v. 10, n. 6, p. e0127502, 2015. DOI: https://doi.org/10.1371/journal.pone.0127502.

LEMMENS, R. *et al*. A Conceptual Model for Participants and Activities in Citizen Science Projects. *In*: VOHLAND, K. *et al*. **The Science of Citizen Science**. Cham: Springer, 2021. p. 159-182.

LIU, H.-Y. *et al*. Citizen Science Platforms. *In*: VOHLAND, K. *et al*. **The Science of Citizen Science**. Cham: Springer, 2021. p. 439-459.

MADDURI, R. *et al*. The Globus Galaxies Platform: Delivering Science Gateways as a Service. **Concurrency and Computation: Practice and Experience**, v. 27, n. 16, p. 4344-60, 2015. DOI: https://doi.org/10.1002/cpe.3486.

MADHAVAN, K.; ZENTNER, M.; KLIMECK, G. Learning and Research in the Cloud. **Nature Nanotechnology**, v. 8, n. 11, p. 786-89, 2013. DOI: https://doi.org/10.1038/nnano.2013.231.

MANSELL, R.; STEINMUELLER, W. E. **Advanced Introduction to Platform Economics**. Cheltenham: Edward Elgar, 2020.

MCLENNAN, M.; KENNELL, R. HUBzero: A Platform for Dissemination and Collaboration in Computational Science and Engineering. **Computing in Science & Engineering**, v. 12, n. 2, p. 48-53, 2010. DOI: https://doi.org/10.1109/MCSE.2010.41.

MCLENNAN, M. *et al*. HUBzero and Pegasus: Integrating Scientific Workflows into Science Gateways. **Concurrency and Computation: Practice and Experience**, v. 27, n. 2, p. 328-443, 2015. DOI: https://doi.org/10.1002/cpe.3257.

MIROWSKI, P. The Future(s) of Open Science. **Social Studies of Science**, v. 48, n. 2, p. 171-203, 2018. DOI: https://doi.org/10.1177/0306312718772086.

MUKHOPADHYAY, S.; BOUWMAN, H.; JAISWAL, M. P. An Open Platform Centric Approach for Scalable Government Service Delivery to the Poor: The Aadhaar Case. **Government Information Quarterly**, v. 36, n. 3, p. 437-448, 2019. DOI: https://doi.org/10.1016/j.giq.2019.05.001.

NOORDEN, R. V. Publishers Threaten to Remove Millions of Papers from ResearchGate. **Nature**, Oct, 2017. DOI: https://doi.org/10.1038/nature.2017.22793.

NOORDEN, R. V. Online Collaboration: Scientists and the Social Network. **Nature**, v. 512, n. 7513, p. 126-129, 2014. DOI: https://doi.org/10.1038/512126a.

NOWOTNY, H.; SCOTT, P.; GIBBONS, M. ntroduction: 'Mode 2' Revisited: The New Production of Knowledge. **Minerva**, v. 41, p. 179-194, 2003. DOI: https://doi.org/https://doi.org/10.1023/A:1025505528250.

OTTO, B.; JARKE, M. Designing a Multi-Sided Data Platform: Findings from the International Data Spaces Case. **Electronic Markets**, v. 29, n. 4, p. 561-580, 2019. DOI: https://doi.org/10.1007/s12525-019-00362-x.

PARSONS, P. *et al*. SGCI Incubator and Its Role in Workforce Development: Lessons Learned from Training, Consultancy, and Building a Community of Community-Builders for Science Gateways. **Practice and Experience in Advanced Research Computing**, p. 491-94, 2020. DOI: https://doi.org/10.1145/3311790.3400850.

PITKOW, J. B.; SCHUETZE, H. **System and method for searching and recommending objects from a categorically organized information repository**. US 20020016786 A1, issued 2006.

PLANTIN, J. et al. Infrastructure studies meet platform studies in the age of Google and Facebook. **New Media & Society**, v. 20, n. 1, p. 293-310, 2018.

POELL, T.; NIEBORG, D.; VAN DIJCK, J. Platformisation. **Internet Policy Review**, v. 8, n. 4, p. 1-13, 2019. DOI: https://doi.org/10.14763/2019.4.1425.

SARACEVID, T. Progress in Documentation. Perception of the Needs for Scientific and Technical Information in Less Developed Countries. **Journal of Documentation**, v. 36, n. 3, p. 214-267, 1980.

SAUERMANN, H. *et al*. Citizen Science and Sustainability Transitions. **Research Policy**, v. 49, n. 5, p. 103978, 2020. DOI: https://doi.org/10.1016/j.respol.2020.103978.

SCHWARZ, J. A. Platform Logic: An Interdisciplinary Approach to the Platform-Based Economy. **Policy & Internet**, v. 9, n. 4, p. 374-94, 2017. DOI: https://doi.org/10.1002/poi3.159.

SIMPSON, R.; PAGE, K. R.; DE ROURE, D. Zooniverse: Observing the World's Largest Citizen Science Platform. **Proceedings of the 23rd International Conference on World Wide Web**, p. 1049-1054, 2014. DOI: https://doi.org/10.1145/2567948.2579215.

SUBER, P. **Open Access**. Cambridge: The MIT Press, 2012.

TABARÉS, R. HTML5 and the Evolution of HTML; Tracing the Origins of Digital Platforms. **Technology in Society**, v. 65, n. May, p. 101529, 2021. DOI: https://doi.org/10.1016/j.techsoc.2021.101529.

TAHA, N. *et al*. **Social Media Shaping E-Publishing and Academia**. Springer International Publishing, 2017.

THOMPSON, M.; VENTERS, W. Platform, or Technology Project? A Spectrum of Six Strategic 'Plays' from UK Government IT Initiatives and Their Implications for Policy. **Government Information Quarterly**, v. 38, n. 4, p. 101628, 2021. DOI: https://doi.org/10.1016/j.giq.2021.101628.

TIM BERNERS-LEE.; FISCHETTI, M. **Weaving the Web**: The Original Design and Ultimate Destiny of the World Wide Web. New York: Harper Business, 2000.

TINATI, R. *et al*. Designing for Citizen Data Analysis. **Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems**, p. 4069-78, 2015. DOI: https://doi.org/10.1145/2702123.2702420.

UNESCO. UNISIST: Study Report on the Feasibility of a World Science Information System. Paris: United Nations Educational, Scientific and Cultural Organization, 1971.

VELETSIANOS, G.; KIMMONS, R. Networked Participatory Scholarship: Emergent Techno-Cultural Pressures toward Open and Digital Scholarship in Online Networks. **Computers & Education**, v. 58, n. 2, p. 766-74, 2012. DOI: https://doi.org/10.1016/j.compedu.2011.10.001.

VELETSIANOS, G.; JOHNSON, N.; BELIKOV, O. Academics' Social Media Use over Time Is Associated with Individual, Relational, Cultural and Political Factors. **British Journal of Educational Technology**, v. 50, n. 4, p. 1713-1728, 2019. DOI: https://doi.org/10.1111/bjet.12788.

VELETSIANOS, G. **Social Media in Academia**: Networked Scholars. Routledge, 2016.

VICKERY, B. A Century of Scientific and Technical Information. **Journal of Documentation**, v. 55, n. 5, p. 476-527, 1999. DOI: https://doi.org/10.1108/EUM0000000007155.

WAGENKNECHT, K. *et al*. 2021. EU-Citizen.Science: A Platform for Mainstreaming Citizen Science and Open Science in Europe. **Data Intelligence**, v. 3, n. 1, p. 136-149, 1999. DOI: https://doi.org/10.1162/dint_a_00085.

WELLER, M. **The Digital Scholar**. How Technology Is Transforming Scholarly Practice. New York: Bloomsbury Academic, 2011.

WORK, J. DN.; BLUE, A.; HOFFMAN, R. **Method and system for reputation evaluation of online users in a social networking scheme**. US 20060042483 A1, issued 2005.

WYNN, J. **Citizen Science in the Digital Age**. Rhetoric, Science, and Public Engagement. Tuscaloosa: The University of Alabama Press, 2017.

YIN, R. K. **Case Study Research**. Design and Methods. Third Edit. Thousand Oaks: Sage Publications Inc, 2003.

2.7 Annex

Table 2.4 – Documents accessed

Platforms	Type of Document	Links	References
ArXiv	Website	https://arxiv.org/	-
	Website	https://arxiv.org/help/policies/code_of_conduct	-
	Corporate Brochure	https://indd.adobe.com/view/fdd63397-e4b0-41af-b479-4845dc1ef48e	(ArXiv 2020)
	Nature Article	http://www.nature.com/articles/476145a	(Ginsparg 2011)
Zooniverse	Website	https://www.zooniverse.org	-
	Corporate Brochure	https://www.zooniverse.org/about/highlights	-
National e-portfolios	Study Report	https://unesdoc.unesco.org/ark:/48223/pf0000064862	(UNESCO 1971)
Cyberinfrastructure networks	Website	https://www.nsf.gov/div/index.jsp?div=OAC	-
Terascale Computing System	NSF Award	https://www.nsf.gov/awardsearch/showAward?AWD_ID=0085206	-
The TeraGrid: Cyberinfrastructure for 21st Century Science and Engineering	NSF Award	https://www.nsf.gov/awardsearch/showAward?AWD_ID=0122296	-
XSEDE: eXtreme Science and Engineering Discovery Environment	NSF Award	https://www.nsf.gov/awardsearch/showAward?AWD_ID=1053575	-
XSEDE 2.0: Integrating, Enabling and Enhancing National Cyberinfrastructure with Expanding Community Involvement	NSF Award	https://www.nsf.gov/awardsearch/showAward?AWD_ID=1548562	-
GHub	Website	https://vhub.org/groups/ghub	-
Science Gateway Catalog	Website	https://catalog.sciencegateways.org/#/home	-
HUBzero	Website	https://hubzero.org/about	-
	Nature Article	https://doi.org/10.1038/nnano.2013.231	(Madhavan, Zentner, and Klimeck 2013)
Lattes Platform	Website	https://lattes.cnpq.br/	
	Nature Article	https://doi.org/10.1038/464488a	(Lane 2010)

Elsevier	Website	https://www.elsevier.com/open-science/science-and-society	-
	Corporate Brochure	https://www.elsevier.com/__data/assets/pdf_file/0010/1143001/Elsevier-corporate-brochure-2021.pdf	-
HASTAC	Website	https://www.hastac.org/	-
Academia.Edu	Website	https://www.academia.edu/letters/about	-
Amazon Mechanical Turk	Website	https://www.mturk.com/	-
Prolific Academic	Website	https://www.prolific.co/about	-
ResearchGate	Website	https://www.researchgate.net/about	-
	Website	https://www.researchgate.net/blog/post/a-note-on-recent-content-takedowns	-
	Nature Article	https://doi.org/10.1038/d41586-018-06945-6	(Else 2018)
	Nature Article	https://doi.org/10.1038/512126a	(Van Noorden 2014)
	Nature Article	https://doi.org/10.1038/nature.2017.22793	(Van Noorden 2017)

Source: Authors' Own. Note: all links were accessed on 05/03/2022

CHAPTER 3 – THE PLATFORMIZATION OF SCIENCE: LATTES PLATFORM AT A CROSSROADS?[42]

3.1 Introduction

Several scholars point to the platformization process that society is currently passing through (DIJCK; POELL; WAAL, 2018; KENNEY; ZYSMAN, 2016; POELL; NIEBORG; VAN DIJCK, 2019). Although there are different perspectives, in general terms, platformization is understood as "the penetration of the infrastructures, economic processes, and governmental frameworks of platforms in different economic sectors and spheres of life" (POELL; NIEBORG; VAN DIJCK, 2019, p. 5-6). The spread of digital platforms has also encompassed the scientific field and many studies focus on the most "tangible" part of digital scientific platforms, that is, academic social networks (ASN) (COZMA; DIMITROVA, 2020; MASON, 2020; ORDUNA-MALEA *et al*., 2017; THELWALL; KOUSHA, 2015; VELETSIANOS, 2016; ZHAO *et al*., 2020). ASNs were mostly developed in the early 2010s – the two largest ones being Academia.edu and ResearchGate[43] – following in the footsteps of conventional social networks – such as MySpace, Facebook, and LinkedIn – and they aimed to create virtual *loci* for individual communication and knowledge exchange, constituting a specific online community for researchers.

The growth of ASN platforms in the last decade has led researchers to question whether they were becoming private scientific infrastructures (PLANTIN; LAGOZE; EDWARDS, 2018). This question is legitimate since it is possible to observe digital platforms being controlled by the private sector, which is positioning them as intermediaries between infrastructure platforms and sectorial platforms (VAN DIJCK, 2020). That intermediate position allows them to control essential flows of data, services, and information across entire platform ecosystems. By becoming critical elements within this ecosystem, intermediary platforms generate a sort of "infrastructuralization of platforms" (PLANTIN *et al*., 2018). *Pari passu*, motivated by the control of data flows, intermediary platforms seek to expand "upstream" and "downstream", occupying spaces of often public infrastructure platforms,

---

[42] Originally published as: SILVA, V., CHIARINI, T. 'The Platformization of Science: Lattes Platform at a Crossroads?', IPEA Discussion Paper, DP 268, 2022. Available in https://www.ipea.gov.br/portal/images/stories/PDFs/TDs/ingles/dp_268.pdf.

[43] HASTAC – Humanities, Arts, Science, and Technology Alliance and Collaboratory – is considered the world's first ASN and it was developed within the research community and was founded in 2002. ResearchGate and Academia.edu were both founded in 2008.

causing what can be called the "platformization of infrastructures" movement (PLANTIN *et al.*, 2018).

From the above, we seek to understand if and how the phenomenon of "infrastructuralization of platforms" is evolving in Brazil; more specifically, we focus on the possibilities and roles of established scientific platforms in the light of the new trends. In this Discussion Paper, we focus our analysis on the Lattes Platform – a public initiative developed endogenously in the late 1990s which was very innovative – since it is a representative case of a scientific platform and is a paramount tool for the Brazilian scientific system. Although there are many academic studies published using data provided by the Lattes Platform, we identified a lack of official reports and articles that address Lattes as a digital platform and its adequacy in the context of the platform economy (SILVA; BONACELLI; PACHECO, 2020). In doing so, not only do we dialogue with the latest literature on scientific "platform infrastructuring" (PLANTIN *et al.*, 2018; PLANTIN; LAGOZE; EDWARDS, 2018), but we also provide new insights for policy-making considering that both the early-stage technology choices made by the National Council for Scientific and Technological Development[44] (early adopter) and the constant budget cuts for its maintenance and upgrading have strongly affected the Lattes Platform's trajectory. Consequently, the Lattes Platform may be locked-in to a technological trajectory that could lead to the stagnation of its functionalities, threatening its relevance in the long run.

Our Discussion Paper is organized as follows. Firstly, we present our research design, and afterwards we make a brief presentation of digital platforms as infrastructures. Subsequently, we present the Lattes Platform (its brief history and its components) using the multilevel framework proposed by Manca (2017) for analyzing ASN; however, we improve her proposition by including new categories. Finally, we close the paper with some concluding remarks and policy implications.

3.2 Research design

The discussion we propose in this article sheds light on digital platforms, showing that they are very pervasive and, thanks to the ever-continuing advances and diffusion of a constellation of interconnected technologies, they are progressively transforming socio-economic and techno-cultural spheres. In that context, they are altering science as well by:

---

[44] *Conselho Nacional de Desenvolvimento Científico e Tecnológico*, CNPq.

a) creating new ways to assess academic and scientific activities,

b) transforming the way scholars conduct their work (BORGMAN, 2007; VELETSIANOS, 2016; VELETSIANOS; JOHNSON; BELIKOV, 2019; VELETSIANOS; KIMMONS, 2012; WELLER, 2011); by shaping e-publishing (TAHA *et al*., 2017); and,

c) allowing policymakers, funders, institutions, and managers to have access to a great deal of data, guiding their policy propositions to allocate resources for the development of scientific and technological activities.

By scientific digital platform, we refer to an assemblage of distinct digital technologies used by policymakers, scientific communities, and profit-seeking companies, for example, to gather scientific information (repository function), to disseminate scientific knowledge (transfer function), and to facilitate interactions (network function). With this previous definition, we do not want to propose a "straightjacket concept"; rather, we intend to use a very broad conceptualization able to cover technologies with different maturity levels.

Manca (2017) proposed a descriptive approach to analysing ASN by drawing up a multilevel framework, blending the contributions of Dijck (2013) and Veletsianos and Kimmons (2012). She considered ASN's systemic/infrastructural and personal/individual dimensions in three layers:

a) socio-economic (i.e., macro-level considering components related to ownership, governance, and business model);

b) techno-cultural (i.e., meso-level considering components related to the technology itself, user/usage, and content); and

c) networked scholar (i.e., micro-level considering components related to networking, knowledge sharing, and identity).

Using the socio-economic and techno-cultural layers proposed by Dijck (2013), Manca (2017) reiterated that platforms encompass "coevolving networks of people and technologies with economic infrastructure and legal-political governance and blends techno-cultural and political economy views" (MANCA, 2017, p. 23). Finally, the networked participatory layer was inspired by Veletsianos and Kimmons (2012) who acknowledge individuals' learning and knowledge in networked spaces.

We use the same three layers; however, we bring forward other components, advancing her framework (Table 3.1). We do it since Manca's proposition was thought of "in the light of theoretical frameworks developed in the educational technology sector and aimed at analyzing social digital scholarship practice" (MANCA, 2017, p. 22). We are also concerned about deepening our knowledge on technological upgrading (considering platforms' infrastructure role and their open innovation mechanisms) and about other socio-economic components (platforms' market scope, economic effects, and the path-dependent dynamics).

Table 3.1 – Multilevel framework for analyzing academic platforms

Levels	**Layers**	**Components**	**Description**
Macro	Socio-economic	Ownership	components that govern commercial and non-profit platforms according to different policies
		Governance	technical and social protocols and sets of rules for managing user activities
		Business model	components that mediate the engineering of connectivity through subscription models
		Market scope	national, regional, or global scope
		Economic effects	scope, network, or feedback effects at work
		Infrastructure	features of the infrastructural role
		Path dependent dynamics	mechanisms of the trajectory that may lead to lock-in effects
Meso	Techno-cultural	Technological components	components that help encode activities into a computational architecture that steers user behavior
		User/usage	components that orient user agency and implicit and explicit participation
		Content	components that determine the standardization of content and the uniform delivery of products
		Open innovation mechanism	components that allow the generation of new ideas, products, and services
Micro	Networked-scholar	Networking components	components that allow connectivity of communication and collaboration
		Knowledge sharing	components that allow knowledge sharing, concerning the collective and distributed learning dimension
		Identity	components that shape academic personae (reputation and trust)

Source: Authors' own based on Manca (2017)

3.3 Digital platforms as infrastructures

It is possible to find in the platform economy components also present in the classical infrastructure literature such as scale, ubiquity, essentiality. The analogy is with the relational nature of what it means to be infrastructure: when platforms offer digital services that become indispensable to other actors, such as other platforms and even public bodies, they have become infrastructures (KENNEY; ZYSMAN, 2016; POELL, 2020), that is, "infrastructuralization of platform" (PLANTIN *et al*., 2018).

Platforms and infrastructures do have some similarities, such as "embeddedness, a degree of invisibility, extensibility, and broad coverage" (PLANTIN *et al*., 2018, p. 14). However, they also have a considerable difference: unlike infrastructure, platforms deliberately avoid building gateways between systems to force users to lock-in. However, this is not an attribute inherent to the platform as an organizational model, but only to private platforms. The bottom line is that the current digital infrastructures are largely private, which is a crucial difference from the public and universal infrastructure of the 20th century (DIJCK; POELL; WAAL, 2018). By the same token, there are claims that platforms differ from traditional infrastructure in terms of the need for a smaller investment to generate network effects, dynamism, and geographic scope[45] (MONTERO; FINGER, 2017).

Practically, it is possible to identify three non-excludable ways in which platforms merge with infrastructure (PLANTIN; PUNATHAMBEKAR, 2019). The first, in terms of user experience, highlights the indispensability/criticality of certain platforms for participating in the digital economy. There are also reports of the practical impossibility of participating in the digital economy without depending on the services of the main U.S. (HILL, 2020) and Chinese platforms (PLANTIN; DE SETA, 2019). For instance, without the rail-hailing platform DiDi Chuxing, "it would be substantially difficult to obtain a taxi service nowadays" (CHEN; QIU, 2019, p. 277).

The second, at the level of traditional infrastructures, highlights how platforms are merging with traditional public utility providers. Chen and Qiu (2019, p. 280) outline DiDi Chuxing's trajectory of intertwining with public transport infrastructure in dozens of Chinese cities, making efforts to "datafy the urban transport ecosystem, including taxis, and occupy the centre of the converging networks of information, traffic, and transactions involving all kinds of vehicles and transport services". In this way, platforms become a mandatory "partner" for local authorities who wish to have access to their data in real-time. This strategy, for instance, "has enabled DiDi to become a central component in the datafied urban transport infrastructure". Other studies investigate the interpenetration of generations of infrastructure in the port (DI VAIO; VARRIALE, 2020), telecommunications (MONTERO; FINGER, 2017; MUKHERJEE, 2019; PLANTIN; DE SETA, 2019), international remittance

[45] McQuire (2019, p. 159) draws attention to the fact that traditional infrastructures had well-defined accountability and control mechanisms, which allowed public agencies to evaluate their performance in offering services within the expected quality and universality standards. The same has not happened so far with the new generation of infrastructuralized platforms.

(RODIMA-TAYLOR; GRIMES, 2019), cartography (MCQUIRE, 2019), education (DIJCK; POELL; WAAL, 2018; SELLAR; GULSON, 2021), energy (MONTERO; FINGER, 2017), scholarly communication (PLANTIN *et al*., 2018), health and media (DIJCK; POELL; WAAL, 2018) sectors.

Finally, the third way in which platforms approach infrastructures is through investments in infrastructure projects. Mosco (2017) presents a collection of initiatives in which Big Tech is involved, from investments in data centres to investments in submarine cables and its satellite constellations.

Given this multilevel phenomenon, researchers have resorted to different perspectives, usually developing insightful case studies. Nechushtai (2018) investigated the relationship between platforms and media companies, and she proposed the effect of the risk of infrastructural capture: when auditing entities must scrutinize entities that provide infrastructure services to them, things can go wrong. Chen and Qiu (2019), when analysing the infrastructural role of DiDi Chuxing in the urban transport sector in China, proposed the concept of digital utility to highlight how some platforms develop a strategy of becoming critical in the digital economy exploiting both regulatory loopholes and informal labour practices. They also distinguished traditional providers of public utilities from platform providers of digital utilities: the latter depends on the extraction of surplus data to become a central node in data networks.

Still, Plantin *et al*. (2018) have pointed out that platforms assume the role of infrastructures at a specific historical moment: when public bodies retreat from the responsibility of offering universal infrastructure, under the aegis of neoliberal ideology[46]. Montero and Finger (2017, p. 223) also stressed that "as network industries [infra] have been fragmented both vertically and horizontally by liberalization policies, platforms can play a role in enhancing coordination." In this respect, a recent report by the European Commission (BUSCH *et al*., 2021) highlights the links between the platforms' infrastructural role and the different types of power they leverage over users, markets, and governments. This range of power stances comes from their ability to position themselves as infrastructures. This

---

[46] Other authors support the interpretation that platforms become infrastructural due to the withdrawal of public power: "To understand how datafication, commodification, and selection tie in with contemporary governance strategies, it is especially important to see how in neoliberal or advanced liberal democracies, calculative regimes of accounting, and financial management have been employed to enable what Miller and Rose (2008, p. 212-13) call a "degovernmentalization of the state" (…) It is in this framework of calculative regimes and deregulation that platform datafication takes shape." (DIJCK; POELL; WAAL, 2018, p. 46); and the same is applied even in the State-led Chinese economy: "(…) in 2016, the Measures for Administration of Urban Taxi Business Operations and Services defined the taxi service as a "supplement" to public transport, which means that the business can be privately owned or operated" (CHEN; QIU, 2019, p. 278).

diagnostic leads almost naturally to the following inquiry: how to protect public values when private platforms dominate the infrastructures of society (DIJCK; POELL; WAAL, 2018; SRNICEK, 2017).

3.4 Lattes platform: a pioneer Brazilian initiative

The Lattes[47] Platform, an indigenous initiative, was officially launched in the late 1990s, in a decade when many governments worldwide were architecting their online presence as a result of the penetration of ICT-based technologies in both public and private sectors.

3.4.1 Brief history

The Lattes Platform was developed in a fruitful collaborative network formed by government agencies, universities, and the private sector. It was funded with public resources coordinated by CNPq, and other public organizations of the Brazilian Innovation System were also involved in the partnership, such as the Coordination for the Improvement of Higher Education Personnel[48], the Studies and Projects Financing Agency[49] and research groups from public universities (*Stela* from Federal University of Santa Catarina – UFSC, and *Cesar* from Federal University of Pernambuco – UFPE, which later became independent[50]). The private sector was also involved (Multisoft)[51] (CNPQ, 2019).

In the early 2000s, the expected budget for running the Lattes Platform programme amounted to BRL 15.08 million (about USD 7.36 million)[52] for just two years (about 0.30% of the total CNPq budget for both 2000 and 2001) (Figure 3.1).

---

[47] The name of the platform is a homage to a Brazilian physicist called Cesare Mansueto Giulio Lattes, who participated in experiments that proved the existence of the pi-meson – a composite subatomic particle – and encouraged the development of experimental physics in the country, including for instance the foundation of the Brazilian Centre for Research in Physics (*Centro Brasileiro de Pesquisas Físicas*, CBPF) (NASCIMENTO, 2015).

[48] *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, CAPES.

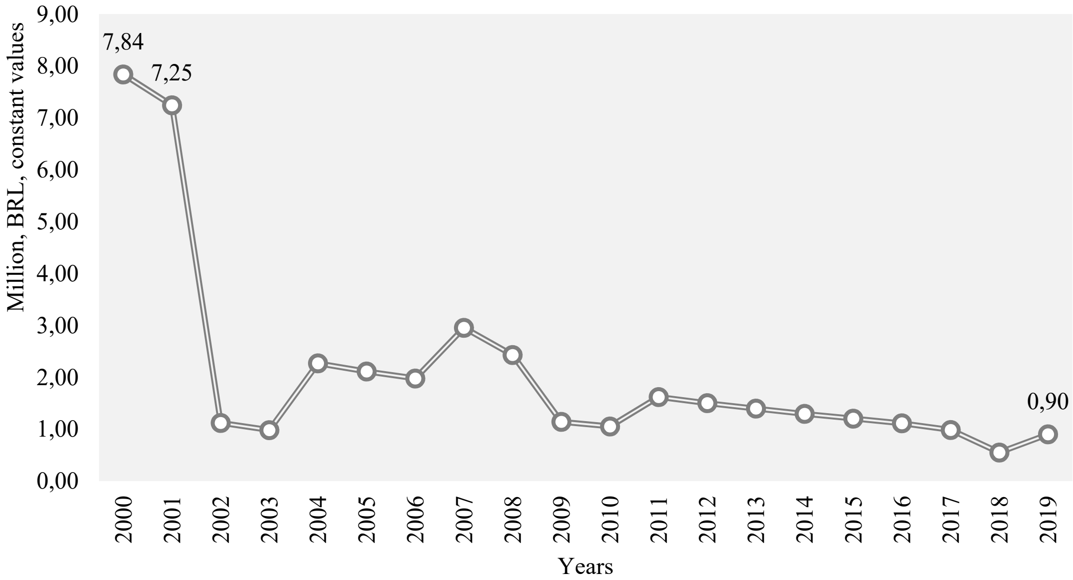
[49] *Financiadora de Estudos e Projetos,* FINEP.

[50] https://www.stela.org.br/ and https://www.cesar.org.br/

[51] https://multisoftwares.com.br/

[52] Constant BRL values were converted into USD using the official exchange rate. It was calculated as an annual average based on monthly averages (local currency units relative to the USD). Data was sourced from the World Bank.

Figure 3.1 – Evolution of Lattes platform budget



Source: The authors' own. Data sourced from *Ministério da Economia – Planejamento, Desenvolvimento e Gestão* (https://www.siop.planejamento.gov.br). Note: Current BRL values were deflated using the GDP deflator (year base 2019) from the World Bank database. (*) Annual Budget Law (*Lei Orçamentária Anual*, LOA) estimates revenues and establishes the estimate of how much the Brazilian Federal Government is authorized by Congress to spend throughout the year. As it is an estimate, LOA can be amended by additional credit bills once Congress approves provisional measures proposed by the president. Government expenditure can therefore vary from what was proposed by LOA, however, we opt to use it, as expenditure estimation. For all years the "Budget Unit" (*Unidade Orçamentária*) considered was *Conselho Nacional de Desenvolvimento Científico e Tecnológico*, CNPq (code number 24201). From 2000 to 2012 we used data from "*Ação Orçamentária 4208 – Sistema Integrado de Informação em Ciência e Tecnologia (Plataforma Lattes)*"; from 2013 to 2018, "*Plano Orçamentário 000A – Sistema Integrado de Informações em Ciência e Tecnologia (Plataforma Lattes)*; and for 2019, "*Plano Orçamentário 000A – Sistemas Integrados de Informações em Ciência, Tecnologia & Inovação (Plataformas Digitais)*", considering all types of expenditure (*Grupo de Despesa: corrente e investimento*).

The main goal with the Lattes Platform was the standardization of Brazilian researchers' curricula vitae. This standardization aimed at constructing a database, making it possible to find specialists and provide statistics on the distribution of scientific research countrywide. Since it was launched, the Lattes Platform has been increasing its scope and its base of users. The Lattes Platform is now being used by most universities and research institutes, and foundations that support scientific and technological activities in the country. According to Günther *et al*. (2020), science, technology and innovation (ST&I) institutions and funding agencies have informational and interest links with the CV-Lattes, one of the Lattes Platform components. In a recent report issued by CNPq, it was stated that:

> The Lattes Platform represents CNPq's experience in database integration of Curricula Vitae, Research Groups, and Institutions in a single system of information. Not only does its current dimension extend to planning and management actions of CNPq, but also to other federal and state agencies, such as state foundations to support science and technology, in addition to higher education institutions and research institutes. Besides that, it has also become strategic for the formulation of policies by the Ministry of Science, Technology, and Innovations (MCTI) and other government agencies in the field of science, technology, and innovation. (CNPQ, 2020, p. 17).

Implicit in the above quote is how network effects have operated to consolidate the Lattes Platform. As more scholars use the platform to publicize their expertise and the results of their research, it becomes more important for all entities in the science and technology (S&T) system to use the same database. It was undoubtedly a disruptive innovation in terms of gathering and organizing information of the Brazilian scientific communities (PACHECO, 2003). It was so successful that the Lattes Platform was licensed through bilateral agreements mainly within the scope of ScienTI[53], and its technology and methodology were adopted and/or adapted such as for the *Sistema de Información de Ciencia y Tecnología Argentina* (SICYTAR) of the *Ministerio de Ciencia, Tecnología e Innovación* (TORRES, 2019), the CvLAC Peru of the *Consejo Nacional de Ciencia, Tecnología e Innovación* (CONCYTEC); the CvLAC Ecuador of *Secretaría de Educación Superior, Ciencia, Tecnología e Innovación* (SENESCYT); the CvLAC Directory of Venezuela of the *Observatorio Nacional de Ciencia, Tecnología e Innovación* (ONCTI); and the CVuy of the *Agencia Nacional de Investigación e Innovación* (ANII) from Uruguay (D'ONOFRIO, 2009).

In 2005, the CNPq nominated a commission to evaluate, reformulate and improve Lattes (*Comissão de Gestão do Lattes* – COMLATTES), however, it seems it was focused more on correcting deviations and designing incremental improvements[54]. This can also be reflected in the CNPq's constant budget reductions (Figure 3.1). After the Lattes Programme was launched, there were drastic budget reductions, and this worsened especially from 2016, probably as a result of the approval of a Constitutional Amendment (95A), capping public spending for 20 years (ROSSI; DWECK, 2016). Its goal was to control public debt and regain

---

[53] The International Network of Information and Knowledge Sources for Science, Technology and Innovation Management (ScienTi) provided a public and cooperative forum for stakeholders of national ST&I systems and communities of the member-countries (Argentina, Brazil, Chile, Colombia, Cuba, Ecuador, Mexico, Panama, Paraguay, Peru, Portugal, and Venezuela) to interact and it was designed to have an updated identification of human resources, institutions, and research projects for development and assessment of national policies and capacities in ST&I (http://www.scienti.net/).

[54] Most of the marginal improvements focused on the inclusion (or exclusion) of more adequate information from scholars, as is demonstrated in *Relatórios de Gestão*. For example, in 2012 CNPq launched its new Lattes Platform with new functionalities in CV-Lattes, where scholars could register information on innovation, education, and the popularization of science and technology and patent records, and in 2019 maternity leave was also included in CV-Lattes.

the confidence of foreign investors by capping the federal budget at the expenditure of the previous year, adjusted for inflation (CHIARINI *et al*., 2020).

There were attempts in 2018 to organize COMLATTES in two different committees (management and technical). One of the attributions of the technical committee was the proposition of new technologies for the solutions and evolution of the Lattes Platform and the conducting of studies, benchmarking, and prospect solutions for its technological innovations[55]. However, in 2019 the two-committee division was undone.

With the ambition of promoting a nationwide discussion on the evaluation of ST&I policies using data from the Lattes Platform, the CNPq, and the Centre for Strategic Studies and Management[56] – a thinktank organization supervised by the Brazilian Ministry of Science, Technology, and Innovation – organized the “First ST&I Policy Evaluation Seminar”[57]. Besides examples of applied research using the databases (CALIARI; RAPINI; CHIARINI, 2018; DUARTE; WEBER; PACHECO, 2018; GUIDINI *et al*., 2018; LEITE *et al*., 2018; ROSSI; DAMACENO; MENA-CHALCO, 2018), the seminar fostered the inventive capacity of Brazilians in a special section called “EXPOLattes” where inventors who proposed technological solutions for data mirroring and data extraction could present their gadgets. Besides the presentation of e-Lattes and Intelligentia Lattes, two other softwares were presented by a start-up (Indeorum[58]) based at *Universidade Federal de Pelotas* (UFPel): Ranquium and Cientum.

COMLATTES demonstrated its intention of inviting “EXPOLattes” winners to present their solutions to CNPq and discuss ways of internalizing their technologies. This is part of a recent strategy to find ways to develop partnerships to make the Lattes Platform sustainable in the long run. Consequently, COMLATTES proposed verifying the possibility of monetizing for the use of Lattes Platform data[59]. This “strategy” may be a result of budget reductions, as presented in Figure 3.3. Consequently, some pressures force the Lattes Platform’s maintainers to find monetizing possibilities. However, it may also reflect a “government as a platform” perspective (O’REILLY, 2010). In a few words, it means that open data from the government is offered publicly for recombination aiming at generating new products and services (O’REILLY, 2010).

---

[55] *Resolução Normativa* 025/2018 (CNPq).
[56] *Centro de Gestão e Estudos Estratégicos*, CGEE.
[57] *I Seminário de Avaliação de Políticas de CT&I.*
[58] https://indeorum.com/
[59] This information was obtained under the rule of Law No. 12,527 called "Brazilian Access to Information Law", which provides access to public information. Under Protocol Number 01217.001918/2021-36, we had access to all meeting minutes organized by COMLATTES from July 2019 to October 2020. CNPq did not provide the meeting records from 2005 to 2019.

Opening up Lattes data creates value; however, if the majority of them are gathered outside CNPq, there is a risk in "(…) turning vital public resources into proprietary assets". Dijck *et al*. (2018, p. 154) highlight this issue in a very clear way: "businesses can profit from open data produced by public institutions, while data and knowledge generated by users but processed by corporations become proprietary. (…) Once again, "open" and "public" are not the same things".

3.4.2 Lattes components

The Lattes Platform is divided into three sub platforms: a) Directory of Institutions and Research Infrastructure (*Diretório de Instituições e Infraestruturas de Pesquisa,* DIIP); b) Directory of Research Groups (*Diretório de Grupos de Pesquisa,* DGP); and c) Lattes Curricula (CV-Lattes) (Table 3.2).

Table 3.2 – Lattes Platform and its sub platforms

	Sub platforms	**Acronym**	**Objectives**
Lattes Platform	Directory of Institutions and Research Infrastructure	DIIP	• monitor and evaluate public policies to promote the national S&T infrastructure, constituting an information system about organizations and their research infrastructures.
	Directory of Research Groups	DGP	• allow information exchange among researchers, • monitor and evaluate public policies, constituting an information system about S&T production and interactions among research groups and the productive sector.
	Lattes Curricula	CV-Lattes	• allow scholars and researchers to create a professional profile adding information about their scientific achievements, • monitor and evaluate public policies, constituting an information system about S&T production and investment allocation.

Source: Authors' own. Information sourced from http://lattes.cnpq.br/

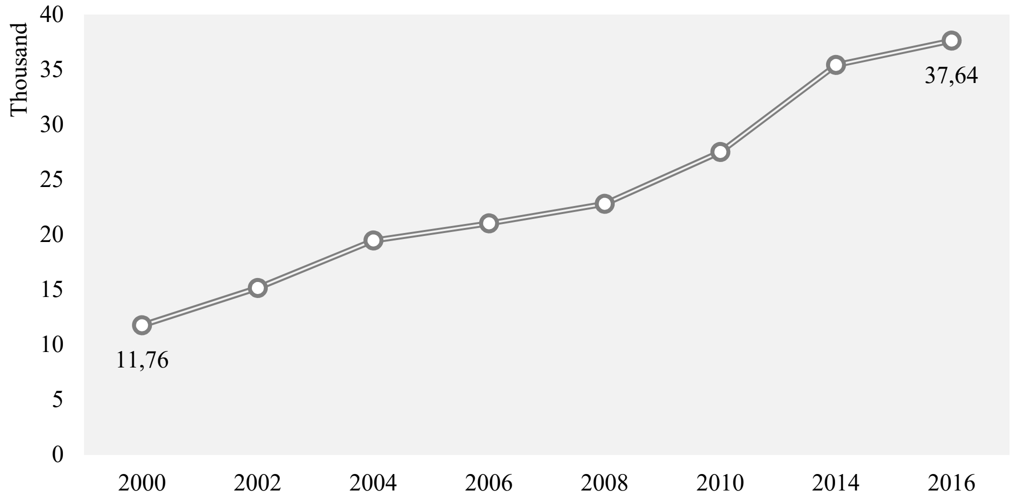
The DIIP was designed to gather information on research infrastructure available in Brazilian institutions and to provide info for comprehensive diagnosis of their conditions[60]. One of its main objectives was to build a dynamic database that allows systematic monitoring and the production of periodic reports on the national research infrastructure; however, the

[60] http://lattes.cnpq.br/web/diip/

information was gathered only in 2012 and only a report was publicly available (NEGRI *et al*., 2013) and a few studies used the dataset (CALIARI; RAPINI; CHIARINI, 2020; NEGRI; SQUEFF, 2016). It is noteworthy that DIIP gathers information about 1,760 scientific infrastructures including laboratories; research ships or floating labs; and plants or pilot plants.

The DGP was developed to gather information about active research groups.[61] and has an increasing registration number, reaching more than 37 thousand groups spread nationwide (Figure 3.2). It includes data about researchers, students, technicians, research streams in progress, and the scientific, technological, and artistic production generated by each group. It automatically provided a few aggregated metrics regarding research groups' characteristics and scientific and technological production; however, it was discontinued in 2010. Notwithstanding that, it still allows XML microdata extraction[62] from the 2000 to 2016 Census. DGP data has been increasing over the years (from a little more than 4 thousand groups in 1993 to over 37 thousand in 2016), and it can be assumed that it is representative of the national scientific community. Many studies build on DGP microdata to describe university-industry interactions through different perspectives (FERNANDES *et al*., 2010; GARCIA; RAPINI; CÁRIO, 2018; RAPINI, 2007; RAPINI *et al*., 2019; SUZIGAN *et al*., 2009; SUZIGAN; ALBUQUERQUE; CARIO, 2011).

Figure 3.2 – Number of research groups registered at DRG

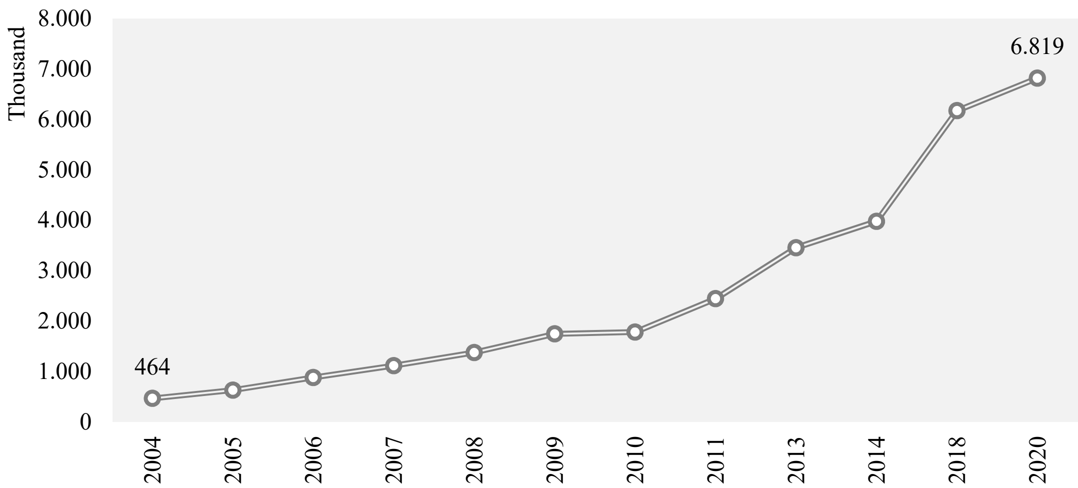


Source: The authors' own. Data sourced from CNPq (http://lattes.cnpq.br/web/dgp/). Note: The last census was done in 2016

[61] http://lattes.cnpq.br/web/dgp
[62] http://lattes.cnpq.br/web/dgp/censos2

Finally, the Lattes Curricula sub-platform gathers public information individually available for each scholar with information about his/her lines of research, projects, and scientific and technological production. Registration with CV-Lattes is voluntary and spontaneous, although scholars are highly stimulated to participate, mainly because having updated curricula is a precondition for accessing public funding and scientific research. Probably, CV-Lattes is the most used sub-platform provided by Lattes and we can define it as an outlet of scholarly work, as checking one's CV-Lattes is a current practice among scholars. It gathers now over 6 million résumés (Figure 3.3).

Tabela 3.3 – Number of scholars registered at CV-Lattes



Source: The authors' own. Data sourced from *Relatórios de Gestão* (https://www.gov.br/cnpq/pt-br/acesso-a-informacao/auditorias/prestacao-de-contas). Note: for the Years 2012, 2015–2017, and 2019, information about updated numbers is not available publicly

Lattes Curricula does not provide automatically reports and metrics of scientific productions; however, it allows the free use of the "*Lattes Extractor*"[63] tool to mine data. Many inventors have also developed software for extracting information and big data mining from CV-Lattes as registered at the Brazilian National Institute of Industrial Property[64] using different languages like Java, Python, and R. There is also open-source software like ScriptLattes[65] – designed and implemented to create academic reports based on Lattes Database (MENA-CHALCO; JUNIOR, 2009) – and e-Lattes[66] – which uses the R language (SAMPAIO, 2018).

63 http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados
64 *Instituto Nacional de Propriedade Industrial*, INPI.
65 http://scriptlattes.sourceforge.net/
66 https://elattes.com.br/

With the development of extractors, it was possible to use Lattes Database to draw a bigger picture of the Brazilian Innovation System. Examples are snapshots of the scientific production (DIAS; MOITA, 2018), scientific regional distribution (SIDONE; HADDAD; MENA-CHALCO, 2016, 2017), knowledge field particularities (ARRUDA *et al*., 2009; HILÁRIO; GRÁCIO, 2017; MCMANUS; BAETA NEVES, 2020), scientific collaboration and coauthorship networks (FONSECA; FERNANDES; FONSECA, 2017; MENA-CHALCO *et al*., 2014), gender inequality in scientific areas (OLIVEIRA; MELLO; RIGOLIN, 2020; PERLIN *et al*., 2017; SANTIAGO; AFFONSO; DIAS, 2020), academic genealogy studies (DAMACENO *et al*., 2019; ROSSI; DAMACENO; MENA-CHALCO, 2018), and other scientometric and network analysis such as the research output in response to the Zika Crisis (SAMPAIO *et al*., 2020) and tuberculosis (FONSECA *et al*., 2017), development of methods to assess researcher quality-oriented to specific purposes (DUARTE; WEBER; PACHECO, 2018), trust analysis in cooperation in research (ADAMATTI; CASTELFRANCHI, 2015) and predatory publishing (PERLIN; IMASATO; BORENSTEIN, 2018).

The Lattes Platform provides valuable data. The studies cited above highlight that Lattes provides value through open, comprehensive, high-quality, and structured data. These data are vital for science policy too as it is a process with many sources of uncertainty. As precisely mentioned by Thomas and Mohrman (2011, p. 261),

> There are endless options for investment in science, each carrying varying levels and kinds of potential payoff and risk. Investment decisions and science policies are themselves hypotheses. They are based on incomplete knowledge of what avenues of science exploration are likely to yield useful knowledge, what areas of science will be adequately funded without policy intervention, and what dynamics will result from a policy and how they will impact science production and the linking of science to the larger innovation and mission system outcomes.

As a consequence, policymakers must deal with a very complex system and they have to combine their insights, ideologies, values, and judgements with data and analytics to guide the strategic and tactical decisions on science policy. Gathering and analysing data creates pieces of evidence to support policy options and their opportunities for impact (THOMAS; MOHRMAN, 2011). Therefore, the creation of databases and the development of analytical and interpretive tools and methodologies are required to “support a robust, evidence-based science policy decision process” (THOMAS; MOHRMAN, 2011, p. 268).

The Lattes Platform gathers substantial information on the stock of knowledge and, to a certain extent, on its flows regarding talent and capabilities (provided at CV-Lattes for example). Notwithstanding that, Lattes data is also an input to identify opportunities for any investment decision.

The questions, however, are: who will appropriate the value emerging from the Lattes Platform's data? Is the open data strategy pursued enough to make the Lattes Platform catch up with the new generation of scientific platforms?

3.5 Lattes Platform multi-layered analysis

Manca's (2017) contribution was in proposing a multilayer approach for analysing ASN according to a networked socio-technical perspective into three levels: networked-scholar, techno-cultural and socio-economic. Notwithstanding that, we believe it is possible to improve her multilayers to analyse other scientific digital platforms such as the state-mandated Lattes Platform. The micro-level – the network-scholar layer – refers to affordances (MAJCHRZAK; MARKUS, 2013), which are the action potential within reach of the platform user; the meso level – techno-cultural layer – concerns the platform controller's decisions about its dynamics, *i.e.*, ability to integrate open innovation, etc.; finally, the macro-level – socio-economic layer – contextualizes the entire platform ecosystem (users + platform) in terms of infrastructure characteristics and economic dynamics. In what follows, we apply an extended version of Manca's (2017) framework with new categories, starting from the micro level and going to the macro level.

3.5.1 Networked-scholar layer

a) Networking components

The Lattes Platform was established before the full development of technologies used in the most advanced ASN such as ResearchGate. Some scientific platforms whose primary focus was on posting and sharing academic-related content were able, subsequently, to add social networking capabilities (such as Mendeley) (JORDAN, 2019). That was not the case with the Lattes Platform so far: it has not included any network of feedback artifacts. The sub-platform DGP, for example, contained at its core a proposal for the visualization and formation of research networks, aiming…

> to boost the creation of knowledge and the innovation process resulting from the *exchange of information* and, above all, from the junction of competencies of groups that join efforts in the search for common goals, with or without sharing facilities. *Not to be confused with ASN* – which are not the target of DGP. These aim to meet the need for a specific environment that is specialized in a particular area of knowledge, with possibilities for discussions, networking, purchase of inputs, dissemination, and marketing of research results and innovations.[67] (CNPq, [s.d.]-a, emphasis added).

In recent years, DGP has started to have many of its functionalities disabled, keeping only a history of data from research groups on its website for a consultation. It is possible though to use a parameterized query tool[68] to find research groups, being the way to allow users to monitor other research groups' activities. There is no other way to foster interaction with peers. In a public report, CNPq affirms that DGP was…

> (…) defined as an inventory of active scientific and technological research groups in the country. Today, the DGP, despite some problems it has been passing through, continues to configure itself as an *instrument for information sharing and exchange*, allowing, accurately and quickly, to *identify and locate expertise* and their current occupation, in addition to recent production, such as the source of information for both scientific societies and the various instances of the political and administrative organization of the country, allowing the analysis of the state-of-the-art of Brazilian Science and Technology and helping to preserve the memory of the national scientific and technological activity. (CNPq, 2020, p. 17, emphasis added)[69].

b) Knowledge sharing

Concerning knowledge sharing, the Lattes Platform, especially the CV-Lattes sub-platform, has since its inception adopted an open interface, in which users fill in relevant information about their scientific and academic activities. If, on the one hand, this reduces the costs of development and maintenance of the platform (and perhaps allows the insertion of idiosyncratic information), the lack of standardization and the total autonomy of the insertion of information hinder the recovery of this information. The exchange of knowledge is also

---

[67] In the original: "visam impulsionar a criação do conhecimento e o processo de inovação resultantes do intercâmbio de informações e, sobretudo, da junção de competências de grupos que unem esforços na busca de metas comuns, podendo ou não haver compartilhamento de instalações. Não confundir com redes sociais de pesquisa que não são o alvo do DGP. Estas visam suprir a necessidade de um ambiente próprio e especializado em uma determinada área do conhecimento, com possibilidades de discussões, networking, compras de insumos, divulgação e comercialização dos resultados de pesquisas e inovações."

[68] http://dgp.cnpq.br/dgp/faces/consulta/consulta_parametrizada.jsf

[69] In the original: "é definido como "inventário dos grupos de pesquisa científica e tecnológica em atividade no País". Hoje, o DGP, não obstante alguns problemas porque tem passado, continua se configurando como instrumento para o intercâmbio e a troca de informações, permitindo, com precisão e rapidez, identificar e localizar expertises e sua ocupação atual, além da produção recente, como fonte de informação tanto para sociedades científicas como as várias instâncias de organização político-administrativa do país, permitindo a análise do estado da Ciência e Tecnologia brasileiras e ajudando a preservação da memória da atividade científico-tecnológica nacional."

affected by the absence of "a controlled vocabulary or a hierarchical tree of terms" (SILVA; SMIT, 2009, p. 86).

c) Identity

Concerning identity formation, as in ASN, the Lattes Platform allows users to shape their profiles, including a headline where scholars present their expertise and skills and they can present themselves; consequently, user identity is mostly conveyed through the profile. However, according to Silva and Smit (2009), the complete autonomy for generating profiles by such a heterogeneous group of users, more or less versed in communication and information tools, leads to a very large dispersion of profiles. Both CV-Lattes and DGP are user-generated contents (UGC) with low standardization (pre-determined categories for metadata).

CV-Lattes and DGP provide indicators of researchers' scientific and technical production, but they are simple indicators that seem to contribute little to the formation of a user's "reputation and trust as elements that shape academic personae" (MANCA, 2017, p. 03), even if there are displays of research results and impacts (CV-Lattes proposes publication metrics based on Web of Science, Scopus and Scielo and JCR impact factor for indexed journals).

3.5.2 Techno-cultural layer

a) User/usage and content

As far as the techno-cultural level is concerned, CNPq implements features to spur users' connectivity based on the stimuli to join all sub platforms. For instance, having updated information in CV-Lattes is a precondition for accessing public funding and scientific research. CV-Lattes' homepage provides an outlet that allows users to monitor other scholars' activities (main researches, publications, filiation, etc.).

Manca (2017, p. 3) presents the technology of a platform as "services that help encode activities into a computational architecture that steers user behavior". In this sense, CV-Lattes users have the option of including its ORCID ID through an application programming interface (API). Although the Lattes Platform created its identification number many years before, the inclusion of ORCID ID indicates its consolidation in the Brazilian scientific community and its increase in popularity. Bello and Galindo-Rueda (2020) demonstrated how ORCID ID has taken the lead[70] as the most pervasive standard for the unique and digital identification of researchers in the world. Concerning the interface, CV-Lattes users can register information in seven modules: General data, Education, Performance, Projects, Production, Education, and Popularization of S&T and Events. Each of these modules also opens up other different categories, for example, the production module unfolds into bibliographic, technical, and artistic/cultural production.

This user-generated content (UGC) is also related to what Manca (2017) calls "content": standards of content and delivery of products. There are three levels of user autonomy when filling fields in the platform: with full autonomy, partial autonomy, or without autonomy. As most fields have total or partial autonomy, the system "allows the filling of fields to be carried out at the mercy of the perception that feeder users have of its operation or of the objectives pursued when filling out the Lattes Platform" (SILVA; SMIT, 2009, p. 85). This opening allowed by the interface leads to lower economic costs, but to greater difficulty in standardizing and retrieving information in the system, as previously mentioned.

b) Technological components and open innovation mechanisms

Technological components present in the Lattes Platform, despite marginal improvements, maintain the main architecture used when it was developed: XML (extendable markup language) and three sets of information sources: a) operational bases; b) warehouse bases; and c) web information repositories. The operational bases are structured (relational) bases resulting from transactions with the different users of the Platform. Warehouse bases include both text bases generated from operational bases (for indexing and searching) such as the data marts produced for the different information units and their domains (PACHECO, 2003; PACHECO; KERN, 2001, 2003)

It was argued that the Lattes Platform's maintainers allow third parties to recombine data aiming at generating new products and services. Not only did the open innovation mechanism result in open source software such as ScriptLattes (MENA-

[70] The "Open Researcher and Contributor Identifier" (ORCID), promoted by the namesake international non-profit organization, appears to have become the prevailing global standard as it is the most disseminated type of identifier used by scientific authors worldwide (more than 60 percent)" (BELLO; GALINDO-RUEDA, 2020, p. 29).

CHALCO; JUNIOR, 2009) and e-Lattes (SAMPAIO, 2018), but also in new technologies filed at INPI (Table 3.3).

Table 3.3 – Application filed of computer software at Brazilian National Institute of Industrial Property (INT) related to Lattes Platform

Code Number	Year of Deposit	Title	Owner	Main Language	Type of Program
BR 51 5051 001146 5	2021	LLattes ProdPPT – Gerador de Relatórios Padrão Lattes da Produção Técnica	Márcio Carneiro dos Santos	Python, XML	DS07, GI04
BR 51 2021 001145 7	2021	LLattesPBiblio – Gerador de Relatórios Padrão Lattes da Produção Bibliográfica	Márcio Carneiro dos Santos	Python, XML	DS07, GI04
BR 51 2020 002612 5	2020	Extract Lattes	Universidade Estadual de Montes Claros (UNIMONTES)	Django, Python	AP02, AV01, GI04, IA01
BR 51 2020 002538 2	2020	Extrator Lattes	Fabiano Peruzzo Schwartz	R	AP03, AV01, GI04
BR 51 2017 001622 4	2017	Intelligentia Lattes Extractor	Instituto Stela	Java	CD01, GI01, GI06, GI08, SO07
BR 51 2017 001621 6	2017	Intelligentia Lattes Annotator	Instituto Stela	Java	CD01, GI01, GI06, GI08
BR 51 2017 001619 4	2017	Intelligentia Lattes Cube	Instituto Stela	Java	CD01, GI01, GI06, GI08
BR 51 2017 001618 6	2017	Intelligentia Lattes Intelectus	Instituto Stela	Java	CD01, GI01, GI06, GI08
BR 51 2017 001617 8	2017	Intelligentia Lattes Viewer	Instituto Stela	Java	CD01, GI01, GI06, GI08
BR 51 2017 001616 0	2017	Intelligentia Lattes Service	Instituto Stela	Java	CD01, GI01, GI06, GI08
BR 51 2016 001291 9	2016	SILQ – Sistema de integração Lattes-Qualis	Universidade Federal de Santa Catariana (UFSC)	Java	GI01, GI04, GI08, SO02
BR 51 2014 000516 0	2014	NILREP	Universidade Federal do Rio Grande do Sul (UFRGS)	Python	GI04
10918-5	2010	SISLattes – Sistema Extrator Lattes	Universidade Estadual do Oeste do Paraná (UNIOESTE)	Java	GI01, GI08

Source: Authors' Own. Data sourced from Brazilian National Institute of Industrial Property (INPI). Search String: "Plataforma Lattes" and "Lattes". Note: AP02 = Planning; AP03 = Controlling; AV01 = Performance evaluation; CD01 = Data communication; DS07 = Documentation support; GI01 = Information management; GI04 = Report generator; GI06 = Data entry and validation; GI08 = Data recovery; IA01 = Artificial intelligence; SO02 = Input and output interface; and, SO07 = Process controlling

3.5.3 Socio-economic layer

a) Ownership and governance

The Lattes Platform is a non-profit scientific digital platform belonging to CNPq. The governance component is mostly managed through the Terms of Service – ToS (*Termo de Adesão e de Condições de Uso*), which was defined after 2004. The term stipulates that, in what regards CV-Lattes, CNPq collects and stores users' personal information for supporting policy making and evaluation, sharing data with third parties except for personal data:

> CNPq, through CV-Lattes, collects and stores curricular information from users, necessary to fulfil its institutional mission: To promote and foster the country's scientific and technological development and contribute to the formulation of national S&T policies. Such information is used to assess the competence of candidates to obtain scholarships and grants; in the selection of consultants, committee members, and advisory groups; in supporting the evaluation of Brazilian research and graduate studies, and in the construction of other databases that support the drafting of indicators and studies of interest to ST&I (…). All curricular information sent to CNPq may be made available for internal access or displayed on the Agency's internal network. They may also be disclosed to the external public, through the Internet or other means, except for the following information regarding user's identification data, by which CNPq undertakes not to publicly disclose it: a) residential address; b) home telephone; c) affiliation; d) year of birth; e) CPF; f) gender; g) colour or race; h) identity; i) passport and j) email addresses. [71] (CNPQ, [s.d.])

Lattes' Terms and Conditions seem to be partially following the Brazilian General Personal Data Protection Law (*Lei Geral de Proteção de Dados* – LGPD) (BRASIL, 2018) which states that the processing of personal data can be carried out by the public administration for the processing and shared use of data necessary for the execution of public policies and by research institutes to carry out studies, ensuring, whenever possible, the anonymization of personal data[72]. However, as Lattes' data "may also be disclosed to the

---

[71] In the original: "*O CNPq, através do Sistema de Currículos Lattes, coleta e armazena informações curriculares dos usuários, necessárias ao cumprimento de sua missão institucional: Promover e fomentar o desenvolvimento científico e tecnológico do país e contribuir na formulação das políticas nacionais de C&T. Tais informações são utilizadas na avaliação da competência de candidatos à obtenção de bolsas e auxílios; na seleção de consultores, de membros de comitês e de grupos de assessoramento; no subsídio à avaliação da pesquisa e da pós-graduação brasileiras e na construção de outras bases de dados que subsidiam a elaboração de indicadores e estudos de interesse da CT&I. (...) Todas as informações curriculares enviadas ao CNPq poderão ser por este disponibilizadas para acesso interno ou exibidas na rede interna da Agência. Poderão também ser divulgadas para o público externo, através da Internet ou de outros meios, exceto as seguintes informações relativas aos dados de identificação do usuário, pelas quais o CNPq se compromete à sua não divulgação pública: a) endereço residencial; b) telefone residencial; c) filiação; d) ano de nascimento; e) CPF; f) sexo; g) cor ou raça; h) identidade; i) passaporte e j) endereços eletrônicos.*"

[72] Law n. 13.709/2018, Article 7, items III and IV.

external public, through the Internet or other means", how can data holders avoid improper use of their data? According to LGPD, the holder has the right to easily access the information about the processing of their data, which must be made available in a clear, adequate, and ostensible way about, for example, the specific purpose of the data processing and responsibilities of the agents who will carry out that processing[73].This information is not available at Lattes' ToS.

In addition, the Lattes Platform does not comply with a recently published Bill that provides principles, rules, and instruments for Digital Government (BRASIL, 2021), which states that e-gov platforms must have transparency and control tools for the processing of personal data that are clear and easily accessible and that allow the citizens to exercise their privacy right.

b) Business model and market scope

The business model is based on public funding, as the Lattes Platform offers free-of-charge services and there are no other supplementary services. Even if the Platform was freely licensed to other countries in the scope of ScienTI Network in the early 2000s and most recently CNPq's Information Technology Master Plan (*Plano Diretor de TI*) the expansion of its use by other countries was plotted in a SWOT Matrix[74] as an opportunity (CNPQ, 2014a). There is no public evidence demonstrating that CNPq has a well-developed strategy to internationalize the artifact, as in its most recent Strategic Plan there is no mention about the Lattes Platform at all (CNPQ, 2014b). Its market scope is national.

c) Economic effects

The value generation dynamic takes place in the platform's interaction with its user community and third parties, that is, with its ecosystem. The mechanism that defines the generation of value through this interaction is network effects. Network effects are characterized by the increase in the value perceived by the user as new users join the network/platform in question. These effects occur both directly and indirectly. In the case of the CV-Lattes, the direct network effect occurs when scholars understand that the entire research community uses the Lattes portfolio as a standard to register their scientific and

[73] Law n. 13.709/2018, Article 9.
[74] *SWOT* stands for Strengths, Weaknesses, Opportunities, and Threats.

academic activities. This effect is based on the very essence of the research community: by sharing values, norms, and practices, the community spreads certain conventions over time. Following these conventions creates value for network participants as they are shared by all other members. While there were few researchers with their CV registered on the Lattes Platform, this practice did not seem to have been institutionalized by the community; today, with over six million résumés registered (Figure 3.3), there has been a "naturalization" of the Lattes Platform as the unquestionable mechanism of the individual research portfolio. In other words, as the number of subscribers to the platform grew, the greater the importance that new members gave to this practice.

The network effect is also present indirectly. As more researchers use the platform to register their portfolios, it becomes more important for S&T institutions to use the Lattes Platform as the source of information about researchers for granting scholarships, offering jobs, etc. The same is true for researchers: the more institutions requesting the CV-Lattes, the greater the value associated with having an up-to-date profile.

While the Lattes Platform has been successful in leveraging network effects, the same cannot be said about feedback effects. The success of machine learning techniques over the last decade has enabled a new way to improve products and services. Services offered via ML-based systems collect user data and digital traces of their use and enhance the service itself based on these inputs. According to Mayer-Schönberger and Ramge (2018, p. 127) "it feels strangely alchemistic: turning a by-product of usage into the raw material of improvement, like converting lead into gold". Feedback effects are at the heart of a new generation of science platforms, as we can see from the excerpt below, taken from the ResearchGate terms of service (ToS)[75]:

> The Service also provides you with functionalities that support your scientific work, your professional life, and your development. To be as helpful as possible, the Service takes information about you as a Member, your Member Submissions, and your activity on the Service into account in emerging content and providing other aspects of the Service. This way we can make recommendations for connections, content, and features that may be useful to you. Keeping your profile information accurate and up-to-date helps us to make these recommendations more accurate and relevant.

The improvement of services based on feedback effects appears as a new dynamic of digital innovation: the transformation of the service occurring almost automatically and at increasingly lower costs as the pool of captured data grows *pari passu* with the increase of the

[75] https://www.researchgate.net/terms-of-service#Scope-of-the-Service

user base (MAYER-SCHÖNBERGER; RAMGE, 2018). Although the Lattes Platform can make use of user data internally, it does not mobilize this data to improve its service.

d) Infrastructure

Regarding the infrastructural role played by the Lattes Platform, according to Günther *et al*. (2020, p. 111), "Brazilian graduate programmes make use of CV-Lattes to mediate their bureaucratic processes. It is an information flow channel that enables social relations, as it organizes bureaucracy, standardizes information, and fosters a collaborative scientific network that uses its information in an operational and strategic way." According to Ramos *et al*. (2017), the Lattes platform is a strategic tool for the formulation of public policies; in addition, the CV-Lattes component is used as an evaluation criterion by "institutions such as Capes and other research bodies; selection of students and candidates for scholarships and grants; evaluation of courses in Postgraduate; selection of faculty in selection processes; monitoring of national scientific and technological production, among other attributions of the kind. Finally, to Silveira (2020, p. 1) "a scientist in the country who does not have his or her vitae in it practically does not exist for peers and funding agencies".

Two current examples demonstrate how the Lattes Platform has become the standard for the country's S&T agencies. For instance, the CNPQ Call 05/2021 – which provides granting for Senior Researchers – requires candidates to have their résumés registered in the Lattes Platform and they must be updated until the deadline for proposal submission. The public notice released by FAPESP, another S&T funding agency, establishes a calendar for submissions of proposals from Research, Innovation, and Dissemination Centres and declares that "CV-Lattes [of the proponents] must be up to date" (FAPESP, 2021a).

The above evidence strengthens the idea of the criticality of the Lattes Platform in the Brazilian S&T system, a trait of infrastructure platforms. Another way to capture its criticality is through the public's interests, concerns, or intentions on the Lattes Platform over time, using "Google Trends" tools[76], which provides data on Google Search usage[77].

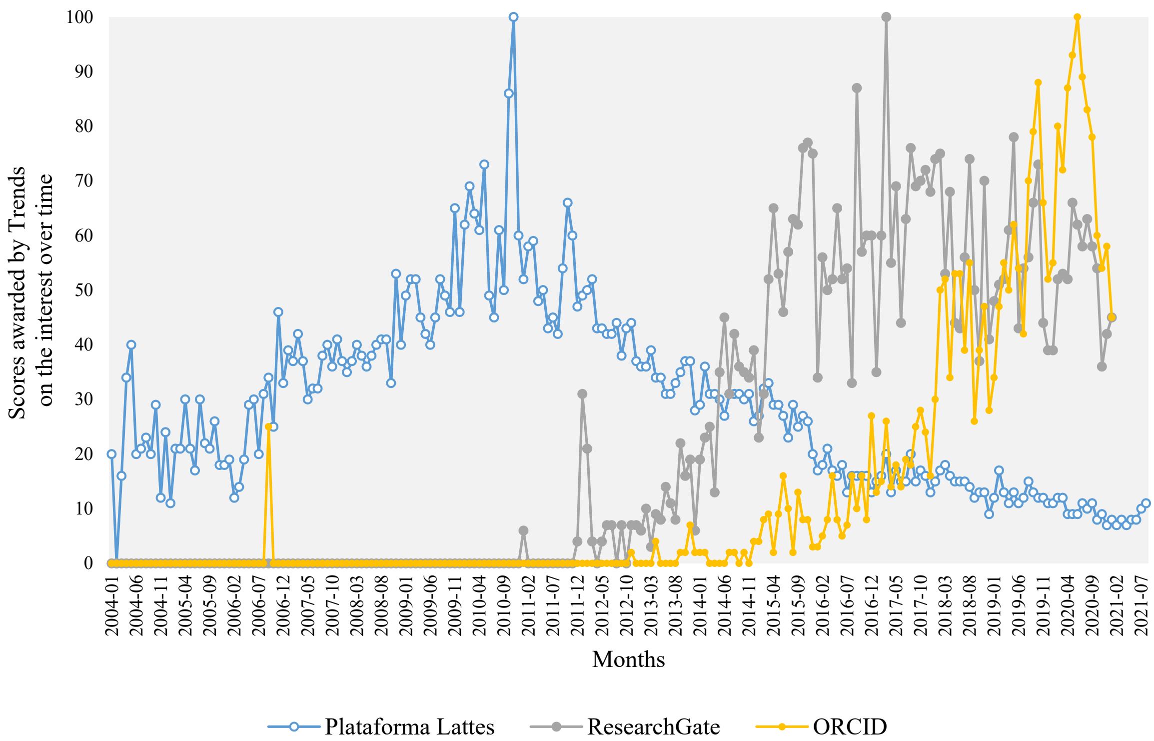
---

[76] For the limitations to the use of Google Trends, see Jun, Yoo, e Choi (2018).

[77] "Google Trends is useful to promptly detect a certain phenomenon and is, therefore, an excellent monitoring tool" (JUN; YOO; CHOI, 2018, p. 71). It aggregates data that can be useful in shedding light on Google Search usage as a proxy for the public's interests, concerns, or intentions about a specific theme. It has been used to analyse many variables in a wide range of areas, including IT, communications, medicine, health, business, and economics (JUN; YOO; CHOI, 2018).

In Google Trends the results are returned on a scale ranging from 0 to 100, with 100 representing the highest proportion for the terms queried and zero the lowest. In a search for the terms “*Plataforma Lattes*” across Brazil month by month from January 2004 to August 2021, Trends assign the highest value (100) to the month with the highest volume (*i.e.*, November 2010 as depicted in Figure 3.4. All other months are then represented as a fraction of that maximum.

Figure 3.4 – Interest over time on the terms “Plataforma Lattes”, “Research Gate” and “ORCID”, Jan. 2004-Aug.2021



Source: Authors’ Own. Data sourced from https://trends.google.com/trends/. Search String: “Plataforma Lattes”; “ResearchGate” and “ORCID”. Note: Parameters used for the search were: a) selected region: “Brazil”; b) time frame: “Jan. 2004 to Jan. 2021”. Scores are based on the absolute search volume for each term separately, relative to the number of searches received by Google for each term separately

The above process also applies to queries that contain other terms, which is the case in Figure 3.4. So, as seen in that Figure, “ResearchGate” and “ORCID” are in a query separately, therefore, 100 is assigned to that highest month for each one of them respectively (*i.e.*, April 2018 and August 2020) and then all results across those terms are proportional to those points. Consequently, in interpreting results, we bear in mind that we can make evaluations about the rate of change and comparisons among the same term; however, we are not able to assess the total number of searches for any term, nor compare them directly as they were in a query separately. The idea is to show their usage tendency.

From Figure 3.4, we can notice that in terms of search engine tendency, *Plataforma Lattes* is decreasing with time starting from 2010. It is also interesting to note that other players like ResearchGate and ORCID have gained more popularity over time in Brazil.

Another typical trait of infrastructures is their invisibility, or the "visibility upon breakdown" (STAR, 1999). Since they become essential utilities for social actors' activities, they blend into their daily lives and are only perceived when they fail to deliver their critical services. This has been the case with the recent Lattes "blackout" (ROCHA, 2021) and in July-August 2021 we can notice a relative increase of *Plataforma Lattes* in Google search.

3.5.4 Path dependent dynamics

The lock-in concept applies especially in the context of technologies that generate positive network effects. ICTs, for example, are more valuable the more people use them. This externality puts pressure on the system towards non-coordinated standardization among several competing technologies. Another element that reinforces the lock-in is switching costs: in the emblematic case of the QWERTY keyboard (DAVID, 1985), typists needed to invest in alternative typing training if they were looking to switch to machines with other designs. The lock-in concept cannot be directly translated into the case of the Lattes Platform, given its public character, as it did not compete directly with other technologies to be established as the main artifact. It was defined by CNPq as the standard and gradually accepted by the two main groups that use it: researchers and scientific and research institutes. However, the analogy is valid for thinking about the future of the Platform.

The Lattes Platform is a digital system composed of its controller (CNPq) and two groups of users, who benefit from network effects as presented before. Directly, the more researchers have a CV-Lattes, the more it establishes itself as the *de facto* standard in the national scientific portfolio, increasing the value for the researcher who has it. Indirectly, for scientific and research institutes, the more researchers registered, the more it makes sense to use it as a source of information for the most diverse processes, from the formation of talent banks to the granting of scholarships and financing. Also indirectly, the more scientific and research institutes use the Platform, the greater the value for the researcher who owns it.

The switching cost, in its turn, is related to the potential of multihoming, that is, the development or consumption of products/services on more than one platform. In the case of Plataforma Lattes, there is currently no platform that is a perfect substitute for its services. Thus, at the moment, multihoming is possible not because of the cost, but because of the lack of reasonable alternatives.

The two elements that guarantee users' lock-in into the Lattes Platform are under transformation. The platformization of science advances with the establishment of new digital platforms dedicated to different phases of scientific activity. Academia.edu, ResearchGate, and ORCID already have millions of users in the whole world and are more and more searched for in Brazil (see Figure 3.4); it seems reasonable to assume that shortly the possibility of multihoming will be real and Lattes Platform users will be able to migrate to these new intermediaries.

Once an eventual migration process starts, negative network effects come into play. Assuming that a large number of researchers started to adopt another platform, or that scientific and research institutes with weight in the national S&T system turn to other solutions, a dynamic of negative network effects would be set in motion. For instance, recently, FAPESP – *Fundação de Amparo à Pesquisa de São Paulo* – declared that ORCID ID will be mandatory for all researchers and it is going to be used to monitor their scientific production and research results (FAPESP, 2021b). Fewer scientific and research institutes requesting the CV-Lattes would discourage researchers from owning/updating their registration. The negative spiral could, in theory, occur at an accelerated speed. The user ecosystem built over two decades could be lost.

3.6 Final comments

The Lattes Platform is crucial for the S&T community in Brazil. Its importance goes beyond *vitae* showcases of Brazilian scholars, as we have exemplified through the multilevel analysis synthesized in Table 3.4 and through the many studies that used its data to provide a big picture of the Brazilian National Innovation System.

Table 3.4 – Multilevel analysis of Lattes Platform

Levels	Category	Lattes
Socio-economic	Ownership	Governmental
	Governance	Managed through the "Terms and Conditions", which stipulate that the company collects and stores personal information from users to be used for supporting policy making and evaluation, sharing data with third parties except for personal data
	Business model	Public funding
	Market scope	National
	Economic effect	Network effect
	Infrastructure	Criticality and invisibility
	Path dependence dynamics	Lock-in effect
Techno-cultural	Technological components	Protocols; operational bases; warehouse bases; and web information repositories
	User/usage	Based on the stimuli to join the platform, mainly because having updated information is a precondition for accessing public funding and scientific research
	Content	UGC (user-generated content) with low standardization (pre-determined categories for metadata)
	Open innovation mechanism	Open data
Networked-scholar	Networking components	Absent
	Knowledge sharing	Based on the free access to scholars' curricula vitae (e.g., DOI for publication) and Research Groups' information; search mechanism based solely on research name
	Identity	User identity is conveyed through the profile and there are indirect reputation indexes (e.g, productivity grants, citations, impact factor, etc.)

Source: Authors' own. Note: (*) Based on Manca (2017)

Lattes is also a relevant source of information for policy decision-making based on pieces of evidence as argued; however, despite its originality and criticality, it seems that CNPq was not able to keep pace with the technological development of the new generation of platforms, such as the inclusion of advanced machine-learning algorithms to generate feedback effects and dashboards of analyses that could be automated, integrating data about talent, capabilities, and opportunities to enable real-time strategic decision-making. It seems that the Lattes Platform is trapped in a less dynamic and less promising technological trajectory and with constant budgetary constraints it will hardly be able to escape from the trap. Below, we comment on both the path dependence and the budgetary constraint effects and finally, we bring some points to reflect on the crossroads that lie in the future path of the Lattes Platform envisioning three possible scenarios.

3.6.1 Path dependence

The Lattes Platform maintains its infrastructural characteristic in the Brazilian context due to its criticality. However, launching only the open data strategy hindered the platform, for example, from providing research IDs (already occupied by ORCID), communication functionalities (taken over by ASN such as ResearchGate or even more conventional social network as LinkedIn), or even new useful functions within the restricted e-portfolio proposal. The Lattes Platform fulfilled its function of providing information about scholars and the scientific communities to the government and S&T institutions. It became a critical platform for all Brazilian scholars, who registered and showcased their e-portfolios. Its success, however, could be the reason for its stagnation (lock-in). Lattes was an early adopter of a new generation of technologies and organizational models but it was not prepared for the digital transformation that has accelerated since the 2010s, with the explosion of big data, the use of algorithms to order this same data, and the formation of collaborative platform ecosystems.

3.6.2 Budgetary constraints

The Lattes Platform's business model is based on public funding as it offers free-of-charge services. After it was launched, there were drastic budget reductions – the average expected budget for 2000-2003 was BRL 4.29 million and for 2017-2019, it was five times less – and this reduction tendency was worsened from 2016, probably as a result of the approval of the Constitutional Amendment (95A), capping public spending for 20 years. There is no public evidence demonstrating that CNPq has a well-developed strategy to maintain and upgrade the Lattes Platform, as in its most recent Strategic Plan there is no mention about the platform at all.

In 2019, in a conference to celebrate the Lattes Platform's 20th anniversary, the then CNPq President mentioned that:

> There is a series of improvements happening (…). The serious thing to mention about Lattes [Platform] and Carlos Chagas Platform (…) is that we need to reformulate both of them. The decision we took at the beginning of this year is that in 2019 we are going to invest in Carlos Chagas Platform because it is more urgent (…) and the expectation, and I am afraid to make forecasts as the crystal ball can be a little cloudy, but the expectation until mid-2020 is that we will make these overall [changes] of Carlos Chagas and then we will start to do the same process for Lattes [Platform]. Of course, marginal changes will still happen (…). (AZEVEDO, 2019, pt. 1h45min)[78]

Indeed, his crystal ball was cloudy. No extra budget was allocated for the Lattes Platform; rather, it was reduced. The National Strategy of Innovation published in 2020 by the Presidential Office (BRASIL, 2020) and approved in 2021 (MCTI, 2021) allocated BRL 485.000 for implementing a new platform substituting Carlos Chagas Platform[79] for 2021. There are no mentions of the Lattes Platform in that plan.

3.6.3 Crossroads

Overcoming the budgetary constraints and dealing with technological modernization to escape path dependence are the challenges facing the Lattes Platform directly. These two obstacles stand in the way of the platform and need to be overcome so that it redefines its new trajectory in the reality of the infrastructuralization of platforms. As it is currently a critical infrastructure, the Lattes Platform would have to be updated to stop relying solely on its enforcement power and, let's say, win over users (scholars and S&T agencies) for the service it provides. We do not believe that this necessarily implies transforming the Lattes Platform into an ASN; rather it needs to be updated and incorporate new tools that allow, for example, feedback mechanisms for users (complying with LGPD, not for surveillance purposes). There are studies on Lattes Platform usability concerned with the aspects of user performance and satisfaction which show that system inconsistencies affect users' experiences (RAMOS *et al.*, 2017); also, Google Trends data presented show that more and more ASN are being searched in Brazil.

What the Lattes Platform seems to confirm is that we are witnessing a historic moment in which public infrastructures fragment and are complemented or supplanted by

[78] In the original: "*Há uma série de aprimoramentos que estão acontecendo (...). A coisa séria de se falar sobre Lattes e Carlos Chagas (...) é que precisamos renovar as duas. A decisão que a gente tomou no início desse ano foi de que, nesse ano, nós vamos investir na Carlos Chagas porque ela é mais urgente (...) e a expectativa, e eu morro de medo de fazer essas previsões porque a bola de cristal pode estar meio nebulosa, mas a expectativa é que até o meio do ano que vem a gente vai fazer todo esse* overall*, não é, da Carlos Chagas e aí a gente vai começar a fazer o mesmo tipo de coisa com o Lattes. Claro, as mudanças incrementais vão continuar acontecendo* (...)".

[79] https://carloschagas.cnpq.br/

private infrastructures (PLANTIN *et al*., 2018; PLANTIN; LAGOZE; EDWARDS, 2018). Van Dijck, (2020) highlights how private intermediary platforms push upwards to occupy sectoral platforms, and downwards, to occupy the space of flows in the infrastructural layers of platforms. Due to this pattern of expansion, one might consider Dai, Shin e Smith (2018, p. 23) warning:

> At the present stage, online platforms mainly target different stages of the scientific process, e.g. providing services for data analysis, publishing, or evaluation. As technologies, standards and protocols develop an integrated online platform that provides services across the whole scientific process […] or integrates various other platform services [...] can be envisioned. This raises important policy questions related to the ownership, control, and access to such platforms.

On other occasions, private platforms feed on data produced on public platforms, which leads the author to consider that civil-democratic principles are being disrespected[80]. Taken together, the evidence presented leads us to question whether the Lattes Platform is on an irreversible trajectory of obsolescence or whether there will be a space for it in the future configuration of digital platforms for science.

3.6.4 Possible scenarios?

Although it is reasonable to celebrate the Lattes Platform as a Brazilian heritage for science and technology, the combination of the lock-in effect and the budget cuts for its maintenance and modernization jeopardize the Platform. From that, we envision three possible scenarios.

In the first one, without adequate investment for its modernization and inclusion of new functionalities, users would start to overcome the lock-in through migration to other new platforms with better services offer. Little by little, the Platform would lose relevance and wither away.

In another negative scenario, even with inadequate investments, CNPq would be able to maintain the mandatory use of the Lattes Platform. However, given the predictable low cost for users of having more than one platform, multihoming would be established: this way, users would use Lattes due to its enforcement and would adopt other platforms for their value/convenience. In this scenario, the Lattes Platform would become a sort of “zombie”: dead for users and live for CNPq.

[80]“Is the incorporation of data flows generated in public sectors (e.g. schools, hospitals) permitted when they can be connected to data flows outside the public realm?” (VAN DIJCK, 2020, p. 16). The same reasoning would apply to the GaaP practices presented in this article for the Lattes Platform.

Finally, in a more positive scenario, with adequate budget and investments, Lattes Platform would be modernized and would start to attract all groups towards its ecosystem for the value it delivers.

3.7 References

ADAMATTI, D. F.; CASTELFRANCHI, C. A Trust Metric to Lattes Curriculum Data. **Artificial Intelligence and Applications**, v. 2, p. 1-10, 2015.

ARRUDA, D. *et al.* Brazilian Computer Science research: Gender and Regional Distributions. **Scientometrics**, v. 79, p. 651-665, 2009. Available at: https://link.springer.com/article/10.1007/s11192-007-1944-0. Accessed: 28 mar. 2022.

AZEVEDO, J. L. F. Mecanismos de Fomento para CT&I do CNPq e os 20 anos da Plataforma Lattes. 2019. Available at: https://www.youtube.com/watch?v=8fdAXHw5GII. Accessed: 28 mar. 2022.

BELLO, M.; GALINDO-RUEDA, F. **Charting the digital transformation of science**: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). OECD Science, Technology and Industry Working Papers, No. 2020/03. Paris: OECD Publishing, 2020. DOI: https://doi.org/10.1787/1b06c47c-en.

BORGMAN, C. L. **Scholarship in the Digital Age**: Information, Infrastructure, and the Internet. Cambridge, Mass. USA: MIT University Press, 2007.

BRASIL. **Decreto nº 10.534, de 28 de outubro de 2020**. Institui a Política Nacional de Inovação e dispõe sobre a sua governança. 2020. Available at: https://www.in.gov.br/en/web/dou/-/decreto-n-10.534-de-28-de-outubro-de-2020-285629205. Accessed: 29 mar. 2022.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Available at: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Accessed: 22 mar. 2022.

BRASIL. **Lei nº 14.129, de 29 de março de 2021**. Princípios, regras e Instrumentos para o Governo Digital e para o Aumento da Eficiência Pública. 2021. Available at: https://www.in.gov.br/en/web/dou/-/lei-n-14.129-de-29-de-marco-de-2021-311282132. Accessed: 29 mar. 2022.

BUSCH, C. *et al*. **Uncovering blindspots in the policy debate on platform power 2**. [online] PLATFORM OBSERVATORY EU, 2021. Available at: https://platformobservatory.eu/app/uploads/2021/03/05Platformpower.pdf. Accessed: 22 mar. 2022.

CALIARI, T.; RAPINI, M. S.; CHIARINI, T. A Cooperação Com Empresas Aumenta a Geração De Tecnologia Nas Universidades? Análise a Partir Do Diretório Dos Grupos De Pesquisa No Brasil Do CNPq. **Parcerias Estratégicas**, v. 23, n. 47, p. 9-28, 2018.

CALIARI, T.; RAPINI, M. S.; CHIARINI, T. Research infrastructures in less developed countries: the Brazilian case. **Scientometrics**, v. 122, n. 1, p. 451-475, 2019.

CHEN, J. Y.; QIU, J. L. Digital utility: Datafication, regulation, labor, and DiDi's platformization of urban transport in China. **Chinese Journal of Communication**, v. 12, n. 3, p. 274-289, 2019. Available at: https://www.tandfonline.com/doi/full/10.1080/17544750.2019.1614964. Accessed: 22 mar. 2022.

CHIARINI, T. *et al*. The Political Economy of Innovation Why is Brazil Stuck in the Technology Ladder?. **Brazilian Political Science Review**, v. 14, n. 2, 2020. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1981-38212020000200200&tlng=en. Accessed: 22 mar. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **20 anos de Plataforma Lattes**. 2019. Available at: https://www.youtube.com/watch?v=JEZmHv4e0Rw. Accessed: 25 mar. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Glossário - Glossário - Plataforma Lattes - CNPq**. 2022. Available at: http://lattes.cnpq.br/web/dgp/glossario. Accessed: 22 mar. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Termo de adesão e compromisso da base de dados Lattes**. s/d. Available at: https://wwws.cnpq.br/cvlattesweb/pkg_cv_estr.termo. Accessed: 22 mar. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO; CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO. **Plano Diretor de Tecnologia da Informação (PDTI 2014-2016)**. 2014a. Available at: https://www.gov.br/cnpq/pt-br/acesso-a-informacao/pdti20142016.pdf. Accessed: 28 mar. 2022.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO; CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO.

Plano Estratégico 2025. 2014b. Available at: https://www.gov.br/cnpq/pt-br/acesso-a-informacao/institucional/planejamento_estrategico_2025.pdf. Accessed: 28 mar. 2022.

COZMA, R.; DIMITROVA, D. Research Gate or Revolving Door? Uses and Gratifications of Academic Social Media Among Communication Scholars. **Journalism & Mass Communication Educator**, p.107769582096503, 2020.

D'ONOFRIO, M.G. The public CV database of Argentine researchers and the "CV-minimum" Latin-American model of standardization of CV information for R&D evaluation and policy-making. **Research Evaluation**, v. 18, n. 2, p. 95-103, 2009.

DAI, Q.; SHIN, E.; SMITH, C. **Open and Inclusive Collaboration in Science**: A Framework. 2018/07. OECD Science, Technology and Industry Working Papers. Paris: OECD, 2018.

DAMACENO, R. J. P. *et al*. The Brazilian Academic genealogy: Evidence of Advisor–advisee Relationships through Quantitative Analysis. **Scientometrics**, v. 119, n. 1, p. 303-333, 2019. Available at: https://link.springer.com/article/10.1007/s11192-019-03023-0. Accessed: 30 oct. 2020.

DAVID, P. A. Clio and the Economics of QWERTY. **The American Economic Review**, v. 75, n. 2, p. 332-337, 1985. Available at: https://www.jstor.org/stable/1805621. Accessed. 28 mar. 2022.

DI VAIO, A.; VARRIALE, L. Digitalization in the sea-land Supply chain: Experiences from Italy in Rethinking the Port Operations within Inter-organizational Relationships. **Production Planning & Control**, v. 31, n. 2-3, p. 220-232, 2020. Available at: https://www.tandfonline.com/doi/full/10.1080/09537287.2019.1631464. Accessed: 28 mar. 2022.

DIAS, T.; MOTTA, F. Um Retrato da Produção Científica Brasileira Baseado em Dados da Plataforma LATTES. **Brazilian Journal of Information Science**, v. 12, n. 4, p. 62-74, 2021. Available at: https://revistas.marilia.unesp.br/index.php/bjis/article/view/7831/5421. Accessed: 28 mar. 2022.

DIJCK, J. V. **The Culture of Connectivity**: A Critical History of Social Media. Oxford Press ed. New York: Oxford University Press, 2013.

DIJCK, J. V.; POELL, T.; WAAL, M. D. The Platform society. Public Values in a Connective world. Kettering: Oxford University Press, 2018.

DUARTE, K. B.; WEBER, R. O.; PACHECO, R. C. S. Um Método Orientado a Propósito Aplicado ao Currículo Lattes para Fins de Concessão de Fomento a pesquisadores Em Grupos Colaborativos. **Parcerias Estratégicas**, v. 23, n. 47, p. 181-196, 2018.

FUNDAÇÃO DE AMPARO À PESQUISA DE SÃO PAULO. **Edital CEPID 2021**. 2021a. Available at: https://fapesp.br/14948/edital-cepid-2021. Accessed: 25 mar. 2022.

FUNDAÇÃO DE AMPARO À PESQUISA DE SÃO PAULO. **FAPESP Adotará Registro Orcid iD na Submissão de Propostas**. 2021b. Available at: https://agencia.fapesp.br/fapesp-adotara-registro-orcid-id-na-submissao-de-propostas/36583/. Accessed: 25 mar. 2022.

FERNANDES, A. C. *et al.* Academy–industry Links in Brazil: Evidence about Channels and Benefits for Firms and Researchers. **Science and Public Policy**, v. 37, n. 7, p. 485-498, 2010.

FONSECA, B. *et al*. Network Analysis for Science and Technology management: Evidence from Tuberculosis Research in Fiocruz, Brazil. **PLOS ONE**, v. 12, n. 8, p. e0181870, 2017. Available at: https://dx.plos.org/10.1371/journal.pone.0181870. Accessed: 25 mar. 2022.

FONSECA, B. P. F.; FERNANDES, E.; FONSECA, M. V. A. Collaboration in Science and Technology Organizations of the Public sector: a Network Perspective. **Science and Public Policy**, v. 44, n. 1, p. 37-49, 2017.

GARCIA, R.; RAPINI, M. S.; CÁRIO, S. **Estudos de caso da interação universidade-empresa no Brasil**. Belo Horizonte: FACE/UFMG, 2018.

GUIDINI, M. B. *et al.* PPSUS/RS: Um Estudo sobre Avaliação de Impacto usando Abordagem quase-experimental. **Parcerias Estratégicas**, 23, n. 47, p. 165-180, 2018.

GÜNTHER, L. L. *et al*. Análise do Sistema de Currículo LATTES segundo Modelo CESM: Perspectivas para um Sistema de Informação para a e-science. **Perspectivas em Gestão & Conhecimento**, v. 10, n. 1, p. 107-130, 2020. Available at: https://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/46092/30063/. Accessed: 29 mar. 2022.

HILÁRIO, C. M.; GRÁCIO, M. C. C. Scientific Collaboration in Brazilian researches: a Comparative Study in the Information science, Mathematics and Dentistry Fields. **Scientometrics**, v. 113, n. 2, p. 929-950, 2017. Available at: http://link.springer.com/10.1007/s11192-017-2498-4. Accessed: 25 mar. 2022.

HILL, K. I **Tried to Live without the Tech Giants**. It Was Impossible. The New York Times. 2020. Available at: https://www.nytimes.com/2020/07/31/technology/blocking-the-tech-giants.html. Accessed: 29 mar. 2022.

JORDAN, K. From Social Networks to Publishing Platforms: a Review of the History and Scholarship of Academic Social Network Sites. **Frontiers in Digital Humanities**, v. 6, n. 5, p. 1-53, 2019. Available at: https://www.frontiersin.org/article/10.3389/fdigh.2019.00005/full. Accessed: 25 mar. 2022.

JUN, S.-P.; YOO, H. S.; CHOI, S. Ten Years of Research Change Using Google Trends: from the Perspective of Big Data Utilizations and Applications. **Technological Forecasting and Social Change**, v. 130, p. 69-87, 2018. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0040162517315536. Accessed: 25 mar. 2022.

KENNEY, M.; ZYSMAN, J. The Rise of the Platform Economy. **Issues in Science and Technology**, v. 32, n. 3, p. 62-69, 2016.

LEITE, D. *et al*. Publish or perish? Avaliação de Redes de Pesquisa e Colaboração Com RNPE. **Parcerias Estratégicas**, v. 23, n. 47, p. 83-102, 2018.

MAJCHRZAK, A.; MARKUS, M. L. Technology Affordance and Constraints in Management Information System. **Encyclopedia in Management Theory**, p. 832-834, 2013. Available at: https://sk.sagepub.com/reference/encyclopedia-of-management-theory. Accessed: 29 mar. 2022.

MANCA, S. An Analysis of ResearchGate and Academia.edu as Socio-technical Systems for Scholars' Networked Learning: a Multilevel Framework Proposal. **Italian Journal of Educational Technology**, v. 25, n. 3, 2017.

MASON, S. Adoption and Usage of Academic Social Networks: a Japan Case Study. **Scientometrics**, v. 122, n. 3, p. 1751-1767, 2020. Available at: http://link.springer.com/10.1007/s11192-020-03345-4. Accessed: 25 mar. 2022.

MAYER-SCHÖNBERGER, V.; RAMGE, T. Reinventing Capitalism in the Age of Big Data. New York: Basic Books, 2018.

MCMANUS, C.; BAETA NEVES, A. A. Production Profiles in Brazilian Science, with Special Attention to Social Sciences and Humanities. **Scientometrics**, v. 126, n. 3, p. 1-23, 2020. Available at: https://doi.org/10.1007/s11192-020-03452-2. Accessed: 25 mar. 2022.

MCQUIRE, S. One Map to Rule Them All? Google Maps as Digital Technical Object. **Communication and the Public**, v. 4, n. 2, p. 150-165, 2019. Available at: http://journals.sagepub.com/doi/10.1177/2057047319850192. Accessed: 25 mar. 2022.

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO DA REPÚBLICA FEDERATIVA DO BRASIL. **Resolução CI n. 1, de 23 de Julho de 2021**. Aprova a Estratégia Nacional de Inovação e os Planos de Ação para os Eixos de Fomento, Base Tecnológica, Cultura de Inovação, Mercado para Produtos e Serviços Inovadores e Sistemas Educacionais. 2021. Available at: https://www.in.gov.br/web/dou/-/resolucao-ci-n-1-de-23-de-julho-de-2021-334125807. Accessed: 26 mar. 2022.

MENA-CHALCO, J. P.; CESAR JUNIOR, R. M. ScriptLattes: an Open-source Knowledge Extraction System from the Lattes Platform. **Journal of the Brazilian Computer Society**, v. 15, n. 4, p. 31-39, 2009. Available at: https://www.scielo.br/j/jbcos/a/DNqND3wQHrTHkCNWQbKx6pt/?lang=en. Accessed: 26 mar. 2022.

MENA-CHALCO, J. P. *et al*. Brazilian Bibliometric Coauthorship Networks. **Journal of the Association for Information Science and Technology**, v. 65, n. 7, p. 1424-1445, 2014. Available at: http://doi.wiley.com/10.1002/asi.23010. Accessed: 26 mar. 2022.

MILLER, P.; ROSE, N. Governing the Present: Administering economic, Social and Personal Life. Cambridge: Polity Press, 2008.

MONTERO, J. J.; FINGER, M. Platformed! Network Industries and the New Digital Paradigm. **Competition and Regulation in Network Industries**, v. 18, n. 3-4, p. 217-239, 2017.

MOSCO, V. Becoming Digital : toward a post-Internet Society. United Kingdom: Emerald Publishing Limited, 2017.

MUKHERJEE, R. Jio Sparks Disruption 2.0: Infrastructural Imaginaries and Platform Ecosystems in “Digital India.” **Media, Culture & Society**, v. 41, n. 2, p. 175-195, 2019.

NASCIMENTO, M. On the “Missing Letter” to Lattes and the Nobel Prize in Physics. **Ciência e Sociedade**, v. 3, n. 2, p. 35-42, 2015. Available at: http://revistas.cbpf.br/index.php/CS/article/view/120/10. Accessed: 26 mar. 2022.

NECHUSHTAI, E. Could Digital Platforms Capture the Media through infrastructure? **Journalism**, v. 19, n. 8, p. 1043-1058, 2017. Available at: http://journals.sagepub.com/doi/10.1177/1464884917725163. Accessed: 26 mar. 2022.

NEGRI, F. de *et al*. Mapeamento da Infraestrutura Laboratorial das Instituições de Pesquisa do MCTI. Brasília, Brazil: MEC, 2013. Available at: http://lattes.cnpq.br/documents/11997/727909/Relatorio-infraestrutura-de-pesquisa-MCTI.pdf/c9800f40-4bba-4bf6-86de-b4d93cbedc18. Accessed: 29 mar. 2022.

NEGRI, F. De; SQUEFF, F. DE H. S. **Sistemas setoriais de inovação e infraestrutura de pesquisa no Brasil**. Brasília: Ipea/Finep/CNPq, 2016.

O`REILLY, T. Government as a platform. **Oper government**, p. 11-50, 2010. Available at: https://www.researchgate.net/publication/257244639_Open_Government_Collaboration_Transparency_and_Participation_in_Practice_Daniel_Lathrop_Laurel_Ruma_Eds_O%27Reilly_Sebastopol_CA_2010_402_pp_4499_paperback_ISBN_978-0-596-80435-0. Accessed: 26 mar. 2022.

OLIVEIRA, J. R. DE; MELLO, L. C.; RIGOLIN, C. C. D. Participação Feminina na Pesquisa sobre Tecnologia da Informação no Brasil: Grupos de Pesquisa e Produção Científica de Teses e Dissertações. **Cadernos Pagu**, v. 58, p. 1-51, 2020. Available at: https://www.scielo.br/j/cpa/a/TmJbqvbcBzdc9hkFGmjgHZK/?lang=pt. Accessed: 26 mar. 2022.

ORDUNA-MALEA, E. *et al*. Do ResearchGate Scores Create Ghost Academic reputations? **Scientometrics**, v. 112, n. 1, p. 443-460, 2017. Available at: https://link.springer.com/article/10.1007/s11192-017-2396-9. Accessed: 26 mar. 2022.

PACHECO, R.; KERN, V. Arquitetura Conceitual e Resultados da Integração de Sistemas de Informação e Gestão da Ciência e Tecnologia. **DataGramaZero - Revista de Ciência da Informação**, v. 4, n. 2, p. 1-10, 2003. Available at: https://brapci.inf.br/index.php/res/v/3877. Accessed: 26 mar. 2022.

PACHECO, R. C. DOS S.; KERN, V. M. Uma Ontologia Comum para a Integração de Bases de Informações e Conhecimento sobre Ciência e Tecnologia. **Ciência da Informação**, v. 30, n. 3, p. 56-63, 2001. Available at: https://www.scielo.br/j/ci/a/Mbjn8vKDsBhS4MXC9LRpg7S/?lang=pt. Accessed: 26 mar. 2022.

PACHECO, R. C. S. (PDF) Uma Metodologia de Desenvolvimento de Plataformas de Governo para Geração e Divulgação de Informações e de Conhecimento. 2003. Available at: https://www.researchgate.net/publication/234168640_Uma_Metodologia_de_Desenvolvimento_de_Plataformas_de_Governo_para_Geracao_e_Divulgacao_de_Informacoes_e_de_Conhecimento. Accessed: 26 mar. 2022.

PERLIN, M. S.; IMASATO, T.; BORENSTEIN, D. Is Predatory Publishing a Real Threat? Evidence from a Large Database Study. **Scientometrics**, v. 116, n. 1, p. 255-273, 2018. Available at: http://link.springer.com/10.1007/s11192-018-2750-6. Accessed: 27 mar. 2022.

PERLIN, M. S. *et al*. The Brazilian Scientific Output Published in journals: a Study Based on a Large CV Database. **Journal of Informetrics**, v. 11, n. 1, p. 18-31, 2017. Available at: https://linkinghub.elsevier.com/retrieve/pii/S1751157716301559. Accessed: 27 mar. 2022.

PLANTIN, J. *et al*. Infrastructure studies meet platform studies in the age of Google and Facebook. **New Media & Society**, v. 20, n. 1, p. 293-310, 2018. Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444816661553. Accessed: 27 mar. 2022.

PLANTIN, J.-C.; DE SETA, G. WeChat as Infrastructure: the techno-nationalist Shaping of Chinese Digital Platforms. **Chinese Journal of Communication**, v. 12, n. 3, p. 257-273, 2019. Available at: https://www.tandfonline.com/doi/full/10.1080/17544750.2019.1572633. Accessed: 26 mar. 2022.

PLANTIN, J.-C.; PUNATHAMBEKAR, A. Digital Media Infrastructures: Pipes, Platforms, and Politics. **Media, Culture & Society**, v. 41, n. 2, p. 163-174, 2018. Available at: http://journals.sagepub.com/doi/10.1177/0163443718818376. Accessed: 27 mar. 2022.

PLANTIN, J.-C.; LAGOZE, C.; EDWARDS, P. N. Re-integrating Scholarly Infrastructure: the Ambiguous Role of Data Sharing Platforms. **Big Data & Society**, v. 5, n. 1, p. 205395171875668, 2018. Available at: https://journals.sagepub.com/doi/pdf/10.1177/2053951718756683. Accessed: 29 Sep. 2019.

POELL, T. Three Challenges for Media Studies in the Age of Platforms. **Television & New Media**, v. 21, n. 6, p. 650-657, 2020. Available at: http://journals.sagepub.com/doi/10.1177/1527476420918833. Accessed: 27 mar. 2022.

POELL, T.; NIEBORG, D.; VAN DIJCK, J. Platformisation. **Internet Policy Review**, v. 8, n. 4, p. 1-13, 2019. Available at: https://policyreview.info/node/1425. Accessed: 27 mar. 2022.

RAMOS, C. *et al*. Usabilidade de Plataforma LATTES Apresenta Níveis Inadequados de Desempenho e Satisfação do Usuário. **Ergodesign & HCI**, v. 5, p. 153-156, 2017. Available at: https://www.proceedings.blucher.com.br/article-details/avaliao-da-usabilidade-da-plataforma-lattes-compreendendo-os-nveis-de-desempenho-e-satisfao-do-usurio-25904. Accessed: 27 mar. 2022.

RAPINI, M. S. Interação Universidade-Empresa no Brasil: Evidências do Diretório dos Grupos de Pesquisa do CNPq. **Estudos Econômicos (São Paulo)**, v. 37, n. 1, p. 211-233, 2007. Available at: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0101-41612007000100008&lng=pt&tlng=p. Accessed: 22 mar. 2022.

RAPINI, M. S. *et al*. The Intensity of Private Funding and the Results of university? Firm interactions: the Case of Brazil. **Innovation & Management Review**, v. 16, n. 2, p. 161-184, 2019. Available at: https://www.revistas.usp.br/rai/article/view/160164. Accessed: 22 mar. 2022.

ROCHA, R. Lattes fora do Ar Pode Gerar uma Tragédia para a Ciência, Aponta Pesquisadora. **CNN Brasil**, 2021. Available at: https://www.cnnbrasil.com.br/nacional/2021/07/28/lattes-fora-do-ar-pode-gerar-uma-tragedia-para-a-ciencia-aponta-pesquisadora. Accessed: 27 mar. 2022.

RODIMA-TAYLOR, D.; GRIMES, W. W. International Remittance Rails as Infrastructures: Embeddedness, Innovation and Financial Access in Developing Economies. **Review of International Political Economy**, v. 26, n. 5, p. 839-862, 2019. Available at: https://www.tandfonline.com/doi/full/10.1080/09692290.2019.1607766. Accessed: 27 mar. 2022.

ROSSI, L.; DAMACENO, J. P.; MENA-CHALCO, J. P. Genealogia Acadêmica: um Novo Olhar Sobre Impacto Acadêmico de Pesquisadores. **Parcerias Estatégicas**, v. 23, n. 47, p. 197-212, 2018. Available at: http://seer.cgee.org.br/index.php/parcerias_estrategicas/article/viewFile/910/828. Accessed: 27 mar. 2022.

ROSSI, P.; DWECK, E. **Impacts of the New Fiscal Regime on Health and Education**. 2016. Available at: https://www.scielo.br/j/csp/a/jXPKhnYnvR4BtZ4LcHDkm4M/?lang=en. Accessed: 27 mar. 2022.

SAMPAIO, R. B. **e-Lattes**: Um Novo Arcabouço em Linguagem R para Análise do Currículo Lattes. 2018. Available at: https://medium.com/data-net-sci/e-lattes-um-novo-arcabou%C3%A7o-em-linguagem-r-para-an%C3%A1lise-do-curr%C3%ADculo-lattes-9ecb2f68574e. Accessed: 27 mar. 2022.

SAMPAIO, R. B. *et al*. Scientometric Analysis of Research Output from Brazil in Response to the Zika Crisis Using e-Lattes. **Journal of Data and Information Science**, v. 5, n. 4, p. 137-146, 2020. Available at: https://sciendo.com/article/10.2478/jdis-2020-0038. Accessed: 27 mar. 2022.

SANTIAGO, M. de O.; AFFONSO, F.; DIAS, T. M. R. Scientific Production of Women in Brazil. **Transinformação**, v. 32, p. 1-11, 2020.

SELLAR, S.; GULSON, K. N. Becoming Information centric: the Emergence of New Cognitive Infrastructures in Education Policy. **Journal of Education Policy**, v. 36, n. 3, p. 309-326, 2019. Available at: https://www.tandfonline.com/doi/full/10.1080/02680939.2019.1678766. Accessed: 27 mar. 2022.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. A Ciência nas Regiões brasileiras: Evolução da Produção e das Redes de Colaboração Científica. **TransInformação**, v. 28, n. 1, p. 15-31, 2016. Available at: https://www.scielo.br/pdf/tinf/v28n1/0103-3786-tinf-28-01-00015.pdf. Accessed: 23 Nov. 2020.

SIDONE, O. J. G.; HADDAD, E. A.; MENA-CHALCO, J. P. Scholarly Publication and Collaboration in Brazil: the Role of Geography. **Journal of the Association for Information Science and Technology**, v. 68, n. 1, p. 243-258, 2016. Available at: https://doi.org/10.1002/asi.23635. Accessed: 27 mar. 2022.

SILVA, F. M.; SMIT, J. W. Organização da Informação em Sistemas Eletrônicos Abertos de Informação Científica & Tecnológica: Análise da Plataforma Lattes. **Perspectivas em Ciência da Informação**, v. 14, n. 1, p. 77-98, 2009. Available at: https://www.scielo.br/j/pci/a/KMc8TfJhPwCLrJCLdw4qTQr/abstract/?lang=pt. Accessed: 27 mar. 2022.

SILVA, V. J.; BONACELLI, M. B. M.; PACHECO, C. A. O Sistema Tecnológico Digital: Inteligência artificial, Computação em Nuvem e Big Data. **Revista Brasileira de Inovação**, v. 19, p. 1-31, 2020. Available at: https://periodicos.sbu.unicamp.br/ojs/index.php/rbi/article/view/8658756. Accessed: 27 mar. 2022.

SILVEIRA, E. da. A Face Perversa da Plataforma Lattes. **Questão de Fato**, 27 nov. 2020. Available at: https://revistaquestaodeciencia.com.br/questao-de-fato/2020/11/27/face-perversa-da-plataforma-lattes. Accessed: 27 mar. 2022.

SRNICEK, N. Platform capitalism. Cambridge; Malden: Polity Press, 2017.

STAR, S. L. The Ethnography of Infrastructure. **American Behavioral Scientist**, v. 43, n. 3, p. 377-391, 1999. Available at: http://journals.sagepub.com/doi/10.1177/00027649921955326. Accessed: 27 mar. 2022.

SUZIGAN, W.; ALBUQUERQUE, E. da M.; CARIO, S. Em Busca da Inovação: Interação Universidade-Empresa no Brasil. Belo Horizonte, Brazil: Autêntica Editora, 2011. Available at: https://www.scielo.br/j/ee/a/74DXgDzzdKGwn6tVGHgTc9P/. Accessed: 27 mar. 2022.

SUZIGAN, W. *et al*. University and Industry Linkages in Brazil: Some Preliminary and Descriptive Results. **Seoul Journal of Economics**, v. 22, n. 4, p. 569-611, 2009.

TAHA, N. *et al*. Social Media Shaping e-Publishing and Academia. Cham Springer International Publishing, 2017.

THELWALL, M.; KOUSHA, K. ResearchGate: Disseminating, Communicating, and Measuring Scholarship? **Journal of the Association for Information Science and Technology**, v. 66, n. 5, p. 876-889, 2014. Available at: http://doi.wiley.com/10.1002/asi.23236. Accessed: 27 mar. 2022.

THOMAS, J.; MOHRMAN, S.A. A Vision of Data and Analytics for the Science of Science Policy. *In*: LANE, J. *et al*. (ed.). Fealing. Stanford, USA: Stanford University Press, 2011. p. 258-281.

TORRES, J. A. La Gestión de la Información como Herramienta de Planificación y Toma de Decisiones in Políticas Universitarias para el Siglo XXI: Perspectivas y Temas de Agenda. **Teseo**, p. 349-366, 2019. Available at: https://www.teseopress.com/politicasuniversitariasparaelsigloxxi/chapter/13-la-gestion-de-la-informacion-como-herramienta-de-planificacion-y-toma-de-decisiones-posibilidades-de-innovacion-y-barreras-culturales-en-las-universidades/. Accessed: 27 mar. 2022.

VAN DIJCK, J. Seeing the Forest for the trees: Visualizing Platformization and Its Governance. **New Media & Society**, p. 1-19, 2020. Available at: http://journals.sagepub.com/doi/10.1177/1461444820940293. Accessed: 27 mar. 2022.

VELETSIANOS, G. Social Media in Academia: Networked Scholars. New York: Routledge, 2016.

VELETSIANOS, G.; KIMMONS, R. Networked Participatory Scholarship: Emergent techno-cultural Pressures toward Open and Digital Scholarship in Online Networks. **Computers & Education**, v. 58, n. 2, p. 766-774, 2012. Available at: https://linkinghub.elsevier.com/retrieve/pii/S0360131511002454. Accessed: 27 mar. 2022.

VELETSIANOS, G.; JOHNSON, N.; BELIKOV, O. Academics' Social Media Use over Time Is Associated with individual, relational, Cultural and Political Factors. **British Journal of Educational Technology**, v. 50, n. 4, p. 1713-1728, 2019. Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/bjet.12788. Accessed: 27 mar. 2022.

WELLER, M. J. The Digital Scholar : How Technology Is Transforming Scholarly Practice. London, UK: Bloomsbury Academic, 2011.

ZHAO, Y. *et al*. The Evolution of Platform Business models: Exploring Competitive Battles in the World of Platforms. **Long Range Planning**, v. 53, n. 4, p. 101892, 2019. Available at: https://www.sciencedirect.com/science/article/pii/S0024630118306368. Accessed: 27 mar. 2022.

THIRD PART: THE ALGORITHMIZATION OF SCIENCE

CHAPTER 4 – FRAMING THE EFECTS OF MACHINE LEARNING ON SCIENCE[81]

"Machine intelligence is the last invention that humanity will ever need to make."
Nick Bostrom
"I am worried that algorithms are getting too prominent in the world. It started out that computer scientists were worried nobody was listening to us. Now I'm worried that too many people are listening."
Donald Knuth

4.1 Introduction

Besides becoming one of the central technologies of the alleged fourth industrial revolution (YU; LIANG; WU, 2021), artificial-intelligence-based systems influence more than commercial trends. They impact diverse social spheres and its influence extends to scientific research (CHUBB; COWLING; REED, 2021). The scholar community recognizes that "Scientific research can lead to technological advance, but technology very much affects advances in science" (STEPHAN, 2010, p. 229). The same applies to recent artificial intelligence (AI) techniques, such as machine learning (ML) and deep learning (DL). Still, available studies offer a partial account of their influence on science. Cockburn *et al*. (2018) investigate bibliometric data and patents. They conclude that there was a reorientation towards applied deep learning solutions from the 2010s onwards. More than that, they conclude that DL systems alters knowledge-producing circles such as science. Vasilescu and Filzmoser (2021) conceptualize *machine invention systems*, which share a similar view on the *invention in the method of invention* (IMI). Bianchini *et al*. (2020) are more specific: they investigate how different areas of research instrumentalized DL systems. They demonstrate its use as a tool and provide insights into the effects of this application for the health sciences research sector. However, they do not address other forms of influence, *e.g*., steering the scientific agenda.

Despite recent works discussing specific impacts of new AI techniques on science, there is no broad and holistic view that synthesizes the channels through which this interaction takes place. Our goal in this article is to advance in this direction, framing the

[81] Originally published as: SILVA, V.; BONACELLI, M., PACHECO, C. Framing the effects of machine learning on science, AI & Society: Journal of Knowledge, Culture and Communication, 2022. Available in https://link.springer.com/article/10.1007/s00146-022-01515-x.

influence ML systems have upon science. We recognize the mutual interaction between science and technology (S&T); however, in this article, we focus on the influence that goes from technology to the sphere of science. Aiming to assess this influence systematically, we have developed a general taxonomic framework called "Rosenberg Effects." Nathan Rosenberg (1927 – 2015) observed the interaction in which technological developments condition further scientific advance (ROSENBERG, 1976, 1982). We coded Rosenberg's historical examples and put together a framework consisting of four categories of technology's effects on science: intellectual, experimental, economic, and instrumental. The intellectual effect occurs when technology acts as a source of unexplained issues to be scientifically examined. The experimental effect occurs when technology behaves as a source of empirical knowledge to be scientifically explored. Economic effects emerge when technology acts as a source of economic incentives that steer the scientific research agenda. Finally, instrumental effects are triggered when technology acts as a source of new scientific "capital goods" that transform scientific practices. These categories are further sub-divided into eight sub-categories.

After the elaboration of the general framework, we apply it to the case of ML/DL systems resorting to primary (*e.g.*, bibliometric data and content analysis of AI manuals) and secondary data from journals (*e.g., AI & Society, Technology in Society*) and business reports. Applying the framework in the case of ML/DL reveals significant effects of this AI approach upon science via all categorized effects, which demonstrates the broad power of AI technology to influence science. The value of this article is twofold. First, it lies in providing a simple framework to assess the aforementioned relations between technology and science. Second, it applies the framework to the case of AI, providing this broad and holistic view of the influence of new artificial intelligence techniques (machine learning, deep learning) on science. More specifically, the article details the channels through which this relationship takes place, the nature of these channels and the *loci* in which the potential effects on science unfolds.

We argue that the panorama outlined in the article branches out to studies of technology and science. Regarding technology, the study offers a simple framework that allows visualizing the pervasiveness of a given technology in relation to science. With regard to science, we argue that the ability to observe its conditioning by technology (and economic interests) should, in theory, lead to a reflection on scientific autonomy (which may be more or less socially desirable depending on the case). Finally, with regard to the development of ML/DL-based systems, the article raises the question: has the use and development of

ML/DL-based systems taken into account the responsibility for the effects it (potentially) causes in science?

We structure the article in six sections. Next is a brief literature review of the effects of technology on science. Then we present the methodology that fundaments the framework. The fourth session goes deeper into the description of the "Rosenberg effects." In the fifth session, we apply the framework to the case of ML/DL. Finally, we present some closing remarks in the discussion session.

4.2 Reviewing the literature on technology effects on science

Nathan Rosenberg is "one of the founding fathers of the Economics of Innovation" (MOWERY *et al*., 2019, p. 283). One of his topics of interest was the complex relations between S&T. He started his research career when the linear model of innovation was dominant. That model framed innovation as an unidirectional process in which scientific research outputs were inputs to the development of new products and innovations. Rosenberg found numerous cases in which this linear relationship was not confirmed (MOWERY; ROSENBERG, 1998; ROSENBERG, 1976). Thus, he understood that "the relationship between science and technology is bidirectional and nonlinear, with the direction of the link often going from technology to science and not always from science to technology" (ROSENBERG, 1982, p. 285).

For Rosenberg, the universe of technology is an autonomous body of knowledge that developed throughout history before modern science. This independence of technology from science, however, does not mean that there is no mutual influence. His argument is that science, especially from the 20th century onwards, becomes more and more influenced by economic considerations[82]. At the same time, technology has become much more science-based. It is within the framework of the capitalist industrial economy that Rosenberg sees the relationship between S&T and, of most interest to us, the effects of technology upon science.

Price (1984) also perceives technology and science as autonomous fields of knowledge. However, he further argues that there is no relationship between the scientific and technological domains in their "normal state." Significant inter-domain relationships would only occur in moments of rupture: the rise of a new scientific paradigm or the emergence of

[82] Even though we agree with Flexner (1939) that curiosity is as much an essential driver of scientific inquiry as to the prospect of use.

an "unexpected" technological innovation[83]. When referring to the historiographical dynamics of science from a Kuhnian perspective, Price's analysis does not reach less disruptive interactions that occur between S&T, whose significance to a paradigm may be negligible but whose economic significance may be enormous. For this reason, even when using historical examples of technologies that influenced science, his emphasis is on *intellectual effects*: "Since the seventeenth century, particular technologies have often been empirically investigated in order to gain a deeper understanding of them and, *if possible*, to increase their efficiencies." (PRICE, 1984, p. 7, emphasis added). Price also works with examples of technologies that influenced science in a recurring category in Rosenberg's work: instrumentation. He proposed the term "instrumentalities" to define practical knowledge in experimental laboratories that lead to new scientific investigations. Instrumentalities would include the traditional instruments, but also "processes" and "effects", or the whole set of "important techniques that help make new science" (PRICE, 1984, p. 13). The emphasis on "new" science also illustrates that Price was more concerned with new frontiers' effects than, for example, productivity enhancing effects on pre-existing scientific activities. Instrumentalities would still have an interface with the economy since, in the laboratory, they encourage the expansion of understanding, but in industry, they foster new products/services, *i.e.*, innovations. As much as it is possible to find examples that have been explored both scientifically and economically, the author himself admitted that he was dealing with "only one class of rather important events, not with the entirety of scientific or technological change, which has already been admitted to proceed endogenously in step-by-step normal changes." (PRICE, 1984, p. 15).

The over-appreciation of the ontological discussion of science (exogenous/endogenous) led Price to disregard clear examples of the cross-domain influence which were published contemporaneously to him. Gazis (1979) proposed two main categories of T-S influence, in his words, "scientific investigation stimulated by the need to understand some problems" and "the provision by technology of new tools for scientific investigation." (GAZIS, 1979, p. 247). Especially concerning the effects generated by the problems, Gazis presented a rich portfolio of cases that, descending to the "factory floor" level, illustrates the everyday interactions between technology and science. More recently, Brooks (1994) also offers examples of how technology acts as a source of challenges for science. This proto

[83] Price probably wanted to allude to a technological innovation that escapes the dominant technological paradigm when he mentions unexpected technological innovation. He could not mobilize this concept because Giovanni Dosi (1984)would still spread it in the years to come.

category refers to the solution of specific problems that unfold into new areas of basic research. He comments on how technological diffusion also acts as a trigger on science within this same section of his article (BROOKS, 1994, p. 483). In addition, he points to instrumentation and measurement techniques as another important channel for the influence of technology on science.

Rosenberg comes closer to a categorization of the effects of technology on science in *How Exogenous is Science?* (ROSENBERG, 1982). He sought to identify “avenues of influence” of economics on science. Its central premise, therefore, is to conceive technology as an economic variable. Having admitted that, it is possible to investigate how “technological concerns shape the scientific enterprise in various ways” (ROSENBERG, 1982, p. 141). By operating in that way, Rosenberg leveraged his argument, given the moment he wrote: the consolidation of *Big Science*, which demands a high volume of economic resources to carry out science. Rosenberg referred to technology as a “base of observations” for science. Given episodes of chronological precedence, technological knowledge accumulates data to be further analyzed scientifically. In addition to data collection, on other occasions, the performance of technological artifacts is the only possible field of observation given the “inability to conduct valid experiments” (ROSENBERG, 1982, p. 148) in controlled experimental environments. There is also the fact that “many aspects of a material are not scientifically explored until the material has been used for a long time.” (ROSENBERG, 1982, p. 152). Finally, Rosenberg mentioned how technological breakthroughs might “validate the possibility of certain classes of phenomena [...] by providing an empirical demonstration of the falseness of conventional scientific wisdom” (ROSENBERG, 1982, p. 157).

In addition to the base of observations, Rosenberg commented on three other channels directly linked to economic interests. The first is the resolution of problems that sometimes impose “limits of further improvement,” something similar to Gazis’ “physical limits of technologies” (GAZIS, 1979). The second is complementary technologies, which are generally bottlenecks to the full development of a technological innovation, generating different demand points for new scientific knowledge. Finally, the perception of the potential economic return of specific technologies drives the scientific agenda, which we understand as a “diffusion potential”. Rosenberg also comments on the centrality of instrumentation as a channel of technological influence upon science, something he approached in later works. One must remember that all channels are “economic,” as this is the author’s assumption. Table 4.1 summarizes the main insights from the literature review.

Table 4.1 – Summary of Literature Review

Author	Major work in the T-S topic	Terminology used for T-S relations	Channels through which technology triggers science
Gazis, D.	Gazis, 1979	Stimulation of science by technology	Problem-solving; new tools
Rosenberg, N.	Rosenberg, 1982	Avenues of influence	Observation base; instrumentation; limits of improvement, complementary technologies, diffusion potential
Solla Price, D.	Price, 1984	Experimental science; applied science	Instrumentalities
Brooks, H.	Brooks, 1994	Impact of technology on science	New scientific challenges; instrumentation and measurement techniques

Source: author's own

4.3 Methodology

Rosenberg (1982) provided the most critical insights for developing our taxonomy. The author raises historical cases and analyzes their relationships between technology and science. Despite his intent of categorization, he only mentions "dimensions" in which technology influences science: technology as an "observation base" (ROSENBERG, 1982, p. 147), the power of technology to "influence the scientific agenda" (ROSENBERG, 1982, p. 147) and "instrumentation" (ROSENBERG, 1982, p. 147). Rosenberg did not homogenize the criteria with which he proposed each of the above groups[84], neither the author proposes a hierarchy of categories and sub-categories. Our effort is to generate a taxonomy for the "avenues of influence" of technology on science, based upon Nathan Rosenberg's work. The main objective of a taxonomic definition "is to establish sufficient shared meaning so that the class of entity can be investigated by a scientific community." (HODGSON, 2019, p. 5). Following the principle of parsimony, we "identify the minimum number of properties that are sufficient to demarcate one group of entities from all other entities." (HODGSON, 2019, p. 8). What Rosenberg named 'dimensions', we understand as categories whose general nature comes from their stimuli on science. We base sub-categories on the specific and original "triggering events" (BROOKS, 1994) that, starting from a phenomenon in the technological sphere, generates the activation of a scientific response (*e.g.,* data generation,

[84] "Base of observations" and "instrumentation" refer to the initial trigger that emerges in the technological sphere and that will influence science: new data from the body of empirical knowledge or new technological artifacts for instrumentation; on the other hand, "influencing the agenda" refers to the result, already in the scientific sphere, of changes that have taken place in the technological sphere. There is some confusion between cause and effect.

generation of unexplained facts, generation of unsolved problems, and generation of new instruments). More details on the *corpus* and our rules for deriving the categories are in the Annex 1.

4.4 Technology-science relations in view of Nathan Rosenberg's work: the "Rosenberg effects"

The first category presented in Table 4.2 is the **intellectual effect**: technology is a source of unexplained issues to be scientifically examined. Historically, technologies worked long before we could explain why. According to Rosenberg, if man expected an *ex-ante* detailed and profound scientific explanation of the facts inherent in each technology's operation, we would still be in the Stone Age. Scientific explanations have often been motivated by the gap in the human understanding of some technologies. Brooks (1994) pointed out that technology often presents itself as a source of new scientific challenges, while Mokyr (2018) emphasized the focus of scientific energy towards explanation-lacking areas, thus directing the research agenda. One of the most representative examples of this class of technological effects on science is the development of thermodynamics.

The second category is the **experimental effects**: technology is a source of empirical knowledge to be scientifically explored. Data is fundamental for scientific research. Sadi Carnot complains that he cannot generalize his findings of the driving force of heat over other mediums (solids and liquids) because "Physics as yet refuses us the necessary data" (CARNOT; THOMSON, 1897, p. 107). Much of the data that scientific research uses comes from outside the laboratory. Technological applications often provide data for scientific research and sometimes they are the only source. We propose three sub-categories that represent specific ways in which technology experimentally informs science.

The first and most general is ***data collection***. Metallurgy is a representative example in this case. Practitioners and engineers developed different alloys through trial-and-error methods (learning by doing), which later informed scientific research. In addition, metallurgy contained an experimental component that was difficult to reproduce in laboratories: the practical applications of the alloys suffered from the passage of time and deterioration, allowing "on-site experiments." In this regard, metallurgy also serves as an example of the second sub-category of experimental effect: the ***incompleteness of scientific experimentation***. In addition to metallurgy, another classic example at this point is the aircraft industry. Test wind tunnels have always been of limited use for observing all the phenomena that occur in an actual flight. Thus, the application of the technology serves as "unique observational platforms from which to observe unusual classes of natural phenomena" (ROSENBERG, 1991, p. 337). Finally, the third sub-category concerns the effect of ***hypotheses validation***. Purposely or accidentally, due to the impossibility of generating experiments in the laboratory or for other reasons, technological applications served as demonstrations of hypotheses that determined trajectories of scientific disciplines. Rutherford's denial that it would be impossible to extract energy from the atom's nucleus and the subsequent nuclear events is an illustrative example (ROSENBERG, 1982, p. 157).

Table 4.2 – 'Rosenberg Effects': categories and sub-categories of technology influence upon science

Category (effect)	Definition	Sub-category (effect)	Sub-category definition	Historical examples
Intellectual	a source of unexplained issues to be scientifically examined	-	-	Carnot and the thermodynamics (Rosenberg, 1982)
Experimental	a source of empirical knowledge to be scientifically explored	Data collection	a source of data	Metallurgy (Rosenberg, 1982)
		Incompleteness of scientific experiments	a source of on-site experimentation	Aircraft industry (Mowery & Rosenberg, 1998)
		Hypothesis validation	a source of evidence that confirms or disproves a hypothesis	Nuclear fission (Rosenberg, 1982)
Economic	a source of economic incentives that steer the scientific research	Problem-solving	a source of demand for science-based solutions to overcome technical limits	Multiple cases of feedbacks in the chain-linked innovation model (Kline & Rosenberg, 1986)
		Complementarity	a source of demand for science-based complementary innovations/adaptations	Transistor development (Rosenberg, 1982)
		Diffusion	an indication of substantial financial returns from scientific research in that area	Steel production (Rosenberg, 1982)
Instrumental	a source of new scientific "capital goods" that transforms scientific practices	Toolbox	A source of new tools for ongoing investigations	New approaches to old problems (Rosenberg, 1982)
		Frontiers'	A source of new tools for unprecedented investigations	Researchers can ask new questions and tackle new problems; Medicine diagnostics drove therapeutic research towards new paths (Rosenberg, 1992)

Source: author's own

The third category is the **economic effects**: technology is as a source of new incentives that steer scientific research. The first economic sub-category is the most common: the ***problem-solving effect***. A product has a failure of unknown causes, or it is necessary to develop a new material to replace another whose price has increased. If the engineering team cannot solve the problems with available "on the shelf" knowledge, then science comes in (KLINE; ROSENBERG, 1986). The second sub-type concerns the effects of technological complementarity, *i.e.*, the ***complementarity effect***: a technological innovation rarely settles smoothly after its development. In general, it requires numerous complementary adaptations and innovations (ROSENBERG, 1976). Occasionally, these complementary innovations are new materials, products, which generate new issues for scientific research in distinct sectors. The lesson is that "a major technological breakthrough really signals the beginning of a series of new developments of great importance, not their culmination" (ROSENBERG, 1982, p. 156). Finally, the third economic sub-category is the ***diffusion effect***: the more a product is used, the greater the potential returns (financial or social) of research. Another way to say this is to observe the advance in engineering knowledge in a specific product: "improvements in the engineering disciplines serves to raise the prospective financial payoff to research into a more purely scientific nature." (ROSENBERG; STEINMUELLER, 2013, p. 115). In the second half of the nineteenth century, steel diffusion (and engineering knowledge on steel) created strong incentives to improve scientific knowledge on that material (ROSENBERG, 1982, p. 157).

The last category is the **instrumental effects**: technology provides a new category of scientific instruments (or scientific "capital goods"[85]). In this case, we propose two main sub-categories: ***toolbox effect*** and ***frontiers' effect***. New instrumentation allows new observation techniques, new data sets, and alternative ways of testing old hypotheses (BROOKS, 1994; MOKYR, 2018; ROSENBERG, 1982). Therefore, it increases research productivity (in itself a disputed concept): "it enlarges the observational, measurement and calculating capabilities of the scientist" (ROSENBERG, 1996, p. 488). Bottom line, the ***toolbox effect*** facilitates the development of research already under development by improved means. Geomorphic research, for example, is allegedly going through productivity increases due to the incorporation of digital technologies (developed for space exploration, video game industry, and smartphone cameras) into its research practices (VILES, 2016). The second sub-category, ***frontiers' effect***, has a wider reach since it allows to "observe or measure

[85] Also referred as an 'economy of research tools.' (COCKBURN; HENDERSON; STERN, 2018)

phenomena that were previously not observable or measurable at all" (ROSENBERG, 1992, p. 382). This effect may emerge even in fields not initially related. New instruments may allow researchers to pose new questions, which even generate completely new branches of research. Therefore, when new technological instruments expand the playbook of science, they foster research that was not being developed before and became feasible only because of the new instruments[86]. The medical science agenda was deeply influenced by the advent of new instruments capable of advanced diagnoses, fostering new areas of scientific inquiry related to treatments not dominated by science (ROSENBERG, 1992). When instruments alter the possibilities of research in more than one field, they seem to fit the description of "inventions in the method of inventing" (IMI) whose vast pervasiveness "would alter the ideas/innovation production function"(COCKBURN; HENDERSON; STERN, 2018). The authors mention the invention of optical lenses and their scientific applications during the 17th century. The lenses could be classified as an IMI since their effect was "opening up the set of problems that can be feasibly addressed, and radically altering scientific and technical communities' conceptual approaches and framing of problems" (COCKBURN; HENDERSON; STERN, 2018, p. 7), not just in one field, but in several.

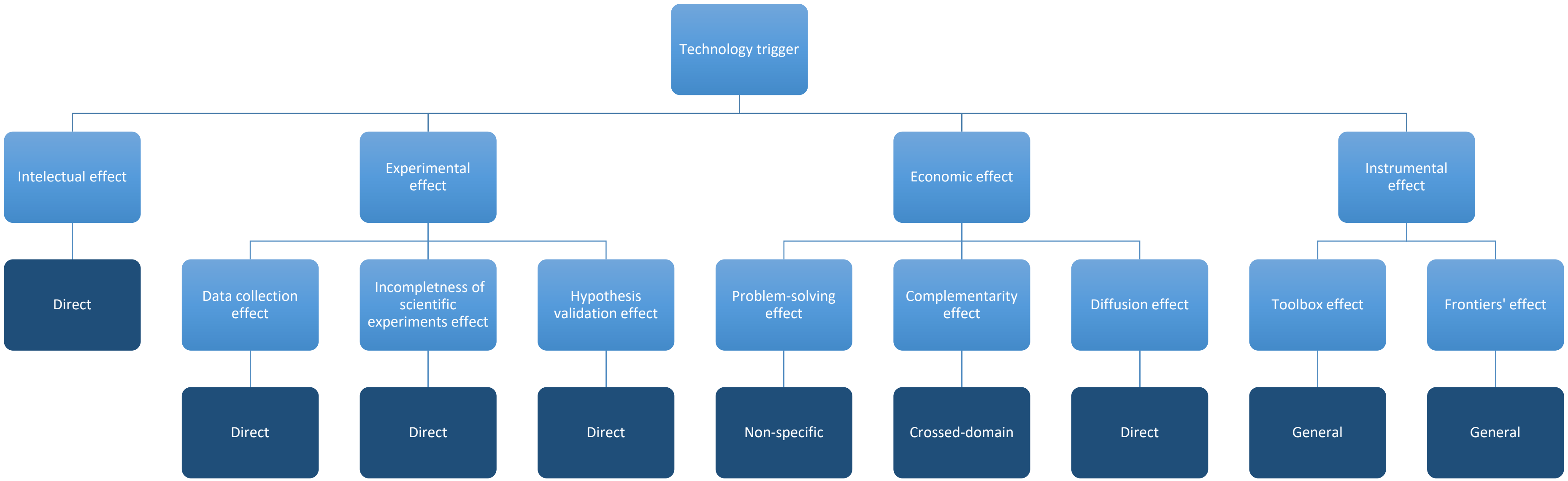
Table 4.2 sums up categories and sub-categories of the 'Rosenberg effects.' All categories constitute possible effects of technological triggers upon science. We base the denomination criterion on the original effect, not on the possible results. Historically, a technological push received by science from one of these channels has had results in different spheres of scientific activity: it generated new theories, advanced the field in which it arose, or even in fields completely apart, due to the serendipity nature (STEPHAN, 2010) of knowledge production. Likewise, a technological trigger has rarely acted exclusively within just one category. We can even raise the hypothesis that the broader the technological breakthrough, the more categories will be triggered, demanding scientific responses[87]. Figure 4.1 outlines (i) a hypothetical technological trigger; (ii) the channels through which this trigger would incur in scientific reactions; and (iii) in which "part" of the scientific sphere (in the same sector, in all sectors, etc.) these reactions would take place. Therefore, the flow of the figure starts with a generic "technological trigger". This concrete event can range from the development of a technological breakthrough (such as machine learning), to the continued use

[86] "Opening up the set of problems that can be feasibly addressed, and radically altering scientific and technical communities' conceptual approaches and framing of problems." (COCKBURN; HENDERSON; STERN, 2018)

[87] According to Brooks (1994), "the more radical the invention, the more likely it is to stimulate wholly new areas of basic research or to rejuvenate older areas of research that were losing the interest of the most innovative scientists."

of a traditional technology (such as metallurgy). The use or development of a technology then acts upon science through any one (or more than one) of the four categories of effects (economic, intellectual, instrumental and experimental) and their respective sub-categories. In addition, we provide the localization of each sub-category effect in the fabric of science. If technology triggers science in the same domain, we classify it as a direct effect *(e.g.*, AI technology fostering AI research). If technology triggers science in domains other than its origin, we classify it as a cross-domain effect (*e.g.*, AI technology fostering semi-conductor research). If technology triggers science in more than one domain, we classify it as a general effect. Finally, if technology triggers science in an unknown domain *ex-ante*, we classify it as a non-specific effect. We now start to analyze a specific case.

Figure 4.1 – Possible science responses for a generic technology trigger in terms of nature (effects, in light blue) and location (in dark blue)



Source: author's own

4.5 Machine Learning and the 'Rosenberg Effects'

4.5.1 Intellectual effects

ML algorithms (especially DL) are "black boxes" because even the experts who build them cannot explain their outputs (CARABANTES, 2020; INNERARITY, 2021; ROBBINS, 2020). Therefore, these models are riddled with epistemic opacity, also framed as opacity *as the way algorithms operate at the scale of application* (BURRELL, 2016), *inherent opacity* (DE LAAT, 2018), or *emerging non-transparency* (INNERARITY, 2021). DL opacity relates to the problem of reconstruction, *i.e.*, the explanation of the causal logic between inputs and outputs[88]. Zittrain (2019) offers a simplified version of the problem of re-construction: "most machine-learning systems don't uncover causal mechanisms. They are statistical-correlation engines. They can't explain why they think some patients are more likely to die, because they don't "think" in any colloquial sense of the word—they only answer". Since ML and DL systems seek to "mechanize reason" (PASQUINELLI; JOLER, 2020), the resulting knowledge presents itself to us as the product of an alien rationale. We know how to steer the machine logic, inducing it to deliver specific pieces of knowledge and better predictions, but we do not know why it proceeds in a specific way. For the first time in history, we have created machines that operate in a way their creators do not understand (INNERARITY, 2021), an instrument capable of applying a new logic of knowledge.

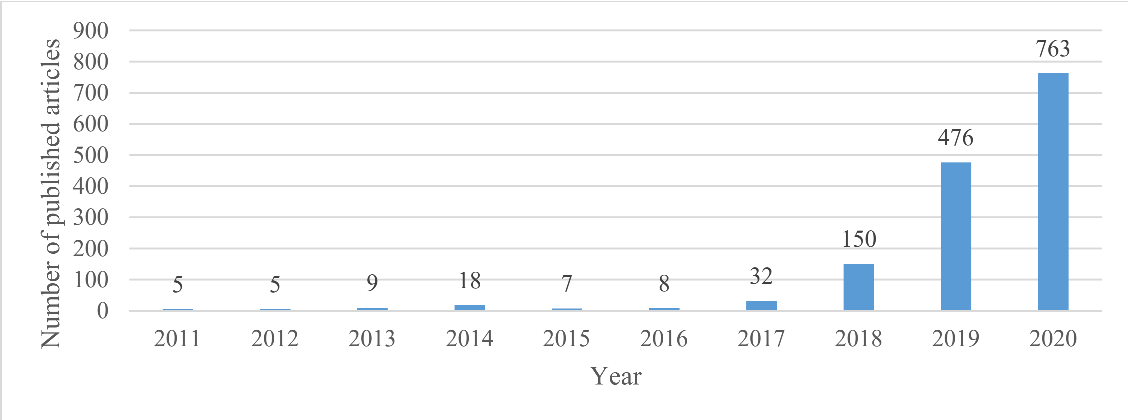
This approach of answers first and explanations later generates *intellectual debt* (ZITTRAIN, 2019). Intellectual debt accumulates when we integrate more and more ML and DL systems to support our decisions. This is more crucial as high-stakes are involved. Our inability to explain these results might lead to confidence issues when one has no prior independent evaluation. Even in trustworthy contexts, there can be malfeasance, not in an intentional sense, but emerging from the operation of the algorithms. This inadequate behavior of algorithms with respect to what is expected of them is one of the types of principal-agent problem (PAP) which happens "due to the cognitive limitation of human experts facing the immensity of big data and the opacity of machine learning algorithms" (KIM, 2020, p. 4). In addition, the system can be vulnerable to attacks that we will not know how to prevent or react to (FINLAYSON *et al*., 2019).

---

[88] "The re-construction problem is serious, because even with complete information about the operations of a system, an ex-post analysis of a specific decision may not be able to establish a linear causal connection which is easily comprehensible for human minds." (WISCHMEYER, 2020, p. 81)

To obtain an approximation of this effect on science, we searched for terms[89] related to the "epistemic opacity" of machine learning in the Scopus database. This lack of explanation has prompted a new stream of research on the explicability of artificial intelligence (BURRELL, 2016; CARABANTES, 2020; DE LAAT, 2018). We can observe an exponential growth of publications on AI opacity in the last decade. Growth accelerated from 2018 onwards when DL-based systems were already commercially established.

Figure 4.2 – Yearly output of research documents on AI opacity, Scopus (2011-2020)



Source: author's own, based on Scopus data

Figure 4.2 provides evidence that the technological breakthrough of DL motivated hundreds of researchers to look for distinct solutions to the intellectual debt that we continue to accrue. However, not all the research community agrees on the need to foster explainable AI. Geoffrey Hinton, pioneer of deep neural networks, believes we can live well with the black box of machine learning, since this is just another heuristic, like so many others, that we can't explain[90]. His view is that we should regulate AI by its performance rather than worrying about its inner workings.

[89] (TITLE-ABS-KEY("xAI" OR "XAI" OR "explainable artificial intelligence" OR "explainable AI" OR "explainable machine learning" OR "explainable deep learning" OR "explainable algorithms" OR "interpretable artificial intelligence" OR "interpretable AI" OR "interpretable machine learning" OR "interpretable deep learning" OR "interpretable algorithms" OR "opaque artificial intelligence" OR "opaque AI" OR "opaque machine learning" OR "opaque deep learning" OR "opaque algorithms" OR "responsible artificial intelligence" OR "responsible AI" OR "responsible machine learning" OR "responsible deep learning" OR "responsible algorithms" OR "transparent artificial intelligence" OR "transparent AI" OR "transparent machine learning" OR "transparent deep learning" OR "transparent algorithms") AND (LIMIT-TO (DOCTYPE,"cp") OR LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"re")))

[90] "People can't explain how they work, for most of the things they do. When you hire somebody, the decision is based on all sorts of things you can quantify, and then all sorts of gut feelings. People have no idea how they do that. If you ask them to explain their decision, you are forcing them to make up a story. Neural nets have a similar problem [...] You should regulate them based on how they perform" Geoffrey Hinton interview in (SIMONITE, 2018).

Still, the articles on Figure 4.2 try moving us from the place where we are today: “AI is now at the same stage as when the steam engine was invented, before the laws of thermodynamics necessary to explain and control its inner workings, had been discovered. Similarly, today, there are efficient neural networks for image recognition, but there is no theory of learning to explain why they work so well and how they fail so badly.” (PASQUINELLI; JOLER, 2020, p. 2).

4.5.2 Experimental effects

AI contests and competitions operate as on-site experiments with multiple functions: they facilitate ***data collection***, provide developers with suitable environments to deploy their “experiments” and enable ***hypotheses validation***. Competitions for embodied AI (robotics) had an essential role in spreading certain lines of research, as well as “increasing rigor by creating standard benchmark problems by which the performance of research can be judged” (ANDERSON; BALTES; CHENG, 2011, p. 12). Competitions “can serve to refocus our attention on algorithms that perform well, but may not otherwise receive their due attention in the research community” (GOODFELLOW *et al*., 2015, p. 62). Neural networks gave their first victorious steps in competitions as early as 1994 (SCHMIDHUBER, 2015).

ImageNet (ImageNet Large Scale Visual Recognition Challenge) competition’s history is representative of the importance of these events for the AI research community. The application and testing of supervised machine learning algorithms depends on labeled databases. Without these applications in actual databases, techniques and theories cannot pass to the stage of commercially viable products. Li Fei-Fei, a Stanford computer professor, developed the ImageNet competition by offering a sophisticated database for testing the algorithms. The results at the 2012 ImageNet edition enshrined DL technique as a major breakthrough:

> ImageNet couldn’t come at a better time for Hinton and his two students. Hinton had been working on artificial neural networks since the 1980s, and while some like Yann LeCun had been able to work the technology into ATM check readers through the influence of Bell Labs, Hinton’s research hadn’t found that kind of home. A few years earlier, research from graphics-card manufacturer Nvidia had made these networks process faster, but still not better than other techniques. Hinton and his team had demonstrated that their networks could perform smaller tasks on smaller datasets, like handwriting detection, but they needed much more data to be useful in the real world. (GERSHGORN, 2017, s/p).

Therefore, competitions and tournaments are not only testing grounds for commercial innovation but also spaces for scientific empiricism, *i.e.*, applied science (PRICE, 1984). Still, has this provoked a fundamental paradigmatic shift in AI science[91]? To observe the rise of the neural network/DL paradigm in the AI canon quantitatively, we provide a brief content analysis of scientific manuals. While article sets may contain emerging topics, manuals supposedly keep the consolidated knowledge of the field (Fleck, n.d.). *Artificial Intelligence: a modern approach* (AIMA), by Russell and Norvig, is one of the most diffused manuals of AI: it is present in more than 1500 universities, in more than 130 countries or regions[92]. Table 4.3 below shows the evolution of term count throughout the four editions of AIMA. We base each group of terms on the keywords proposed by Cockburn *et al.* (2018).

Table 4.3 – Evolution of selected AI-keyword citations (grouped under learning, symbols, and robotics) throughout the editions of Artificial Intelligence: A Modern Approach (AIMA)

Terms	**1st edition - 1995**	**2nd edition - 2002**	**3rd edition - 2010**	**4th edition - 2020**
		Learning		
Reinforcement Learning	57	76	104	200
Machine-learning	54	75	93	239
Neural Networks	102	70	56	133
Deep learning	0	0	0	131
		Symbols		
Natural language processing	37	29	14	25
Symbolic	26	33	33	35
Natural languages	19	11	13	13
		Robotics		
Computer Vision	30	38	36	68
Robotic	104	211	177	188

Source: author's own

The decreasing number of citations of neural networks (NN) until the third edition of AIMA on Table 4.3 draws attention to the priorities of the scientific community, *i.e.*, which approaches are worth teaching at the undergraduate level. By 2010, DL concepts and theories were ripe, but the majority of the scientific community still ignored them (LECUN; BENGIO; HINTON, 2015). After winning several contests and competitions and guiding the technological development of commercially viable products, neural networks received wide

[91] We understand AI science as Gazis (1979, p. 252) understands computer science/software science: "the search for knowledge, or the development of a methodology, that goes beyond satisfying the needs of a single application, but forms the basis for new applications."
[92] http://aima.cs.berkeley.edu/adoptions.html

recognition from the AI scientific community. The number of citations that neural networks (from 56 to 133) and DL (from 0 to 131) received in the last edition of AIMA corroborate this paradigmatic change. The editors mention that a whole chapter is dedicated to DL and that, in addition, “the coverage of natural language understanding, robotics, and computer vision has been revised to reflect the impact of deep learning.” (RUSSELL; NORVIG, 2020). This is evidence of the experimental effect. AI technology, tested in many competitions[93], pushed the science of AI, as proposed by our framework: stimulating changes in the AI domain itself.

4.5.3 Economic effects

The most common type of economic effect of technology on science is the ***problem-solving effect***. Engineering teams leverage the available knowledge (intra/extra firm) to solve problems. When the available knowledge is not enough, science becomes the last resort. Still, the “available knowledge shelf” has limits (KLINE; ROSENBERG, 1986). DL-based systems development faces many practical problems (KUWAJIMA; YASUOKA; NAKAE, 2020; SCULLEY *et al*., 2015), *e.g.*, the training of recurrent neural networks (RNN). These networks seemed beneficial for the sequential recognition of speech or images. However, when training the algorithms, the backpropagation technique ran into problems, since its gradients “either grow or shrink at each time step, so over many time steps they typically explode or vanish” (LECUN; BENGIO; HINTON, 2015). This practical problem of training RNN was addressed by researchers, who developed new architectures based on long short-term memory (LSTM), a component that has become crucial for the training of neural networks since then. Still, a long list of technical challenges for AI guarantees many practical problems in the years to come: AI applications often lack explainability; many learning algorithms are highly inflexible in their functionality; labels are a scarce resource, but also a precondition for many AI systems; AI systems struggle with the extraordinary; and, building secure AI applications is nearly impossible (HAGENDORFF; WEZEL, 2020).

The second sub-category of economic effects is the ***complementarity effect***. The viability of AI “will in fact require significant advances in virtually all areas of computing, including areas that are not traditionally recognized as being important to AI research and

[93] According to Russell and Norvig (2020) “Shared benchmark problem sets became the norm for demonstrating progress, including the UC Irvine repository for machine learning data sets, the International Planning Competition for planning algorithms, the LibriSpeech corpus for speech recognition, the MNIST data set for handwritten digit recognition, ImageNet and COCO for image object recognition, SQUAD for natural language question answering, the WMT competition for machine translation, and the International SAT Competitions for Boolean satisfiability solvers.”

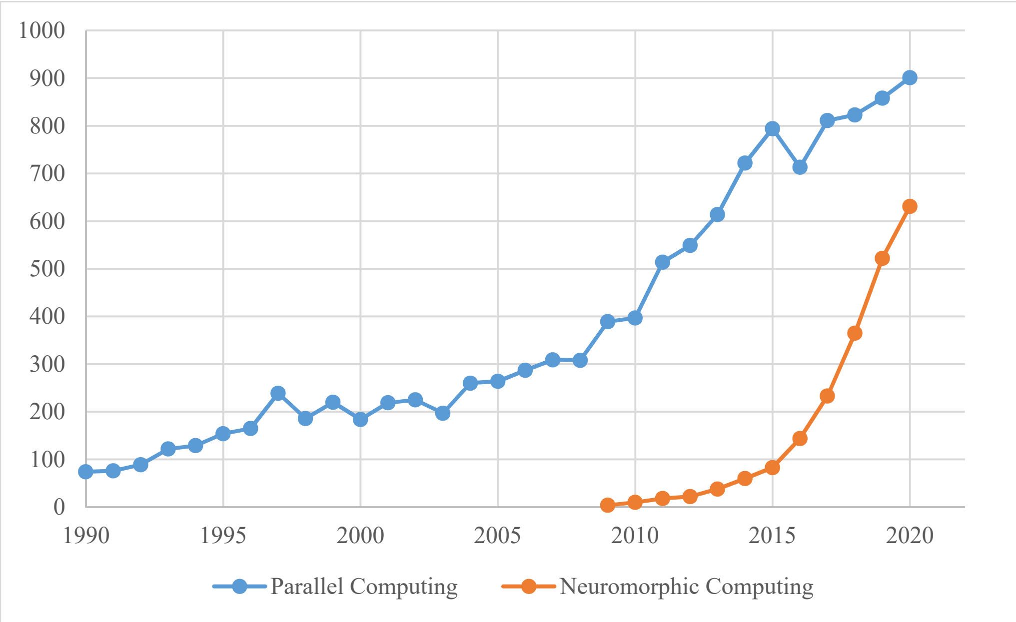
development" (HAGER *et al.*, 2017, p. 1). Concerning processing capacity[94], motivated by the establishment of DL technology, companies and countries joined a race for AI chips: "a class of microprocessor or computer system designed as co-processors to support AI applications, in particular artificial neural networks, machine vision, and machine learning" (ERNST, 2020, p. 27). These chips are directly linked to the specific processing demands of DL algorithms (LEE, 2018; RUSSELL; NORVIG, 2020). Considering this scenario, research that seeks a new horizon for microprocessors is gaining traction. The parallelism revolution has been touted for some time. Since 1990, the technical requirements of machine learning[95] pressure for a transition from sequential to parallel computing (PELÁEZ, 1990). Parallel computing research, which had been on an uptrend since the 1990s, received a very strong boost in the 2010s, most likely related to the success of ML and DL techniques.

A possible path for breaking with von Neumann's sequential architecture lies in hardware inspired by neuromorphic computing: "the hardware that mimics neuro-biological architectures to implement models of neural systems" (CHEN *et al.*, 2018). Netware (hardware for neural networks) is the hardware counterpart to what neural networks are to algorithms. Figure 4.3 demonstrates the growth of articles in the Scopus database for the terms "parallel computing" and "neuromorphic computing". Both received considerable boosts in the 2010s.

94 "Deep learning relies heavily on powerful hardware. Whereas a standard computer CPU can do 10^9 or 10^{10} operations per second. a deep learning algorithm running on specialized hardware (e.g., GPU, TPU, or FPGA) might consume between 10^{14} and 10^{17} operations per second, mostly in the form of a highly parallelized matrix and vector operations." (RUSSELL; NORVIG, 2020).

95 "Von Neumann architecture is seen by its critics as a major obstacle to good programming in general. In one area, however, the shortcomings of the conventional approach have a particular importance. This is the area of artificial intelligence." (PELÁEZ, 1990, p. 68).

Figure 4.3 – Evolution of articles on 'parallel computing' (blue) and 'neuromorphic computing' (orange) (1990 – 2020)



Source: author's own, based on Scopus data

The growth of research in alternative processing paradigms[96] is another piece of evidence of how the technological development of DL has triggered a scientific response in a crossed domain, as specified by the *Rosenberg Effects*. Through the effects of complementarity, powerful economic incentives have generated the demand for research on this alternative paradigm, mobilizing researchers in computing, physics, and biology. Although still in its early stages (ERNST, 2020), there are already several AI chips based on neuromorphic computing[97].

The ***diffusion effect*** is the third sub-category of economic effects. Its logic is simple: the more a technology spreads through the economy, the more attractive the research on that technology will be. The rapid diffusion of ML and DL algorithms in society is the result of infusing these new algorithms into pre-existing systems to improve or modify their services (ENGSTRÖM; STRIMLING, 2020). The World Intellectual Property Organization (WIPO) provides a detailed overview of DL's diffusion: considering patents and scientific articles on AI in general, "the ratio of scientific papers to inventions has decreased from 8:1 in 2010 to 3:1 in 2016 – indicative of a shift from theoretical research to the use of AI

[96] The development of NetWare (hardware for neural networks) was high on the agenda in the 1980s (VAN RAAN; TIJSSEN, 1993). However, as already mentioned, all things related to neural networks were marginalized and would only return to the spotlight after AlexNet's breakthrough in 2012.

[97] Intel researchers have released Loihi, the company's fifth generation of chips inspired by neuromorphic technologies (DAVIES *et al*., 2018).

technologies in commercial products and services” (WIPO, 2019, p. 14). The expansion of the cloud computing market is also a good *proxy* for the expansion of AI services. DL algorithms today are marketed on-demand via cloud[98]. The cloud computing market grew from $ 15.08 billion in 2010 to $ 236 billion in 2020, and Amazon, Microsoft, and Google control 60% of this market (Richter, 2021). The success of DL in 2012 “has caused most major technology companies, including Google, Facebook, Microsoft, IBM, Yahoo!, Twitter and Adobe, as well as a quickly growing number of start-ups to initiate research and development projects” (LECUN; BENGIO; HINTON, 2015).

While private participation in research decreases in other areas, in AI there is a steep growth in corporate publications from 2012 onwards: “Our estimates suggest that Fortune 500 Global technology firms are publishing 44 additional papers annually for AI conference than the counterfactual.”(AHMED; WAHED, 2020, p. 4). This increase is linked to higher participation of Big Tech companies and elite universities and a decrease in the participation of mid-tier universities. This growing imbalance is due to the complementary assets of these organizations, characteristic of the current phase of AI: the tacit knowledge of researchers from elite universities joins computational power and Big Tech’s proprietary databases. Their findings demonstrate how the diffusion of AI prompted corporations to launch a research race. Funding from public sources also increased. Through all its agencies, the “US government expected to spend almost $5 billion on unclassified AI research in 2020 [...] The UK’s Engineering and Physical Sciences Research Council gave out nearly £160 million (US$212 million) in 2020. The European Commission says public and private investment in AI should reach €20 billion (US$23 billion) across Europe by the end of 2020” (SAVAGE, 2020).

4.5.4 Instrumental effects

Within the instrumental effects, the first sub-category is the ***toolbox effects***. The ability of ML systems to increase individual productivity is one of the effects most estimated by researchers (CHUBB; COWLING; REED, 2021). Besides, “deep learning programs trained on large amounts of experimental data and chemical literature will help the emergence of novel compounds.” (KRAUS, 2020, p. 501). The effect of increasing the productivity of certain tasks is also perceived by research carried out by corporations. In the pharmaceutical

[98] See Google’s Cloud TPU: https://cloud.google.com/tpu?hl=pt-br.

industry, the use of recognition AI techniques has enhanced intramural R&D prospects, particularly for small and medium-sized enterprises (KULKOV, 2021). A paradigmatic case is an advance in the structural identification of proteins through their amino acid sequence. DeepMind (acquired by Alphabet) developed the AlphaFold algorithm: “AlphaFold greatly improves the accuracy of structure prediction by incorporating novel neural network architectures and training procedures based on the evolutionary, physical and geometric constraints of protein structures” (JUMPER, 2021, p. 584). By predicting protein structures, scientists will be able to develop new types of drugs and treatments with speed and accuracy previously impossible. For this reason, some already put the feat of AlphaFold alongside the discovery of Watson and Crick’s double helix and the CRISPR genetic editing technique developed by Doudna and Charpentier (DHAR, 2020). “AlphaFold—and computational approaches that apply its techniques for other biophysical problems—will become essential tools of modern biology” (JUMPER, 2021, p. 589).

The ***frontiers’ effect***, the second sub-category, accounts for this type of application. AI is a candidate for a general-purpose technology (GPT) and an IMI: the effect of AI is likely to have a greater reach, as it alters the logic of knowledge production, opening up new possibilities of inquiry. One of the fields in which this kind of leap in research capacity is anticipated is medicine. The field of radiology, for instance, was transformed by image analysis, but the field of pathology did not follow the same trajectory. With the advent of deep learning, the field of pathology could undergo a similar transformation in cancer treatment (COCCIA, 2020). This author raises evidence that this process is ongoing and, if organizational barriers are resolved, “deep learning technology in oncology can pave the revolution in the management of cancer based on new and effective diagnostic approaches in clinical practice to support appropriate anti-cancer treatments” (COCCIA, 2020, p. 9). Regarding the effect of ML on chemical engineering, there are some fields where the fruits on the lower branches were actually harvested: process operations, materials design. However, in other fields of this area there are important challenges, such as the quantity and quality of data, the need to develop specific ontologies for the field of chemical engineering and the explanatory capacity of ML systems (VENKATASUBRAMANIAN, 2019). These challenges should, in theory, moderate our expectations of ML’s ability to promote a significant frontiers’ effect.

4.6 Discussion

Although there are works discussing specific impacts of AI technologies on science, there is no broad and holistic view that synthesizes the channels through which this interaction occurs. The value of this article is twofold. First, it provides a simple framework to assess the relations between technology and science. Second, it applies the framework to the case of AI, providing this broad and holistic view of the influence of new artificial intelligence techniques on science. More specifically, the article details the channels through which this relationship occurs, the nature of these channels and the *loci* in which the potential effects on science unfolds.

Before addressing the ramifications of both contributions, it is necessary to recall the starting point, Nathan Rosenberg's vision of the relationship between economics, technology and science. His central question is: how exogenous is science? *i.e.*, how much is science determined by economic factors? His way out to offer an answer to this question is to frame technology as an economic variable. The movements of the technological sphere, therefore, are mediators between economic incentives and scientific reactions. That is a view, if not reductionist, at least partial of the nature of technology (and science). However, it is a useful lens to understand this relationship between a techno-economic sphere and its scientific counterpart.

Considering that technology is conditioned by social aspects and that science has its own internal dynamics, the framework that we have consolidated in this paper can be useful. The "Rosenberg effects" framework enables outlining the extent of the effects of any technology on science and to qualitatively characterize this relationship. From the perspective of studies on technology, the use of the framework enables the identification of technologies with wide pervasiveness. Technologies that activate different categories of effects on science tend to be "breakthroughs". When there is evidence of a robust direct economic effect (three economic sub-categories triggered), it is worth investigating whether the technology in question is a GPT (ROSENBERG; TRAJTENBERG, 2004). When there is evidence of a robust instrumental effect (two instrumental sub-categories triggered), it is worth investigating whether the technology in question is an IMI (COCKBURN; HENDERSON; STERN, 2018). From the point of view of science studies, the framework would lead to the reflection of what kind of science is socially desirable: a science more or less autonomous from the techno-economic sphere?

Regarding the effects of ML/DL on science, finding active triggers in all categories reinforces the thesis that ML/DL was a major technological breakthrough. These triggers are ongoing processes of adaptation and response of science to new technological conditions. This response is not automatic in any of the categories, as can be seen by the counterexamples (Hinton on the intellectual effect, the barriers to the frontiers' effect). They are mediated by the socio-economic conditions in which science is embedded. Nevertheless, the triggers are active and have ramifications for the research community of ML/DL concerned with its social impacts. We emphasize two themes that emerge at the intersection of the effects analyzed in this article.

The first is between the experimental and the economic effect. The more companies use applied AI systems due to its diffusion, more data and knowledge on these systems will emerge. This means that this highly experimental science is progressively taking place within private walls. That science takes place within companies is not new, but the degree of concentration due to the characteristics of AI (massive data sets, proprietary code and high-performance computing) might lead to a scenario of knowledge monopoly (RIKAP; LUNDVALL, 2020). A small group of companies would have access to the data that is ultimately the field of scientific experimentation of the latest AI techniques. Technologies have a proprietary character, but science, especially basic science (or blue sky) has a public character. How to deal with this entrenchment of frontier science in a few private organizations will be a challenge for States and research communities.

The other point refers to the relation between the intellectual and the instrumental effect. The multiple triggers demonstrate the magnitude of the ongoing transformation. However, as we have seen, there are huge knowledge gaps accumulating about AI systems that take a structural character not only in the economic realm, but also in the scientific realm itself (*i.e.*, in the generation of new knowledge). This implies that ML/DL-based knowledge (instrumental effects) is born with the imprint of our ignorance. This explains the urgency of moving forward on the flank already opened by those concerned with our intellectual debt (intellectual effect) and discussing algorithmic epistemology. Together, these challenges pose the question: have the developers and users of ML/DL systems taken responsibility for these possible effects on science?

Considering the limitations of the article, we base our conceptual instruments on the perspective of one specific author. A broader view can be derived if other perspectives are added, eventually revealing additional interactions between technology and science. It would be especially interesting to adopt a sociomaterial lens (ORLIKOWSKI, 2007) or the social construction of science and technology (SCOT) approach (PINCH; BIJKER, 1984). These approaches would allow investigating the role of the scientific community (as well as other social actors) in the mediation and substantiation of each effect: which social actors participate actively in the transmission of economic effects? Which social actors have fostered a scientific response to the intellectual effect? It would also be possible to observe the list of possible behaviors (acceptance, rejection, reformulation) of various social actors in the face of the effects categorized by our paper. For example, how have different research communities positioned themselves in the face of the increasingly intensive (and perhaps ostensible?) presence of the prediction paradigm (GILL, 2020) in science? Are there active resistances that question how this system of technologies impacts epistemology?

Besides, we only started to systematize S&T relations regarding ML/DL, which is only a specific technique of a specific approach of AI. Much more work needs to be done, especially since "knowledge concerning the science behind AI is sorely lacking" (YUAN *et al*., 2020, p. 993). Some of the sub-categories analyzed deserve a much deeper treatment, capable of refining the methods of measuring the impact of technology on science and of observing it qualitatively and quantitatively. This is necessary if we want to understand how paradigmatic are the changes on science lead by the effects of ML/DL.

4.7 References

AHMED, N.; WAHED, M. ***The De-democratization of AI***: Deep Learning and the Compute Divide in Artificial Intelligence Research, 2020.

ALT, R. Electronic markets on transaction costs. **Electronic Markets**, v. 27, n. 4, p. 297-301, 2017. DOI: https://doi.org/10.1007/s12525-017-0273-2.

ALBUQUERQUE, E. da M. e. Nathan Rosenberg: historiador das revoluções tecnológicas e de suas inquietações econômicas. **Revista Brasileira de Inovação**, v. 16, n. 1, p. 9-34, 2017.

ANDERSON, J.; BALTES, J.; CHENG, C. T. Robotics competitions as benchmarks for AI research. **Knowledge Engineering Review**, v. 26, n. 1, p. 11-17, 2011. DOI: https://doi.org/10.1017/S0269888910000354.

ARNOLD, C. Cloud labs: where robots do the research. **Nature**, v. 606, n. 7914, p. 612-613, 2022. DOI: https://doi.org/10.1038/d41586-022-01618-x.

BAGNOLI, V. Digital Platforms as Public Utilities. **IIC International Review of Intellectual Property and Competition Law**, v. 51. n. 8, p. 903-905, 2020. DOI: https://doi.org/10.1007/s40319-020-00975-2.

BELLO, M.; GALINDO-RUEDA, F. **Charting the digital transformation of science**: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). OECD Science, Technology and Industry Working Papers, No. 2020/03. Paris: OECD Publishing, 2020. DOI: https://doi.org/10.1787/1b06c47c-en.

BIANCHINI, S.; MORITZ, M.; PELLETIER, P. **Deep Learning in Science**. 2020. DOI: https://doi.org/10.48550/arXiv.2009.01575.

BROOKS, H. The relationship between science and technology. **Research Policy**, v. 23, p. 477-486, 1994.

BURRELL, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. **Big Data and Society**, v. 3, n. 1, p. 1-12, 2016. DOI: https://doi.org/10.1177/2053951715622512.

CARABANTES, M. Black-box artificial intelligence: an epistemological and critical analysis. **AI and Society**, v. 35, n. 2, p. 309-317, 2020. DOI: https://doi.org/10.1007/s00146-019-00888-w.

CARNOT, N. L. S.; THOMSON, S. W. **Reflections on the Motive Power of Heat**. Accompanied by An Account of Carnot's Theory. Limited: Chapman & Hall 1897.

CHEN, Y. *et al*. Neuromorphic computing's yesterday, today, and tomorrow – an evolutional view. **Integration**, v. 61, n. July 2017, p. 49-61, 2018. DOI: https://doi.org/10.1016/j.vlsi.2017.11.001.

CHUBB, J.; COWLING, P.; REED, D. Speeding up to keep up: exploring the use of AI in the research process. **Ai & Society**, p. 0123456789, 2021. DOI: https://doi.org/10.1007/s00146-021-01259-0.

COCCIA, M. Deep learning technology for improving cancer care in society: New directions in cancer imaging driven by artificial intelligence. **Technology in Society**, v. 60, p. 101198, 2020. DOI: https://doi.org/10.1016/j.techsoc.2019.101198.

COCKBURN, I. M.; HENDERSON, R.; STERN, S. **The Impact of Artificial Intelligence on Innovation** (NBER Working Paper No. 24449). 2018. Available at: http://www.nber.org/papers/w24449%0Ahttp://www.nber.org/papers/w24449.ack. Accessed: 28 mar. 2022.

COLYVAS, J. *et al*. How Do University Inventions Get Into Practice ? **Management Science**, v. 48, n. 1, p. 61-72, 2002.

CUSUMANO, M.; GAWER, A.; YOFFIE, D. **The business of platforms**: strategy in the age of digital competition, innovation and power. New York: Harper Business, 2019.

DAIBERL, C. F. *et al*. Design principles for establishing a multi-sided open innovation platform: lessons learned from an action research study in the medical technology industry. **Electronic Markets**, v. 29, n. 4, p. 711-728, 2019. DOI: https://doi.org/10.1007/s12525-018-0325-2.

DAVIES, M. *et al*. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. **IEEE Micro**, v. 38, n. 1, p. 82-99, 2018. DOI: https://doi.org/10.1109/MM.2018.112130359.

DE LAAT, P. B. Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability? **Philosophy and Technology**, v. 31, n. 4, p. 525-541, 2018. DOI: https://doi.org/10.1007/s13347-017-0293-z.

DES. PRICE, D. The science/technology relationship, the craft of experimental science, and policy for the improvement of high technology innovation. **Research Policy**, v. 13, n. 1, p. 3-20, 1984. DOI: https://doi.org/10.1016/0048-7333(84)90003-9.

DHAR, P. **AlphaFold Proves That AI Can Crack Fundamental Scientific Problems**. IEEE Spectrum, 2020.

DOSI, G. **Technical Change and Industrial Transformation** - The Theory and an Application to the Semiconductor Industry. MacMillan, 1984.

ENGSTRÖM, E.; STRIMLING, P. Deep learning diffusion by infusion into preexisting technologies – Implications for users and society at large. **Technology in Society**, v. 63, 2020. DOI: https://doi.org/10.1016/j.techsoc.2020.101396.

ERNST, D. **Competing in Artificial Intelligence Chips**: China's Challenge amid Technology War. March, 1-60. 2020.

FEALING, K. *et al*. **The Science of Science Policy**: a handbook. Stanford Business Books, 2011.

FINLAYSON, S. G. *et al*. Adversarial attacks on medical machine learning. **Science**, v. 363, n. 6433, p. 1287-1289, 2019. DOI: https://doi.org/10.1126/science.aaw4399.

FLECK, L. **Genesis and development of a scientific fact**. The University of Chicago, n.d.

FLEXNER, A. The usefulness of useless knowledge. **Harpers**, v. 179, 1939. DOI: https://doi.org/10.18601/01245996.v22n42.03.

FOCACCI, C. N.; PEREZ, C. The importance of education and training policies in supporting technological revolutions: A comparative and historical analysis of UK, US, Germany, and Sweden (1830–1970). **Technology in Society**, v. 70, n. May, p. 102000, 2022. DOI: https://doi.org/10.1016/j.techsoc.2022.102000.

GAWER, A. Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. **Innovation**, v. 24, n. 1, p. 110-124, 2021. DOI: https://doi.org/10.1080/14479338.2021.1965888.

GAZIS, D. C. Influence of technology on science: a comment on some experiences at IBM research. **Research Policy**, v. 8, n. 3, p. 244-259, 1979. DOI: https://doi.org/10.1016/0048-7333(79)90036-2.

GEELS, F. **Technological Transitions and System Innovations**. Cheltenham, Reino Unido: Editora Edward Elgar Ltda, 2005. DOI: https://doi.org/10.4337/9781845424596.

GERSHGORN, D. **The data that transformed AI research** - and possibly the world. Quartz, 2017.

GILL, K. S. Prediction paradigm: the human price of instrumentalism. **AI and Society**, v. 35, n. 3, p. 509-517, 2020. DOI: https://doi.org/10.1007/s00146-020-01035-6.

GOODFELLOW, I. J. *et al*. Challenges in representation learning: A report on three machine learning contests. **Neural Networks**, v. 64, p. 59-63, 2015. DOI: https://doi.org/10.1016/j.neunet.2014.09.005.

HAGENDORFF, T.; WEZEL, K. 15 challenges for AI: or what AI (currently) can't do. **AI and Society**, v. 35, n. 2, p. 355-365, 2020. DOI: https://doi.org/10.1007/s00146-019-00886-y.

HAGER, G. *et al*. Advances in Artificial Intelligence Require Progress Across all of Computer Science. 2017 (Issue February, pp. 1-7). DOI: https://doi.org/10.48550/arXiv.1707.04352.

HAUTAMÄKI, A.; OKSANEN, K. Digital Platforms for Restructuring the Public Sector. *In*: SMEDLUND, A.; LINDBLOM, A.; MITRONEN, L. (Eds.). Collaborative Value Co-creation in the Platform Economy. Translational Systems Sciences, vol 11. Springer, Singapore, 2018. p. 91-108. DOI: https://doi.org/10.1007/978-981-10-8956-5_5.

HODGSON, G. M. Taxonomic definitions in social science, with firms, markets and institutions as case studies. **Journal of Institutional Economics**, v. 15, n. 2, p. 207-233, 2019. DOI: https://doi.org/10.1017/S1744137418000334.

INNERARITY, D. Making the black box society transparent. **AI and Society**, v. 36, n. 3, p. 975-981, 2021. DOI: https://doi.org/10.1007/s00146-020-01130-8.

JANG, H. A decision support framework for robust R&D budget allocation using machine learning and optimization. **Decision Support Systems**, v. 121, p. 1-12, 2019. DOI: https://doi.org/10.1016/j.dss.2019.03.010.

JANG, H. Predicting funded research project performance based on machine learning. **Research Evaluation**, v. 31, n. 2, p. 257-270, 2022. DOI: https://doi.org/10.1093/reseval/rvac005.

JUMPER, J. *et al*. Highly accurate protein structure prediction with AlphaFold. **Nature**, v. 596, 2021. DOI: https://doi.org/10.1038/s41586-021-03819-2.

KATZ, M.; SHAPIRO, C. Network externalities, competition and compatibility. **The American Economic Review**, v. 75, n. 3, p. 424-440, 1985.

KIM, E. S. Deep learning and principal–agent problems of algorithmic governance: The new materialism perspective. **Technology in Society**, v. 63, p. 101378, 2020. DOI: https://doi.org/10.1016/j.techsoc.2020.101378.

KING, R. D. *et al*. Functional genomic hypothesis generation and experimentation by a robot scientist. **Nature**, n. 427, p. 247-252, 2004. DOI: https://doi.org/10.1038/nature02236.

KING, R. D. *et al*. The automation of science. **Science**, v. 324, p. 85-89, 2009. DOI: https://doi.org/10.1126/science.1165620.

KLINE, S. J.; ROSENBERG, N. An Overview of Innovation. **European Journal of Innovation Management**, v. 38, p. 275-305, 1986. DOI: https://doi.org/10.1108/14601069810368485.

KNOBEL, M. The critical role of communication in a post-truth world. **International HIgher Education**, v.100, n. Winter, p. 9-10, 2020.

KNOBEL, M.; BERNASCONI, A. Latin American Universities: Stuck in the Twentieth Century. **International Higher Education**, v. 88, p. 26-28, 2017. DOI: https://doi.org/10.6017/ihe.2017.88.9693.

KNOBEL, M.; MOHAMEDBHAI, G. Speaking Out for Science and Democracy. **International Higher Education**, n. 110, p. 5-6, 2022.

KRAUS, J. L. Can artificial intelligency revolutionize drug discovery? **AI and Society**, v. 35, n. 2, p. 501-504, 2020. DOI: https://doi.org/10.1007/s00146-019-00892-0.

KULKOV, I. The role of artificial intelligence in business transformation: A case of pharmaceutical companies. **Technology in Society**, v. 66, p. 101629, 2021. DOI: https://doi.org/10.1016/j.techsoc.2021.101629.

KUWAJIMA, H.; YASUOKA, H.; NAKAE, T. Engineering problems in machine learning systems. **Machine Learning**, v. 109, n. 5, p. 1103-1126, 2020. DOI: https://doi.org/10.1007/s10994-020-05872-w.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436-444, 2015. DOI: https://doi.org/10.1038/nature14539.

LEE, K.-F. **AI superpowers**: China, Silicon Valley, and the new world order. Houghton Mifflin Harcourt, 2018.

MOKYR, J. The past and the future of innovation: Some lessons from economic history. **Explorations in Economic History**, v. 69, p. 13-26, 2018. DOI: https://doi.org/10.1016/j.eeh.2018.03.003.

MOWERY, D. C.; ROSENBERG, N. **Paths of Innovation: technological change in 20th century america**. Cambridge University Press, 1998.

MOWERY, D. C. *et al*. Introduction: Nathan rosenberg as a founding father of the economics of innovation. **Industrial and Corporate Change**, v. 28, n. 2, p. 283-288, 2019. DOI: https://doi.org/10.1093/icc/dtz012.

MOWERY, D. C.; NELSON, R. R.; STEINMUELLER, W. E. Introduction. In honour of Nathan Rosenberg. **Research Policy**, v. 23, p. iii-v, 1994.

OECD. **Digital platforms for facilitating access to research infrastructures**. OECD Science, Technology and Industry Policy Papers, No. 49. Paris: OECD Publishing, 2017. DOI: https://doi.org/10.1787/8288d208-en.

OECD. **Fostering Science and Innovation in the Digital Age**. OECD Going Digital Policy Note, March, 2019. DOI: https://doi.org/10.1787/9789264312012-en.

OECD. **The Digitalisation of Science, Technology and Innovation**: key developments and policies. Paris: OECD Publishing, 2020. DOI: https://doi.org/10.1787/b9e4a2c0-en.

PASQUINELLI, M.; JOLER, V. The Nooscope manifested: AI as instrument of knowledge extractivism. **AI and Society**, v. 2013, 2020. DOI: https://doi.org/10.1007/s00146-020-01097-6.

PELÁEZ, E. Parallelism and the crisis of von Neumann computing. **Technology in Society**, v. 12, n. 1, p. 65-77, 1990. DOI: https://doi.org/10.1016/0160-791X(90)90029-C.

PEREZ, C. Unleashing a golden age after the financial collapse: Drawing lessons from history. **Environmental Innovation and Societal Transitions**, v. 6, p. 9-23, 2013. DOI: https://doi.org/10.1016/j.eist.2012.12.004.

RICHTER, F. **Amazon Leads $150-Billion Cloud Market**. Statista, 2021.

RIKAP, C.; LUNDVALL, B.-Å. Big tech, knowledge predation and the implications for development. **Innovation and Development**, p. 1-28, 2020. DOI: https://doi.org/10.1080/2157930x.2020.1855825.

ROBBINS, S. AI and the path to envelopment: knowledge as a first step towards the responsible regulation and use of AI-powered machines. **AI and Society**, v. 35, n. 2, p. 391-400, 2020. DOI: https://doi.org/10.1007/s00146-019-00891-1.

ROSENBERG, N. **Perspectives on Technology**. Cambridge University Press, 1976.

ROSENBERG, N. **Inside the black box**: technology and economics. Cambridge University Press, 1982.

ROSENBERG, N. Critical issues in science policy research. **Science and Public Policy**, v. 18, n. 6, p. 335-346, 1991. DOI: https://doi.org/10.1093/spp/18.6.335.

ROSENBERG, N. Scientific instrumentation and university research. **Research Policy**, v. 21, n. 4, p. 381-390, 1992. DOI: https://doi.org/10.1016/0048-7333(92)90035-3.

ROSENBERG, N. Science, technology and society. **Rivista Internazionale Di Scienze Sociali**, v. 4, n. 4, p. 479-496, 1996.

ROSENBERG, N.; STEINMUELLER, W. E. Engineering knowledge. **Industrial and Corporate Change**, v. 22, n. 5, p. 1129-1158, 2013. DOI: https://doi.org/10.1093/icc/dts053.

ROSENBERG, N.; TRAJTENBERG, M. A General-Purpose Technology at Work: The Corliss Steam Engine in the Late-Nineteenth-Century United States. **The Journal of Economic History**, v. 64, n. 1, p. 61-99, 2004.

ROSENBERG, N.; NELSON, R. American universities and technical advance in industry. **Research Policy**, v. 23, p. 323-348, 1994.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: a Modern Approach (Fourth). Pearson, 2020.

SAVAGE, N. The race to the top among the world's leaders in artificial intelligence. **Nature**, v. 588, p. S102-S104, 2020. DOI: https://doi.org/https://doi.org/10.1038/d41586-020-03409-8.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, v. 61, p. 85-117, 2015. DOI: https://doi.org/10.1016/j.neunet.2014.09.003.

SCHOT, J.; KANGER, L. Deep transitions: Emergence, acceleration, stabilization and directionality. **Research Policy**, v. 47, n. 6, p. 1045-1059, 2018. DOI: https://doi.org/10.1016/j.respol.2018.03.009.

SCHREIECK, M.; WIESCHE, M.; KRCMAR, H. Governing nonprofit platform ecosystems–an information platform for refugees. **Information Technology for Development**, v. 23, n. 3, p. 618-643, 2017. DOI: https://doi.org/10.1080/02681102.2017.1335280.

SCULLEY, D. *et al*. Hidden technical debt in machine learning systems. **Advances in Neural Information Processing Systems**, v. 2015-January, p. 2503-2511, 2015.

SIMONITE, T. **Google's AI Guru Wants Computers to Think More Like Brains**. Wired, 2018.

SMEDT, K. De; KOUREAS, D.; WITTENBURG, P. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. **Publications**, v. 8, n. 2, p. 21, 2020. DOI: https://doi.org/10.3390/publications8020021.

STEPHAN, P. The Economics of Science. *In*: HALL, B.; ROSENBERG, N. (eds.). **Handbook of the Economics of Innovation**. Elsevier, 2010. p. 217–273.

VAN BILJON, J.; MARAIS, M.; PLATZ, M. Digital platforms for research collaboration: using design science in developing a South African open knowledge repository. **Information Technology for Development**, v. 23, n. 3, p. 463-485, 2017. DOI: https://doi.org/10.1080/02681102.2017.1328654.

VAN RAAN, A. F. J.; TIJSSEN, R. J. W. The neural net of neural network research - An exercise in bibliometric mapping. **Scientometrics**, v. 26, n. 1, p. 169-192, 1993. DOI: https://doi.org/10.1007/BF02016799.

VASILESCU, D. C.; FILZMOSER, M. Machine invention systems: a (r)evolution of the invention process? **AI and Society**, v. 36, n. 3, p. 829-837, 2021. DOI: https://doi.org/10.1007/s00146-020-01080-1.

VENKATASUBRAMANIAN, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? **AIChE Journal**, v. 65, n. 2, p. 466-478, 2019. DOI: https://doi.org/10.1002/aic.16489.

VICENTE-SAEZ, R.; GUSTAFSSON, R.; VAN DEN BRANDE, L. The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world. **Technological Forecasting and Social Change**, v. 156, n. March, p. 120037, 2020. DOI: https://doi.org/10.1016/j.techfore.2020.120037.

VILES, H. Technology and geomorphology: Are improvements in data collection techniques transforming geomorphic science? **Geomorphology**, v. 270, p. 121-133, 2016. DOI: https://doi.org/10.1016/j.geomorph.2016.07.011.

WIPO. **WIPO Technology Trends 2019**: Artificial Intelligence. In Geneva: World Intellectual Property Organization, 2019.

WISCHMEYER, T. **Regulating Artificial Intelligence**. Regulating Artificial Intelligence, 2020. DOI: https://doi.org/10.1007/978-3-030-32361-5.

YU, Z.; LIANG, Z.; WU, P. How data shape actor relations in artificial intelligence innovation systems: an empirical observation from China. **Industrial and Corporate Change**, p. 1-17, 2021. DOI: https://doi.org/10.1093/icc/dtaa063.

YUAN, S. *et al.* Science behind AI: the evolution of trend, mobility, and collaboration. **Scientometrics**, v. 124, n. 2, p. 993-1013, 2020. DOI: https://doi.org/10.1007/s11192-020-03423-7.

ZITTRAIN, J. **The Hidden Costs of Automated Thinking**. The New Yorker, 2019.

4.8 Annex 1

The *corpus* consists of 32 documents distributed over fifty years of the author's scientific production (1963-2013). The criterion for defining the *corpus* was to include the principal works by Rosenberg (ALBUQUERQUE, 2017; MOWERY; NELSON; STEINMUELLER, 1994). In addition to these works, we included documents (preferring, when possible, peer-reviewed articles) dealing with the relationship between S&T, seeking to balance works from different phases of the author's scientific production. Our fundamental unit of analysis is the historical examples regarding the influence of technology on science, regardless of the nature of this influence at first. We consider as a codifiable unit of analysis specific historical examples that mention any effect of technology on science. An example is "the agricultural experiment stations conducted research aimed at improving the productivity of all types of agriculture enterprise and attempting to solve problems of plant and animal disease" (ROSENBERG; STEINMUELLER, 2013). Generic mentions of these same effects are not considered units of analysis; therefore, they were not codified. An example of generic mention is "The problems encountered by sophisticated industrial technologies and the anomalous observations or unexpected difficulties they produced as powerful stimuli to scientific research in the academic community" (MOWERY; ROSENBERG, 1998). A unit of analysis can receive more than one code; however, we sought to minimize this multiple encoding. We searched and mapped units of analysis in the corpus until reaching a point of apparent saturation. Table 4.4 sums up the encoding results.

Starting from the initial analysis (ROSENBERG, 1982) and following terminological guidelines (HODGSON, 2019), whenever possible we adopted the terms and definitions already proposed by Rosenberg. We propose new categories and sub-categories when we find units (historical cases) that do not fit into any sufficiently defined *ex-ante* category; one that stands out is the intellectual effect, absent from the *corpus* in the meaning we adopt here.

Table 4.4 – Code application per categories (4) and sub-categories (8) of effects

Year of publication, Document	Economic effects				Experimental effects				Instrumental effects			Intellectual Effect	Totals
	Complementarity	Diffusion	Problem-solving	Subtotal	Data collection	Hypothesis validation	Incompleteness	Subtotal	Frontiers'	Toolbox	Subtotal		
2013, Engineering knowledge	0	3	0	3	1	0	1	2	0	0	0	1	6
2009, The economic impact of scientific instrumentation developed in academic laboratories	0	0	3	3	0	0	0	0	1	3	4	0	7
2009, Some critical episodes in the progress of medical innovation	0	0	0	0	0	0	0	0	1	2	3	0	3
2004, A GPT at work: the Corliss Steam Engine	0	0	0	0	0	0	0	0	0	0	0	0	0
1998, The role of electricity in industrial development	0	0	0	0	0	0	0	0	0	0	0	0	0
1998, Paths of Innovation	0	3	0	3	0	0	0	0	0	0	0	0	3
1996, Science, technology and society	1	3	8	12	0	0	0	0	0	0	0	2	14
1994, The dynamics of technological change in medicine	0	0	0	0	0	0	0	0	0	1	1	0	1
1994, How the developed countries became rich	0	0	2	2	3	0	0	3	0	0	0	0	5
1993, American universities and technical advance in industries	0	1	8	9	0	0	0	0	0	0	0	0	9
1992, Scientific instrumentation and university research	1	3	4	8	1	0	0	1	4	7	11	0	20
1991, Critical issues in science policy research	0	0	9	9	0	0	1	1	0	2	2	0	12
1990, Why do firms do basic research	0	0	4	4	0	0	0	0	0	0	0	0	4
1990, Science, technology and the western miracle	0	0	0	0	0	0	0	0	0	0	0	0	0
1988, Why are Americans such poor imitators	0	0	0	0	0	0	0	0	0	0	0	0	0
1982, How Exogenous is Science?	1	7	11	19	4	2	3	9	0	11	11	8	47
1981, Technical Change in the Commercial Aircraft Industry, 1925-1975	0	1	0	1	1	0	0	1	0	0	0	0	2
1980, The economic implications of the VSLI revolution	0	0	0	0	0	0	0	0	0	0	0	0	0
1979, The influence of market demand upon innovation	0	0	1	1	0	0	0	0	0	0	0	0	1

1979, Technological Interdependence in the American Economy	0	0	0	0	0	0	0	0	0	0	0	0	0
1977, American technology: imported or indigenous	0	0	0	0	0	0	0	0	0	0	0	0	0
1976, Perspectives on Technology	1	0	0	1	0	0	0	0	0	4	4	0	5
1976, On technological expectations	0	0	0	0	0	0	0	0	0	0	0	0	0
1974, Science, invention and economic growth	1	0	0	1	0	0	0	0	0	4	4	0	5
1974, Karl Marx on the Economic Role of Science	0	1	1	2	4	0	0	4	0	0	0	0	6
1973, Innovative responses to material shortages	0	0	0	0	0	0	0	0	0	0	0	0	0
1972, Factors affecting the diffusion of technology	1	0	0	1	0	0	0	0	0	0	0	0	1
1971, Technology and the environment: an economic exploration	0	0	0	0	0	0	0	0	0	0	0	0	0
1970, Economic development and the transfer of technology	0	0	0	0	0	0	0	0	0	0	0	0	0
1969, The direction of technical change: inducement mechanisms and focusing devices	0	0	0	0	0	0	0	0	0	0	0	0	0
1963, Technological change in the machine tool industry	0	0	0	0	0	0	0	0	0	0	0	0	0
1963, Capital goods, technology and economic growth	0	0	0	0	0	0	0	0	0	0	0	0	0
Totals	6	22	51	79	14	2	5	21	6	34	40	11	**151**

Source: author’s own

CONCLUSION

Synthesis and contribution of the thesis

This work investigated the consolidation of the system of digital technologies in the scientific field. The perspective adopted, based on the evolutionary economics and economic history of science and technology literature, is one of long-term transformations, which are informed by exchanges, intersectoral mimicry and contextual adaptations. This approach emphasizes the pervasiveness of epoch-making techno-economic paradigms. However, far from being deterministic, this approach considers the contextual/sectoral specificities in the diffusion of technology and the role of institutions in the absorption and innovative co-evolution.

The first part of the thesis, “Long-term technological and institutional transformations” historically contextualized the 2010-2020’ technology system in relation to other informational technological systems that preceded it. By understanding the digital technological system as part of a deeper and more significant transformation, policies can facilitate institutional recomposition in the market and beyond. The streams of technical advances listed in the first chapter (digitization, algorithmization and platformization) provide us with a sketch of a map of possible and plausible technological and organizational trajectories through which spheres such as the State and Science will transform. How concrete this will be is a context-dependent question: how will each national state face, react to, or plan for platformization? How will the scientific community interpret the algorithmization of scientific activity?

In addition, the periodization proposed in chapter 1 inserts the 2010-2020 decade as an integral part of a longer movement, which has unfolded since the 1970s/1980s. By elucidating the technical and institutional relationships between the technological systems that have been added and converged since that date, it is argued that the Third Industrial Revolution (or the fifth perezian “great surge of development”) is in full swing. The voices claiming that we have entered an alleged Fourth Industrial Revolution certainly overlook the fact that (i) revolutions of this size take many decades to establish and supplant old paradigms; and that (ii) institutional recomposition accompanies the technological revolution.

We can recover the questions stated in the introduction to this thesis: *why, at the same time, do different sectors seem to be invaded by attempts to digitize information or to algorithmize practices? Is this a unique phenomenon with common origins? Are we going* *through a Fourth Industrial Revolution or a special phase of a longer process?* Different sectors seem invaded by the same approach, because this is part of the consolidation of organizational and technological best-practices of an era, or in Perezian terms, of a techno-economic paradigm. Hence the perception that the second decade of the 21st century witnessed a *platformania* (CUSUMANO; GAWER; YOFFIE, 2019). This process, in which trajectories become more predictable, is part of a larger phenomenon and marks the consolidation phase of the Third Industrial Revolution. In the first chapter we explained what we mean by “digital technologies”. We elaborate on the main technologies and explain why and how they form a system. Finally, it was possible to achieve specific objective 1 (SO1): to establish a reference on relevant digital technologies with high transformative potential for organizations, through a long-term view of technological and organizational innovation standards.

The second part of the thesis investigates the platformization of Science. The historical approach allows us to verify the participation of the scientific community in the process of gestation of the digital technological system. Certain pioneering platforms in the scientific sphere preceded their private counterparts. It should be noted that the State acted as a funder or executor of some of these scientific initiatives. This is further evidence that the construction of a techno-economic paradigm, as well as technological innovations, depends to a great extent in the role played by other actors in addition to the typical private entrepreneurs. Perez has already pointed to the “under-recognition of the crucial role played by State in the funding and promotion of nanotechnology, biotechnology and the whole information revolution” (PEREZ, 2013, p. 18). We have seen that not only at the technological level, but also at the organizational level, the State or non-market actors such as scientific organizations contribute to innovation.

It should be noted that the scientific digital platforms come in different flavors. There are some that provide useful, additional services, therefore they are well characterized as tools. Crowdwork platform that facilitate the application of surveys are a good example[99]. However, some other platforms configure a new generation of scientific infrastructures. This new generation of scientific infrastructure demands, from science and technology policy managers, an attentive look at the position of platforms and possible competitors; data flows; interactions between platforms with different objectives and forms of governance. Such

99 SociaLab , the brazilian example of scientific digital platform showed in the introduction, characterizes as an auxiliary tool: it connects research laboratories across the country and reallocates reagents and cells, avoiding waste, reducing expenses and researchers’ waiting time for supplies.

infrastructures run the risk of being monopolized by the private sector, due to their competitive dynamics and the political-economic trend of privatization of infrastructures. Especially pioneering scientific platforms, because they are attached to a trajectory, may suffer to adapt to new technical possibilities and new organizational needs. This positional battle of scientific digital platforms occurs not only between different forms of governance (private, associative, public), but also across the phases of scientific research: publication, experimentation, outreach, etc. In other words, platformization affects the organization of scientific activity widely. Given the above, specific objective 2 (SO2) was reasonably addressed: to elucidate the organizational transformations in science permeated by new digital technologies, by proposing a historically-informed taxonomy of these digitized scientific organizations.

The third part of the thesis investigated the algorithmization of science. Chapter four provides a framework for understanding how machine learning and deep learning algorithms trigger answers from science. This influence occurs on several levels. At the directly economic level, AI reorients the scientific research agenda towards problem solving and to financially harness the diffusion of technology. At the experimental level, AI competitions resemble applied science. At the instrumental level, AI techniques offer new tools for scientists. And at the intellectual level, AI offers an enormous challenge for epistemology: what kind of knowledge is this generated by algorithms and where does it stand in our hierarchy of knowledge? How does it complement or displace our traditional sources of knowledge?

The instrumental and intellectual triggers, especially, pose the question: how are scientific and social institutions going to restructure to absorb this technology? It is unclear how these tools will alter traditional institutions of science, such as the authorship of a scientific discovery, or the validity of a non-algorithmic scientific assertion. The more the paradigm consolidates, the more important the resolution of the intellectual deficit becomes. Soon, regulatory agencies will be regulating via algorithms; governments will be ruling via algorithms; and social life will be being ordered, mostly, via algorithms. Therefore, to understand the nature of the relationship between science and digital technology in terms of its directionality, location and impact, through a case study of a core digital technology, i.e., specificic objective 3, was thoroughly addressed in chapter 3.

In view of the findings of parts 2 and 3 of the thesis, we can recover other subsidiary questions that were listed in the introduction to this thesis: *Have digital technologies caught up with science just now? Is science a loci for the application or* *development of these technologies (user or innovator)? Do the latest digital technologies only affect practices on the "factory floor" of science, or do they affect higher levels in terms of the principles and values of the scientific community?*

Digital technologies have not reached the scientific sphere just now. The scientific community was one of the forerunners in the use of the internet and in the development of digital platforms. Besides, the scientific sphere is not just a recipient of innovations: it also innovates and there is an exchange between the market and the scientific sector (COLYVAS *et al.*, 2002; ROSENBERG; NELSON, 1994). Finally, digital technologies enable changes not only on the science "factory floor". It's not just robots automating bench work (KING *et al.*, 2004, 2009). The consolidation of the digital technological system poses high-level challenges for science, in relation to its more traditional institutions, such as authorship and the validity of knowledge.

In general, our investigation allowed us to find some similarities and differences between the processes of absorption of digital technologies between the private sector and science. Successful scientific platforms have understood that they need to provide value to participants, just as their private counterparts do. They also offer reductions in search, information and transaction costs (ALT, 2017); new ways of carrying out traditional activities, such as crowdwork platforms for conducting surveys; entirely new services, such as recommending authors and articles based on individual preferences that academic social networks are able to offer via AI. However, scientific platforms have characteristics that distance them from the context in which private platforms operate. This is observed in the case of the Lattes platform and its enforcement capacity over its users; citizen science platforms do not respond to the same incentives as commercial platforms: while they also benefit from network externalities (KATZ; SHAPIRO, 1985), they do not have an incentive to grow above all else. The interactions between the parties that make up a scientific platform ecosystem are different when they are associated with the State or the scientific community: the motivation that binds the participantes is non-commercial, so there is no competition in commercial terms. Different governance styles (BAGNOLI, 2020; DAIBERL *et al.*, 2019; SCHREIECK; WIESCHE; KRCMAR, 2017; VAN BILJON; MARAIS; PLATZ, 2017) generate different platform dynamics.

Considering these differences, the organizational isomorphism mentioned by Perez (2002) must be nuanced. There is an emulation of technologies and organizational models, but there are considerable differences related to the objectives and the distinct nature of the social spheres. The following warning, regarding the transposition of these trajectories to the State, is also valid for science:

> In order to apply the platform concept to the public sector, an experimental approach is needed; public platforms cannot be built by transposing mechanical models from the private sector to the public sector, because the market logic of public services is quite different than open markets. (HAUTAMÄKI; OKSANEN, 2018, p. 91).

A precise answer to the question of isomorphism requires a more detailed investigation into the processes of digital transformation in science and other domains.

Finally, we return to our central research question: **How has science interacted with the digital technologies and how did it transformed itself in this process?** Understanding this broad research question involves linking together what we learn from each chapter, with each subsidiary research question. First of all, we can say without hesitation that science has been, is and will continue to be very close to frontier technological developments. Digital technology had many of its principles - principles that have become paradigms of the "platform economy" - gestated and tested in science spaces. It is necessary to recognize this so that science policies do not conceptualize a science unaware of digital technologies. It should also be noted that the interaction between the scientific community and digital technologies often takes place alongside partnerships with the state or the private sector. The development of e-portfolios or grid computing are examples of the involvement of multiple actors. Therefore, science policy must recognize that, irrespective of the form and objective of bringing digital technologies closer to science, a multisectoral approach has, at least in historical terms, a wide advantage over an isolated approach. Finally, we must emphasize that currently the transformation of science lags behind the transformation of other areas, such as the private sector, in the use of digital technology. We can infer this from the difficulty of some legacy scientific systems to absorb the potential of new technologies, but also from the institutional resistance that is put in front of more radical transformations. This fact is not necessarily bad: as we have seen, allowing the reformulation of scientific logic via algorithmization raises numerous issues, such as algorithmic opacity, intellectual debt, and many other issues that our time constraint prevented us from addressing (such as data usage and issues associated with privacy). The broad organizational redesign also poses challenges in terms of controlling the digitized scientific infrastructure: an "opening up" to the market can accelerate transformation processes, but at what cost? Science policy must consider how to balance digital transformation with the associated risks.

Although the technologies and organizational principles are clear, there is room for maneuver to define how and why to mobilize such tools. This room for maneuver is constrained by the pre-existing institutional framework in each case. Habits and customs of scientists, deans, directors of funding agencies and science policy managers will influence the absorption of digital technologies by science: “Each technological revolution is received as a shock, and its diffusion finds strong resistance both in established institutions and in the people themselves” (PEREZ, 2002, p. 23). For science, now is the time to rethink its institutional foundations so that digital technologies enable a collectively prosperous future. Science policy has a major task ahead.

Research limitations

A limitation of our investigation concerns the detailed role of institutions in the transformation of science before the digital technological system. Which institutions are more important in this case: formal institutions (laws and norms) or informal institutions (habits and customs)? Is there greater resistance to adaptation than in the private sector? A more detailed institutional analysis will be necessary to assess these issues.

There are limitations inherent to our methodological choice based in case studies: the difficulty of generalizing the findings. The unavailability of data on the subject contributes to this, as well as the idiosyncrasy in the adoption of digital technologies by the organizations in this field.

A very important limitation is the lack of treatment of datafication in this thesis[100]. The issue of datafication, the use, manipulation, storage of scientific data, is intrinsically related to the themes addressed in this thesis. In fact, datafication is one of the three innovation streams mentioned in the first chapter, alongside algorithmization and platformization. There are numerous questions shall become increasingly relevant that pertain to the interface of data and science: what should the legal nature of data be? Would it be necessary to define public data to allow access by researchers and scientific advances? How to balance the right to privacy with the imperative of scientific research? How to securely store scientific data? However, due to time constraints, it was not possible to include a specific analysis of the data dimension.

[100] Thanks to professor Marcelo Knobel for highliting this gap.

A crucial limitation concerns the intentionality of the adoption of digital technologies by the scientific organizations. Although we have mapped several initiatives, it is not clear why decision makers chose to follow the digitalization path. By not incorporating these issues, the work implicitly assumes that decision-making to innovate naturally follows the trajectory of the digital technological system. This lack of attention to agents who have the power of choice is one of the recurring criticisms of the Perezian framework (GEELS, 2005; SCHOT; KANGER, 2018).

Close to this issue is the problem of public convenience: is it convenient to digitize the entire fabric of science? Or are there areas that would be better off without the "readability" that digital allows? As for the convenience of digitizing scientific practices, the fourth chapter touches on this discussion when it points to the epistemological risks of basing knowledge generation on inexplicable artificial intelligence techniques. But further progress would be needed at this point. This problem is intrinsically linked to the nature of the TEP: it becomes a heuristic exclusion mechanism, *i.e.*, all problems are solved with the same set of best practices, even if in some cases this is not the best decision[101].

Finally, there is an occidental bias to our research[102]. Most of our examples are of occidental platforms, or occidental uses of technology. The evolutionary path of these technologies and its interaction with science might assume a totally different configuration in oriental countries, China being one of the most important. Since these countries offer a distinct institutional environment, science might just be taking a different configuration – possibly more advanced in terms of use and generation of digital technologies than its counterparts in the occident.

Possible paths to a research agenda

The research agenda that unfolds from this initial investigation is large. We can divide the possible avenues of research into two: one continues to develop the macro view, directly connected with what was attempted in this thesis. This avenue encompasses broad themes associated with a paradigm shift in science. The second avenue of research is focused on specific elements of the scientific system: universities, research institutes and science policy.

[101] "these favorable conditions become a powerful exclusion mechanism for all possible innovations that are incompatible or not well geared to the existing framework." (PEREZ, 2002, p. 28).
[102] Thanks to professor Marcelo Knobel for bringing up this issue during the thesis defence.

What are the broad themes associated with a paradigm shift in science? Future research axes can be derived from the elements that, according to Freeman and Perez (1988), are part of a restructuring of the entire productive system, making the necessary reservations for the scientific system: “a new ‘best-practice’ form of organisation”, “a new skill profile in the labor force”, “a particular wave of infrastructural investment designed to provide appropriate externalities throughout the system.”

Digital platforms emerge as a strong candidate to become the predominant organizational model (GAWER, 2021). Understanding how this model adapts to each instance of the scientific universe is a possible investigative path: have universities already used digital platforms? Research on the best ways to use digital platforms in science is in its infancy (REICHENBACH; EBERL; LINDENMEIER, 2022). Or maybe research institutes should have more organizational flexibility? It would also be interesting to observe which elements foster organizational learning in scientific organizations: would it be the qualification of managers, or the hiring of professionals with experience in organizational restructuring in other contexts?

The qualification of the workforce, which is a general issue (FOCACCI; PEREZ, 2022), has its own developments in the scientific context. Throughout the thesis we mentioned some areas of research in which algorithmization is present. There are great chances that this process will deepen in the coming years. How to qualify a new scientific workforce? How to make scientists also “data scientists”? Is this necessary in the first place? Can a division of scientific labor be feasible, with a new class of researchers trained in digital technologies and distributed across research areas? Do we run the risk of digital technologies being used to precarize the scientific workforce, in a scientific version of the gig economy (KWOK, 2017)?

Finally, a new generation of infrastructure is built to support the full development of the new paradigm. In addition to the past generation of cyberinfrastructure, there is a movement towards the formation of digital infrastructure tuned to new techno-organizational trajectories. The European Union put in place the European Open Science Cloud[103] (EOSC), to integrate continental research in an interoperable and accessible way. Its scale and scope are so comprehensive that EOSC is called a hyperinfrastructure (SMEDT; KOUREAS; WITTENBURG, 2020). Brazil has an extensive network of scientific infrastructure (NEGRI; SQUEFF, 2016). It is noteworthy that there is no research that identifies how it will be

[103] https://marketplace.eosc-portal.eu/

affected by digitization. Elsewhere, individual, dedicated labs have been centralized into automated labs, in which researchers can perform experiments remotely on demand (ARNOLD, 2022). Which infrastructural units in Brazil may be affected by the outsourcing allowed by cloud computing platforms? What are the updating needs that these infrastructures demand? Or yet, how to transform the threat of the private sector (of dominating the platformed scientific infrastructure) into productive partnerships? We don't know, and this is a central theme for 21st century science policy (BELLO; GALINDO-RUEDA, 2020; FEALING *et al*., 2011; OECD, 2017, 2019, 2020).

In the second avenue of proposed research, it would be possible to study the digital transformation of specific elements of the science system. For instance, how have universities been digitally transformed? Knobel and Bernasconi alert to the "obsolescence of the governance structures and practices of most universities that hinders the development of new thinking."(KNOBEL; BERNASCONI, 2017, p. 27). Even so, the authors call for the transformation of universities (which goes beyond mere digital transformation, it should be clarified), after all "While the most prestigious public and private universities (usually the oldest) represent the small component of each national system , what happens in them, with them, and to them has critical relevance to the [scientific] system as a whole" (KNOBEL; BERNASCONI, 2017, p. 27). There are a number of paths and actions that universities can take to take advantage of the potential of this new paradigm, *e.g.*, "open data sharing, open access publishing, open repositories, open physical labs, participatory design, and transdisciplinary research platforms" (VICENTE-SAEZ; GUSTAFSSON; VAN DEN BRANDE, 2020, p. 1).

Universities can establish general digital transformation plans. One way of trying to understand how best to go through this transition may be to look at the internal and external innovations that the university can develop. Internal innovations would transform the management of universities, from their hiring systems to the way they carry out their selection processes. It doesn't take much imagination to see how digitization, platformization and algorithmization could contribute to increasing the efficiency of existing services and even creating new services. Universities could also implement external innovations, for example, inter-organizational platforms that bring together the national university system; or yet, develop new digital tools for university extension; or how to use digital technologies to improve university business incubation activities (CHAN; KRISHNAMURTY; SADREDDIN, 2022); or even, how to engage in public debate and communication in the digital age in a more assertive way given the urgency of the university's voice to be present in the arena of public debate (KNOBEL, 2020; KNOBEL; MOHAMEDBHAI, 2022).

Another important element of the scientific system is science policy. It is possible to change science policy practices through data-based assessment (FEALING *et al*., 2011); it is also possible to foresee that financing mechanisms will be more automated and investment portfolios informed by data, whether at ministerial or innovation promotion agency levels (JANG, 2019, 2022).

Finally, this thesis brings anguish and relief at the same time. The anguish results from the observation that the digital technological system brought the technical elements that were missing for the consolidation of the informational age. By enabling this technical transformation, institutional stress is incurred for a few years in various social spheres and in less time than in past technological revolutions (BALDWIN, 2019). Currencies go digital, but then financial rules must adapt; digital platforms reach health and so it is necessary to legislate on patient data rights; Algorithms drive our cars (an eternal promise?), but then the traffic code becomes a dead letter – and whose accident victims will demand culpable accountability? The penal code should be updated with a view to blaming the algorithms. This process is "painful" not only because we are accommodated, rooted in traditional institutions. It generates stress because institutions crystallize political interests and ideologies (STRACHMAN, 2002). Its redefinition, therefore, does not only follow the criteria of what is technically efficient, but involves political and ideological positions that are often conflicting.

In addition to legal professionals (who will have plenty of work), not many citizens are excited about the prospect of institutional instability that will accompany us for years to come. The relief, on the other hand, comes from the historical reading made for moments similar to this one, of institutional transformation: after the troubled moments of institutional recomposition, society enjoys a more peaceful and prosperous time. According to Perez (2009, p. 200): "The mutual adaptation of technology and society through the social learning of the paradigm and the adaptive redesign of the institutional framework allows us to reap the maximum benefit from the wealth creation potential contained in each great wave". Hopefully, after these painful adjustments, science will emerge even stronger and more capable than it has been in the past centuries.

REFERENCES

ALT, R. Electronic markets on transaction costs. **Electronic Markets**, v. 27, n. 4, p. 297-301, 2017. DOI: https://doi.org/10.1007/s12525-017-0273-2.

ARNOLD, C. Cloud labs: where robots do the research. **Nature**, v. 606, n. 7914, p. 612-613, 2022. DOI: https://doi.org/10.1038/d41586-022-01618-x.

BAGNOLI, V. Digital Platforms as Public Utilities. **IIC International Review of Intellectual Property and Competition Law**, v. 51. n. 8, p. 903-905, 2020. DOI: https://doi.org/10.1007/s40319-020-00975-2.

BALDWIN, R. **The globotics upheaval**: globalization, robotics and the future of work. New York, NY: Oxford University Press, 2019.

BELLO, M.; GALINDO-RUEDA, F. **Charting the digital transformation of science**: Findings from the 2018 OECD International Survey of Scientific Authors (ISSA2). OECD Science, Technology and Industry Working Papers, No. 2020/03. Paris: OECD Publishing, 2020. DOI: https://doi.org/10.1787/1b06c47c-en.

BIANCHINI, S.; MORITZ, M.; PELLETIER, P. Artificial intelligence in science: An emerging general method of invention. **Research Policy**, v. 51, n. 10, p. e104604, 2022.

CHAN, Y.; KRISHNAMURTY, R.; SADREDDIN, A. Digitally-enabled university incubation processes. **Technovation**, p. e102560, 2022.

COLYVAS, J. *et al*. How Do University Inventions Get Into Practice? **Management Science**, v. 48, n. 1, p. 61-72, 2002.

CUSUMANO, M.; GAWER, A.; YOFFIE, D. **The business of platforms**: strategy in the age of digital competition, innovation and power. New York: Harper Business, 2019.

DAIBERL, C. F. *et al*. Design principles for establishing a multi-sided open innovation platform: lessons learned from an action research study in the medical technology industry. **Electronic Markets**, v. 29, n. 4, p. 711-728, 2019. DOI: https://doi.org/10.1007/s12525-018-0325-2.

DRECHSLER, W.; KATTEL, R.; REINERT, E. **Techno-Economic Paradigms**: Essays in Honour of Carlota Perez. Anthem Press, 2009.

FEALING, K. *et al*. **The Science of Science Policy**: a handbook. Stanford Business Books, 2011.

FOCACCI, C. N.; PEREZ, C. The importance of education and training policies in supporting technological revolutions: A comparative and historical analysis of UK, US, Germany, and Sweden (1830–1970). **Technology in Society**, v. 70, n. May, p. 102000, 2022. DOI: https://doi.org/10.1016/j.techsoc.2022.102000.

FRANK, A. G.; DALENOGARE, L. S.; AYALA, N. F. Industry 4.0 technologies: Implementation patterns in manufacturing companies. **International Journal of Production Economics**, v. 210, p. 15-26, 2019. DOI: https://doi.org/10.1016/j.ijpe.2019.01.004.

FREEMAN, C., PEREZ, C. Structural crises of adjustment: business cycles and investment behavior. *In:* DOSI, G. *et al.* (ed.). **Technical change and economic theory**. London, New York: Printer publishers, 1988. p. 38-66.

FURTADO, A. Long term scenarios for Latin America and the new technological revolution. **History and Technology**, v. 3, n. 1, p. 1-23, 1986. DOI: https://doi.org/10.1080/07341518608581659.

GAWER, A. Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. **Innovation**, v. 24, n. 1, p. 110-124, 2021. DOI: https://doi.org/10.1080/14479338.2021.1965888.

GEELS, F. **Technological Transitions and System Innovations**. Cheltenham, Reino Unido: Editora Edward Elgar Ltda, 2005. DOI: https://doi.org/10.4337/9781845424596.

HAUTAMÄKI, A.; OKSANEN, K. Digital Platforms for Restructuring the Public Sector. *In*: SMEDLUND, A.; LINDBLOM, A.; MITRONEN, L. (Eds.). Collaborative Value Co-creation in the Platform Economy. Translational Systems Sciences, vol 11. Springer, Singapore, 2018. p. 91-108. DOI: https://doi.org/10.1007/978-981-10-8956-5_5.

JANG, H. A decision support framework for robust R&D budget allocation using machine learning and optimization. **Decision Support Systems**, v. 121, p. 1-12, 2019. DOI: https://doi.org/10.1016/j.dss.2019.03.010.

JANG, H. Predicting funded research project performance based on machine learning. **Research Evaluation**, v. 31, n. 2, p. 257-270, 2022. DOI: https://doi.org/10.1093/reseval/rvac005.

JUSTMAN, M.; TEUBAL, M. Technological infrastructure policy (TIP): creating capabilities and building markets. **Research Policy**, v. 24, p. 259-281, 1995.

KATZ, M.; SHAPIRO, C. Network externalities, competition and compatibility. **The American Economic Review**, v. 75, n. 3, p. 424-440, 1985.

KING, R. D. *et al*. Functional genomic hypothesis generation and experimentation by a robot scientist. **Nature**, N. 427, p. 247-252, 2004. DOI: https://doi.org/10.1038/nature02236.

KING, R. D. *et al*. The automation of science. **Science**, v. 324, p. 85-89, 2009. DOI: https://doi.org/10.1126/science.1165620.

KNOBEL, M. The critical role of communication in a post-truth world. **International HIgher Education**, v.100, n. Winter, p. 9-10, 2020.

KNOBEL, M.; BERNASCONI, A. Latin American Universities: Stuck in the Twentieth Century. **International Higher Education**, v. 88, p. 26-28, 2017. DOI: https://doi.org/10.6017/ihe.2017.88.9693.

KNOBEL, M.; MOHAMEDBHAI, G. Speaking Out for Science and Democracy. **International Higher Education**, n. 110, p. 5-6, 2022.

KON, A. Inovação nos serviços públicos: condições da implementação do governo eletrônico. **Planejamento e Políticas Públicas**, n. 52, p. 489-529, 2019.

KWOK, R. Flexible working: Science in the gig economy. **Nature**, v. 550, n. 7676, p. 419-421, 2017. DOI: https://doi.org/10.1038/nj7676-419a.

NAMBISAN, S.; WRIGHT, M.; FELDMAN, M. The digital transformation of innovation and entrepreneurship: Progress, challenges and key themes. **Research Policy**, v. 48, n. 8, p. 103773, 2019. DOI: https://doi.org/10.1016/j.respol.2019.03.018.

NEGRI, F. De; SQUEFF, F. DE H. S. **Sistemas setoriais de inovação e infraestrutura de pesquisa no Brasil**. Brasília: Ipea/Finep/CNPq, 2016.

NELSON, R. Bringing institutions into evolutionary growth theory**. Journal of Evolutionary Economics**, v. 18, p. 17-28, 2002.

NELSON, R. R.; SAMPAT, B. N. Making sense of institutions as a factor shaping economic performance. **Journal of Economic Behavior and Organization**, v. 44, n. 1, p. 31-54, 2001. DOI: https://doi.org/10.1016/S0167-2681(00)00152-9.

NELSON, R. The co-evolution of technology, industrial structure, and supporting institutions. **Industrial and Corporate Change**, v. 3, n. 1, p. 47-63, 1994.

OECD. **Digital platforms for facilitating access to research infrastructures**. OECD Science, Technology and Industry Policy Papers, No. 49. Paris: OECD Publishing, 2017. DOI: https://doi.org/10.1787/8288d208-en.

OECD. **Fostering Science and Innovation in the Digital Age**. OECD Going Digital Policy Note, March, 2019. DOI: https://doi.org/10.1787/9789264312012-en.

OECD. **The Digitalisation of Science, Technology and Innovation**: key developments and policies. Paris: OECD Publishing, 2020. DOI: https://doi.org/10.1787/b9e4a2c0-en.

PEREZ, C. Microelectronics, Long Waves and World Structural Change: New Perspectives for Developing Countries. **World Development**, v. 13, n. 3, p. 441–463, 1985. DOI: https://doi.org/10.1016/0305-750X(85)90140-8.

PEREZ, C. Structural change and the assimilation of new technologies in the economic and social systems. **Futures**, v. 15, n. 5, p. 357-375, 1983.

PEREZ, C. **Technological revolutions and financial capital**: the dynamics of bubbles and golden ages. Cheltenham: Edward Elgar, 2002.

PEREZ, C. Technological revolutions and techno-economic paradigms. **Cambridge Journal of Economics**, v. 34, n. 1, p. 185-202, 2009.

PEREZ, C. Unleashing a golden age after the financial collapse: Drawing lessons from history. **Environmental Innovation and Societal Transitions**, v. 6, p. 9-23, 2013. DOI: https://doi.org/10.1016/j.eist.2012.12.004.

PLANTIN, J. *et al*. Infrastructure studies meet platform studies in the age of Google and Facebook. **New Media & Society**, v. 20, n. 1, p. 293-310, 2018.

REICHENBACH, R.; EBERL, C.; LINDENMEIER, J. Online platforms for research data: A requirements and cost analysis. **Science and Public Policy**, v. 49, p. 598-608, 2022.

REINERT, E.; SAGALOVSKY, B. Carlota Perez – Her Biography and the Origins of her Ideas. *In*: DRECHSLER, W.; KATTEL, R.; REINERT, E. (eds.). **Techno-Economic Paradigms**: Essays in Honour of Carlota Perez. Anthem Press, 2009. p. 395-406.

ROSENBERG, N.; NELSON, R. American universities and technical advance in industry. **Research Policy**, v. 23, p. 323-348, 1994.

SCHOT, J.; KANGER, L. Deep transitions: Emergence, acceleration, stabilization and directionality. **Research Policy**, v. 47, n. 6, p. 1045-1059, 2018. DOI: https://doi.org/10.1016/j.respol.2018.03.009.

SCHREIECK, M.; WIESCHE, M.; KRCMAR, H. Governing nonprofit platform ecosystems–an information platform for refugees. **Information Technology for Development**, v. 23, n. 3, p. 618-643, 2017. DOI: https://doi.org/10.1080/02681102.2017.1335280.

SHAPIN, S. Discipline and bounding: The history and sociology of science as seen through the externalism-internalism debate. **History of Science**, v. 30, n. 4, p. 333-369, 1992. DOI: https://doi.org/10.1177/007327539203000401.

SMEDT, K. De; KOUREAS, D.; WITTENBURG, P. FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. **Publications**, v. 8, n. 2, p. 21, 2020. DOI: https://doi.org/10.3390/publications8020021.

SRNICEK, N. **Platform capitalism**. Cambridge, UK: Polity Press, 2017.

STRACHMAN, E. Instituições: uma caracterização crítica. **Economia**, v. 3, n. 1, p. 113-157, 2002.

STURGEON, T. Upgrading strategies for the digital economy. **Global Strategy Journal**, v. 11, n. 1, p 1-24, 2019. DOI: https://doi.org/10.1002/gsj.1364.

TEECE, D. J. Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. **Research Policy**, v. 47, n. 8, p. 1367-1387, 2018. DOI: https://doi.org/10.1016/j.respol.2017.01.015.

VAN BILJON, J.; MARAIS, M.; PLATZ, M. Digital platforms for research collaboration: using design science in developing a South African open knowledge repository. **Information Technology for Development**, v. 23, n. 3, p. 463-485, 2017. DOI: https://doi.org/10.1080/02681102.2017.1328654.

VAN DER VLEUTEN, E. Infrastructures and societal change. A view from the large technical systems field. **Technology Analysis and Strategic Management**, v. 16, n. 3, p. 395-414, 2004. DOI: https://doi.org/10.1080/0953732042000251160.

VAN DIJCK, J.; POELL, T.; WAAL, M. de. **Platform society**: public values in a connective world. New York: Oxford University Press, 2018.

VICENTE-SAEZ, R.; GUSTAFSSON, R.; VAN DEN BRANDE, L. The dawn of an open exploration era: Emergent principles and practices of open science and innovation of university research teams in a digital world. **Technological Forecasting and Social Change**, v. 156, n. March, p. 120037, 2020. DOI: https://doi.org/10.1016/j.techfore.2020.120037.

ZHU, K. *et al*. Innovation diffusion in global contexts: Determinants of post-adoption digital transformation of European companies. **European Journal of Information Systems**, v. 15, n. 6, p. 601-616, 2006. DOI: https://doi.org/10.1057/palgrave.ejis.3000650.

ZUBOFF, S. **The Age of Surveillance Capitalism**: The Fight for a Human Future at the New Frontier of Power. New York: Public Affairs, 2019.