



Innovation Commons for the Data Economy

Sara Guidi¹ 

Received: 12 September 2022 / Accepted: 3 July 2023 / Published online: 1 August 2023
© The Author(s) 2023

Abstract

Data-driven innovation entails an overall positive effect on society. Innovation is a central policy goal in the EU, and the regulation of the data economy tends to elect innovation as a primary objective. However, considerably less attention is devoted to the identification of the qualitative characteristics of the desired innovation. From a technological point of view, (data-driven) innovation can be cumulative, combinatorial, or generative. In all three instances, innovation commons are crucial. The design of successful data commons demands the analysis of the relational dimension of the data economy, which can be conducted through the framework of business ecosystems. Incentives for data-based competition or cooperation in ecosystems are inspired by a metaphorical cognition of the economic function of data: whether data is considered a resource or an infrastructure ultimately affects the design of innovation commons. To conclude, the paper draws the policy implications of this framework. Policymakers and regulators may select one narrative over another, thus molding the features of future innovation.

Keywords Innovation · Data economy · Cumulative · Combinatorial · Generative innovation

1 Introduction

Innovation is a central policy goal. In the data economy, the quest for more innovation is supported by a growing number of regulatory instruments. Widely varying in scope, such instruments do nonetheless paint a shared frame for the “innovation problem”: moving from the assumption that innovation and economic growth are deeply interlinked, the total welfare generated by the data economy increases with the rise of available data. Public policies, thus, are meant to solve the wicked problem of data

✉ Sara Guidi
sara.guidi@eui.eu

¹ Faculty of Law, European University Institute, Fiesole, Italy

access.¹ A case in point are the data-sharing provisions encompassed, for instance, in the General Data Protection Regulation (GDPR), the Open Banking Directive (PSD2), the Public Sector Information Directive (PSI), the Digital Markets Act (DMA), and the proposed Data Act. The hard truth inspiring policy-making seems to be that making more data available exponentially increases the chances of developing novel, disruptive technologies. But policies pursuing sharing-led increments of innovation in the data economy are at great risk of fallouts. Widening data access requires a certain degree of cooperation among a broad range of economic players; at the same time, the combination of data entails risks for competition (Lundqvist, 2018). Regulatory provisions crucially impact the complex net of competitive and cooperative relations that characterize the digital economy. Cooperation and competition draw the lines of innovation trajectories. Hence, policies aimed at increasing innovation in the data economy may alter the qualities of such innovation.

Indeed, not all innovation is born equal. It is an empirically grounded assumption that private incentives to innovate have positive social repercussions; it is pairwise true that technological change appoints new winners and losers.² As the term itself suggests, *in novare* means to introduce novelty: new ideas, methods, products as well as new social roles, relations, and systems. To stimulate innovation means to feed such change. Redistributive policy instruments can adjust the payoffs of innovative processes ex-post. But change can rarely be reverted. For this reason, policy-makers face a daunting challenge: not only should they ponder how to encourage innovation, but they also need to push change in the right direction. The regulation of the data economy comprehends both a quantitative and a qualitative dimension. Data-driven transformations need to be encouraged and channeled. How can policy foster the best possible innovative scenario?

This paper contributes to the debate on the regulation of data-driven innovation by framing the “innovation problem” in terms of innovation commons. The concept of innovation commons was retrieved from the institutional political studies of Ostrom (Hess & Ostrom, 2003; Ostrom, 1990, 2005) and became an established economic concept thanks to the work of Potts and Allen (Allen & Potts, 2015, 2016; Potts, 2018). In the dynamic context of the data economy, the political design of innovation commons sets the ground for future cooperation and competition. Ultimately, it determines the qualities of the resulting innovation. How should policy-makers design innovation commons? I suggest the answer to be context-dependent. Successful innovation commons channel existing incentives to cooperate and compete, and steer innovation in the legislator’s desired direction. Useful lenses for the scrutiny of digital contexts are offered by the study of business ecosystems. By

¹ The term “wicked problem” was introduced by Horst Rittel and Melvin Webber to describe the nature of many problems arising in complex societies, differing from the traditionally scientific “tame” problems. Wicked problems cannot be definitively described nor unequivocally solved, as the multitude of possible solutions escape the dualism correct/false: better or worse solutions depend on the definition of the problem provided by the parties involved, their individual preferences, and the (largely unpredictable) effects of the solution on the future development of the observed phenomenon. See (Rittel & Webber, 1973).

² The debate around the social repercussions of innovation is introduced in Section 2.

focusing alternatively on the relations among players in different positions of the system it is possible to gain a multidimensional and functional understanding of the digital context. More specifically, the ecosystem metaphor supports the identification of two different economic functions for data: data as a resource and data as an infrastructure. These two data paradigms represent implicit assumptions driving policymakers in the design of innovation commons. The qualitative dimension of innovation depends on whether data is treated as a resource or as an infrastructure.

The attentive reader has already noted that the paper makes exhaustive use of metaphors. Metaphorical thinking may not appear as the first choice framework to analyze technologically-driven changes in society. As Ricoeur elegantly observed, technical and poetic language are at two ends of a single scale (2003). While technical terminology is meant to be precise, clearly defining a univocal meaning to a specific word, analogies work by suggestion, evoking a similarity between two concepts without specifying the distinctive elements that make the comparison possible. Nonetheless, analogies and metaphors are often used to describe technological innovations. Computer desktops, artificial intelligence, and computer viruses are only some of the examples that relate the digital domain to the tangible world. By associating two seemingly unrelated concepts, the author transfers the set of implications commonly attributed to the subsidiary subject (the second term of the analogy) to the subject matter of the analysis. A desktop is organized and kept in order, intelligence allows inductive and deductive reasoning, and a virus spreads when two objects enter into contact. One single word substitutes a long list of attributes, presenting them cohesively and coherently.

Metaphors, in this way, allow us to “identify the similar in the dissimilar” (Black, 1955). Their role is not limited to embellishing an article by way of rhetoric figures: they have the power to redescribe reality, influencing how we approach a new concept. Analogies are not only a matter of language but also a matter of cognition (Vedder, 2002). They play a major role in the shaping of a narrative around the object of study. Narratives, in turn, have the power to deeply influence the behaviors of society, even more so if the unrealistic assumption of pure rationality which characterizes neoclassical economic models is abandoned (Shiller, 2019).

By influencing users’ and companies’ incentives to make their data available for re-use, narratives moving from different characterizations of data have a strong effect on innovation patterns. The direction of innovation in the digital sector is sensitive to the perception of players themselves, which often paints the frontiers of future possibilities. In the case of data, the heavy recourse to analogies highlights a common understanding: data are not simply representational resources (Gray, 2017). Metaphorical thinking draws “data imaginaries” and “data speak,” carving visions and rhetoric that try to encompass the molding effects data exercise on society. In this sense, metaphors and analogies contribute to shaping the data infrastructure. Metaphorical thinking (co)directs innovation.

Analogies also provide policymakers with mental models to rationalize the complexity of the environment. They determine the characteristics of the data economy that they deem relevant. Consequently, they drive the identification of legislative priorities. The characterization of data adopted by lawmakers influences the adoption of a protective or optimistic attitude toward data. It influences the balance that

will be struck between protective measures aimed at safeguarding citizens from new harms and novel rules promising to unlock all the potentialities of data-driven innovation. This, in turn, determines the direction taken by investments in R&D. Eventually, how the technology is conceptualized decides the future developments of the technology itself.

The following sections build up a metaphorical framework for the design of innovation commons. Section 2 expands on the effects of innovation and society, and motivates the need for innovation commons. A first taxonomy of innovation is outlined in Section 3: cumulative, combinatorial, and generative innovation represent distinct goals and require ad hoc policies. Subsequently, the concept of business ecosystem is offered as a tool to read the context in which the regulator wishes to intervene (4). Section 5 identifies two (coexisting) economic functions for data in ecosystems and outlines the differences between the two perceptions of reality. An attempt to point at possible consequences stemming from the adoption of one perspective instead of another is presented in Section 6. Section 7 concludes.

2 Data-Driven Innovation Commons

Innovation benefits society.³ The assumption is certainly simplistic: in reality, it is only true given several co-occurring conditions.⁴ Notwithstanding the debates nuancing the statement, the EU pursues innovation as a social goal. The European “innovation agenda” was first compiled in the mid-1990s (Borrás, 2003). The term innovation entered the political arena substituting and expanding the previously prominent couple “science and technology,” to whom is dedicated Title XIX of the TFEU.⁵ Since the articles contained in Title XIX replace their analogues in the TEC, it is safe to say that the pursuit of technological progress is a foundational goal of the EU.⁶ Advances in technology are expected to promote competitiveness—and competitiveness fosters growth. However, reframing scientific and technological policy in terms of innovation policy marks a shift in the understanding of the problem. A new light is shed on the socio-organizational dynamics concomitant to the

³ To dilute this claim, note that inequalities are often exacerbated by innovation processes. For instance, van den Hoven and Rooksby observe an uneven distribution of the “informational wealth” (Van Den Hoven & Rooksby, 2008). As such, a regulatory reflection on the direction of desired innovation is pivotal.

⁴ For instance, Nelson defines innovation as a form of problem-solving. In this sense, demand is an important factor in determining which problems have been tried to solve; richness influences the benefit extracted from innovation (Nelson, 1962). On a different but related note, Bender et al. have investigated the tremendous costs of language models, stemming from environmental costs to financial costs, opportunity costs, and substantial harms. The authors invite careful weighing of costs and benefits before pushing innovation in the direction of very large language models (Bender et al., 2021).

⁵ For a brief overview of the differences between science and technology, and the relation between them, see Brooks (1994). Regarding Title XIX of the TFEU, it goes under the title “Research and Technological Development and Space”; its first words explicitly reference science and technology. Art. 190 (1) indeed states that “the Union shall have the objective of strengthening its scientific and technological bases.”.

⁶ Treaty Establishing the European Community. The relevant articles are Arts. 163–176 TEC.

production of knowledge (Borrás, 2003). The pursuit of innovation is no longer an apaanage of a single policy area: it is a transversal policy goal.

The EU growing concern for innovation is supported by the literature pointing at innovation as a major source of economic growth (Gilbert, 2006). Ignoring the reasons driving firms to invest in research and development (R&D), the outcome of such activities entails a positive effect on society. This perspective is confirmed by empirical evidence showing that the social positive return from R&D investments exceeds the private one (Griliches, 1992). A pivotal study in this field was carried out by Mansfield in 1977. Measuring both the private and social returns of seventeen industrial innovations, he confirmed previous literature results in appreciating a higher median rate of social returns compared to private ones.⁷ Within his sample, he estimated a 56% rate of social return from industrial innovation. Additionally, he observed that private returns are characterized by extreme variability. Investments in innovation are risky. Lastly, in 30% of the cases, the private returns were so low that no firms able to make an accurate prediction would have invested in the innovation. Nonetheless, social returns were consistent. With the complete information about the success of the technology, it would have still made sense from a societal perspective to innovate; but firms' incentives would have been null.

More recent research on social returns of innovation flags the multiple spillovers that stem from innovating activities. Hence, it is unlikely for R&D investments alone to drive all of the productivity gains. Jones and Summers highlight that productivity gains from innovation are driven by a wider set of innovative efforts (Jones & Summers, 2020).⁸ Incorporating more variables into the analysis, they found out that the magnitude of social gains from innovation might have been consistently overestimated by models purely based on R&D investments. To estimate more accurately the average social returns from innovations, they consider the case in which gains from R&D pay off slowly, delaying the achieved benefits and thus reducing their present value. Moreover, they account for the non-R&D-related costs of innovation: among them, investments in capital assets such as equipment, machinery, and software. The resulting analysis, however, confirms that social gains from innovation remain large. The result is strengthened by the observation that other factors can lead to an underestimation of social returns of innovation when only R&D investments are taken into account: inflation bias, gains in health and longevity, and international spillovers are usually not considered. The conclusion is that, even under less simplified assumptions than the ones Mansfield relied upon, investments in innovation promote large average social gains.

⁷ Note that Mansfield defines the social rate of return as the sum of the savings incurred by customers due to the product's cost reduction (a consequence of both process or product innovation) and the innovator's (adjusted) profits. See Mansfield et al. (1977), p. 224.

⁸ Observe that Jones & Summers' model is rooted in endogenous growth theory. The authors estimate social returns from innovative investments based on the growth rate of GDP per capita, under the assumption that, first, the latter will be equivalent to the growth rate in total factor productivity (Solow, 1956), and, second, that total factor productivity in advanced economies comes from investments in new ideas (e.g., Romer, 1990; Aghion & Howitt, 1992). See Jones and Summers (2020), p.4.

Incentives to innovate, however, depend on the expected returns for the single firm investing in innovation. Private expected returns might be lower than social benefits—in this case, missing investments harm society more than firms. Such risk of underinvestment is coherent with a characterization of innovation as a commons. The term commons can indeed be used with two meanings: first, it may refer to a resource that is both rival and non-excludable; second, it can indicate the institutions that “govern the appropriation and provisioning of the resources among the community” (Potts, 2018).⁹ Innovation commons, thus, jointly define the resources that, if shared with the community, would enable innovation, and the conditions under which such resources are shared. In this paper, data-driven innovation is assumed to be a common. This entails that (1) more data-driven innovation is desirable, and (2) the innovation problem can be construed as a problem of combined knowledge (Potts, 2018).

Data-driven innovation undoubtedly yields transformative effects on society. The term data-driven innovation defines the use of data and analytics to improve or foster new products, processes, organizational methods, and markets (OECD, 2015).¹⁰ Machine learning techniques, artificial intelligence, and always-online interconnected objects are just some of the technological innovations made possible by data exploitation. The data economy continues to increase in size (MIT Technology Review Insight & Infosys Cobalt, 2021). Undisputedly, economic growth does not automatically translate into an even increase in well-being across all the strata of society. The altered social landscape brought upon by the recent rapid surge in the use of pervasive technologies is fraught with risks; the wealth generated by this flourishing sector of the economy may be unevenly distributed. Nonetheless, technological progress in the collection, storage, and analysis of digital information allows for the creation of new value (Buchholtz et al., 2014; OECD, 2015, 2017). It permits to augment the size of the pie to be shared between the participants in the economy. This adds to the direct positive effect of innovations in sectors such as health, science, and education, and to the benefits consumers receive in terms of increased variety (OECD, 2015).

Data drives innovation, but the mere existence of data does not necessarily bring about more innovation. The effective exploitation of the value creation potential of data demands, to begin with, appropriate technologies for their analysis and adequate knowledge management systems to ensure that the information they carry is not lost (Bresciani et al., 2021). Data science, machine learning, artificial intelligence (AI), and computing technologies empower data-driven innovation (Luo, 2023). Changes in how information is stored and made available for use, and in the technology adopted to exploit it, determine the progress of innovation. Ultimately, data flows draw the frontiers of the data economy. Data generation makes innovation

⁹ The theory of the innovation commons draws upon Hayek, Williamson, and Ostrom to present the innovation problem as a combined knowledge problem (Potts, 2018).

¹⁰ The concept of data-driven innovation, in this and subsequent publication of the OECD, is tied to the availability of “Big Data,” that is “the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value” (De Mauro et al., 2016).

possible; rules for data access and use define the direction of the innovation trajectories. The implications across society stemming from the distribution and use of data are wide-ranging (Sadowski, 2019). They dictate which players participate in the economy, the distribution of power among them, and which actors are simply left out.¹¹ In other words, it marks the characteristics of data-driven innovation.

Hence, the design of innovation commons is a necessary but daring task. New spaces for the combination of data-embedded information are key to facilitating the transmission of knowledge and fostering innovation. At the same time, the architecture of said spaces has long-lasting effects on the society of the future. Posing that the sharing of informational resources will produce innovation, and thus economic growth, how does the resulting society look like? The next section provides an overview of the three main modes in which data drives innovation, clearing the field for a successive evaluation of data-driven innovation's socio-relational dimension.

3 Which Innovation for the Common Good?¹²

Among the many definitions that have been offered for the term innovation, one is particularly convenient to investigate innovation commons: innovation is “a new pattern of bits of information” (Macdonald, 1998). In the digital sector, such bits take a binary form. Data is, indeed, machine-readable encoded information (Zech, 2016).¹³ New patterns of data enable data-driven innovation; they—to phrase it better—are data-driven innovations. Variations in such patterns may occur based on the rules governing their creation. Which instructions can be followed in the quest for innovation? The strategy adopted to explore the potentially infinite space of new information (i.e., the *search space*) substantiates diverging results. Players' limited and value-infused perception of the search space infuses the choice of one strategy over another.

Of course, other factors count in the selection of any innovation strategy. A prominent place among them is reserved for the availability of data (and information), the costs and benefits of sourcing additional one, and more generally the fit with players' overarching goals and declared mission. The analysis of the socio-relational

¹¹ In this regard, note that a growing body of literature, referred to as critical data studies, is devoted to the exploration of the cultural, ethical, and critical challenges posed by Big Data. See, for instance, Iliadis & Russo (2016), Crawford (2014), and Dalton et al. (2016).

¹² The title of this section hints at the book “Economics for the common good” (Tirole & Rendall, 2017). In this paper, innovation is conceptualized in a similar way to what Tirole and Rendall did with economics: a fundamentally positive force, whose transformational potential comes with challenges that cannot be ignored. The authors define the “common good” in terms of general interest, which may be opposed to the interests of individuals. To identify the common good, one needs to place themselves behind a veil of ignorance and, pretending not to know their position in society, point out what is desirable. The notion builds on an intellectual tradition originating with Hobbes and Locke, continued by Rousseau, and more recently refined by Rawls and Harsanyi (Rawls, 1999, p. 2).

¹³ For a more technical definition of data, consider the one provided by UNECE, which describes data as a “reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing” (UNECE, 2020). This definition is the one adopted by ISO and OECD.

dimension of innovation commons offered in Section 4 is intended to facilitate a deeper comprehension of these factors. Before plunging into the depth of the network dynamics of digital complex systems, however, it is fundamental to offer a taxonomy of the possible data patterns constituting the output of the economic players—and the subject of policy-making. Bits of information can prompt cumulative, combinatorial, or generative innovation (3.1). Data, specifically, can drive all three kinds of patterns (3.2): luckily, it is possible to identify a baseline regulatory goal capable of fostering innovation in all these forms (3.3).

3.1 Cumulative, Combinatorial, and Generative Innovation

Innovation takes multiple shapes, and is thus described in multiple ways. Setting aside its socio-relational dimension for the time being, the focus is here kept on technological innovation. A widely adopted definition of technological innovation is that of “a new or improved product or process whose technological characteristics are significantly different from before” (OECD, 1992). Focusing on the word “improved,” one can get the idea that innovation is a cumulative process: incremental changes follow one another, to the point that the resulting product/process is so different from the original one to be defined as new. Indeed, the accumulation of knowledge is a possible driver of innovation. The basic principle behind the functioning of carts and cars is the same and has been known to humanity since the dawn of time: a round object in movement dissipates less energy than an angled one. Over thousands of years, new knowledge was accumulated and applied to the cart. The development of the combustion engine determines the invention of cars; nonetheless, the functioning of cars capitalizes on all the knowledge cumulated by human experience with carts. The collection of information enables *cumulative* innovation. The more information is retained in a system, the higher the potential for growth (Winters, 2020).

Innovation does not exclusively rely on the accumulation of information. A second kind of innovation was theorized by Schumpeter.¹⁴ In *The Theory of Economic Development*, he introduces the concept of *combinatorial* innovation. Economic change, in his view, stems from new combinations of productive means. Production is the combination of available materials and forces. Consequently, “to produce other things, or the same things by a different method, means to combine these materials and forces differently” (Schumpeter, 1926). Producers of vehicles do not reinvent the wheel every time a new means of transportation hits the market: the wheel is simply assembled in combination with different components. Bicycles, cars, motorcycles, trucks, and electric scooters are all the result of the combination of wheels with something else. Note that combinatorial innovation is not possible if the information is not easily exchangeable. Thus, information needs to be organized. Combinatorial innovation involves the combination of different components as much as the combination of the clusters of knowledge associated with them.

¹⁴ The concept was probably already introduced by Jean-Baptiste Say, credited by Schumpeter himself in the second German edition of *The Theory of Economic Development* (Harper, 2020).

Moreover, combinatorial innovation is robust to high levels of information loss from one generation of products to another (Winters, 2020). Simplification of information may even help in reducing the complexity of the search space, that is, the number of solutions available to a certain problem. In other words, less information generates simpler, more reachable questions. The combination of multiple solutions to simple questions is what permits breakthrough, complex, innovation.

Consider an oversimplified history of the wheel and its applications. The first known application of the wheel dates back to the Copper Age: domesticated horses were trained to move wheeled carts. Cumulative improvements over millennia led to better carts. Plenty of other inventions, however, concurred to ease the lives of humanity. Among them, animal-led mills, constituted by two overlapping stones, assured the production of high quantities of flour. The integration of such structures with wheels rotating with the force of water cascades allowed the construction of water-powered mills. The functioning of the latter is based on the property of rounded objects of conserving more energy than other shapes, the very same property that makes carts with rounded wheels an efficient means of transportation. The wide leap in innovation constituted by the water-powered mill was enabled by the application of a known technology in a different context. Similarly, cumulative innovation gave the wheel in water mills the shape of a turbine. Water mills became increasingly efficient. Yet, this increase in efficiency loses relevance when compared to the leap in technology that followed the application of the turbine to steamboats. Brian Arthur theorizes such changes of context as shifting of the domain. Domains are defined as “any cluster of components drawn from in order to form devices or methods, along with its collection of practices and knowledge, its rules of combination, and its associated way of thinking” (Arthur, 2014, p. 22). Technologies provide solutions to specific problems, they define products or processes; domains are a collection of mutually supportive technologies. Each domain possesses a unique set of accumulated knowledge, practices, and mindset. They own their proper grammar (Arthur, 2014, p. 47).

Changes in the domain are the main way in which technology progresses. Sometimes, such changes originate from the illuminate minds of gifted individuals. Other times, it is chance that provides unique opportunities. More often, however, it is the spill out of information among different domains that enables the change. When unprompted change is driven by large, varied, and uncoordinated participants, innovation is said to be *generative* (Zittrain, 2005). Generative innovation is easier when the function of technological components is not well-defined (Arthur, 2014). The reason is simple: a kid playing with a red brick will come up with infinite roles for it; should the brick be given the shape of a car, the possibilities become limited. Similarly, multi-purpose technologies foster creative recombination (Murmann & Frenken, 2006).

Multi-purpose technologies are, indeed, technologies that possess a generative capacity. The concept of generative capacity was debuted in the context of social policy research by Schön and referred to the ability of metaphors to continuously generate new perspectives on the world by carrying familiar meanings in new

domains (Schön, 1993).¹⁵ Imported in linguistics, it is traditionally associated with the ability of alphabets to constantly generate new meanings through the recombination of sounds (Chomsky et al., 2006). The meaning of the term “generative” was retained when translated into innovation studies: here, generative capability came to define an overarching capability that enables continuous innovation (Guo et al., 2022). Generative innovation is unending and self-sustaining.

In some instances, the availability of information constrains the trajectory of innovation: the ebbs and flows of human technological progress are affected by the physical movement of individuals carrying expertise, ideas, and knowledge. In cyberspace, physical limits can be relaxed. It remains to be ascertained whether other limitations are in place in the digital world. Does the nature of data itself nudge to the promotion of cumulative, combinatorial, or generative innovation specifically? The question is bound to draw in respondents’ perceptions about the essence of data. Pending further assessments, however, the following Section 3.2 attempts to offer a first irrefragable answer on the potentialities of data for innovation.

3.2 Data Drives Just Anything

The effective management of innovation commons is contingent on the identification of the target type of innovation. Indeed, slightly divergent institutions encourage respectively cumulative, combinatorial, and generative innovation (West, 2009). An appropriate starting point for any policy evaluation stands in the recognition that data potentially enable all three of the above-mentioned innovation paradigms.

Data-driven innovation can be cumulative. At a basic level, the availability of large quantities of data continuously enables the discovery of novel insights. This appears to be the assumption behind the release by the Novartis Institute for Biomedical Research, in 2007, of an incredibly wide amount of data retrieved from the analysis of the genome of more than 3000 type 2 patients (West, 2009). Besides, the technologies adopted in data analyses improve cumulatively as well. The accurate targeting of advertisement services offered by Google or Meta, for instance, builds over decades of systematic data collection. The algorithm adopted by Netflix to provide personalized movie recommendations constantly improved over time as the company gained access to a wider audience. A quantitative study conducted in 2021 on innovation in AI used to mitigate and adapt to climate change showed that new AI patents in mitigation and adaptation technologies are associated with an exponential number of subsequent innovations (Verendel, 2023).

Combinatorial innovation can be data-driven, too. It is common practice among medical scientists, for instance, to mine literature and open data to facilitate diagnostic decision-making in cancer treatment (Ding & Stirling, 2016). Data-driven technologies are often combined: again in the medical sector, blockchain technology can be combined with machine learning to protect personal, highly sensible data collected by

¹⁵ Schön notes that “problem settings are mediated [...] by the “stories” people tell about troublesome situations” (Schön, 1993, p. 138). Metaphors, in his view, are an instrument for policy evolution that allows the framing of ever-new social situations. His reflections appear more relevant than ever in respect of the regulation of the data economy.

medical devices (Snow, 2021). The resort to combinatorial strategies to explore the space of new possibilities bears the considerable advantage of reducing the uncertainty intrinsic to the innovation process. By combining known components, inventors can sensibly reduce the variation in the expected success of their efforts (Fleming, 2001). Completely new components can lead to spectacular failures or triumphant breakthroughs; a combination of old components brings upon more modest but less uncertain results. The concept is exemplified plainly by innovative digital products resulting from the application of data-driven technologies to previously analogical domains (Hylving & Schultze, 2013). The knowledge accumulated in the domain of app development spills onto wearables as much as smart TVs, autonomous vehicles, or IoT appliances.¹⁶ Full digital combinatorial innovation is possible too. In this sense, Application Programming Interfaces (APIs) are a keystone. Thanks to APIs and agile development methods, multiple services can quickly be integrated into a single-user application (Yildiz, 2022). Interoperability facilitates combinatorial innovation. The fungibility of data investments promotes inter-sectoral jumps. As a result, few big players controlling common APIs can expand in multiple sectors: the markets of competition blur, and new risks materialize (Sharon, 2021).¹⁷

Last, the characteristics of the data economy facilitate generative innovation. Data-enabled technologies such as data analytics, data mining, Artificial Intelligence (AI), and the Internet of Things (IoT) can be easily transferred from one domain to another. Their decision problem is mostly defined in wide terms and can quickly adapt to the context. Machine learning, for instance, is used to tailor Netflix's recommendation as well as to identify unknown influences among historical painters—and for uncountable other applications. Coherently, the data economy is a dynamic environment. Big data have a generative capacity (Scholz, 2017, p. 70). Generative algorithms are eyed as the actual more promising development in the data economy (World Economic Forum, 2023; Minevich, 2023). Generative innovation involves a potentially infinite amount of economic players, although strong variations can occur in their degree of awareness and the share of value they capture.¹⁸ Petabytes of data provide information that answers unposed questions (Anderson, 2008). Whether or not this signifies “the end of theory,” as algorithms generate more insightful, useful, accurate, or true results than specialists crafting targeted hypotheses and strategies (Graham, 2012), it is undeniable

¹⁶ At this point it is necessary to clarify the relation between data, information, and knowledge. Data are the syntactic format through which semantic content, *id est* information, is carried. With an example drawn from the analogic world, data represent the letters printed in a book that physically carry the information. The latter, on the other hand, define the meaning that the combination of printed characters delivers. Note that one single character carries, by itself, little information. Additionally, information theory illustrates how the identification of regularities in a set of data is used to predict patterns and guide behaviors, factually turning information into knowledge. In a similar guise, companies in the data economy recur to data analytics to turn data into an intangible asset (knowledge).

¹⁷ To get an idea of the magnitude of those “sphere transgressions,” see the Sphere Transgression Watch at <https://www.ihub.ru.nl/project/spheretransgressionwatch.page>.

¹⁸ One can argue that anyone contributing to one of the webpages composing the 45/B compressed plaintext training dataset of GPT-3 participated in the generative algorithm (Brown et al., 2020). However, each author's contribution was involuntary and no compensation is foreseen for them.

that innovation is increasingly the result of inductive, rather than deductive, reasoning (Mazzocchi, 2015).¹⁹

Consider, for instance, the evolution of watches. The knowledge accumulated in the domain of application development integrates the know-how of producers of watches and their physical components. In 1972, Hamilton released the first digital watch under the name Pulsar Time Computer. The product was a success, contributing to shaping social imaginaries and expectations about the future (Kent, 2021). It represents, mostly, a leading example of combinatorial innovation. The subsequent and frequent releases of new versions, updated only in the graphical interface, can undoubtedly be classified as cumulative improvements. Generative innovation only happens when wearables are integrated with software and operating systems. Technologies matured in the context of smartphone development, applied to the hardware of a digital watch, gave life to an entirely new product with use not comparable with watches'. Wearables permit reading the time; a large share of fitness fiends among consumers suggest that measuring exercise and monitoring sleeping time are more appealing than simply checking the time (The Economist, 2015). Mostly, the wide set of applications available to integrate the smartwatch offers novel and ever-changing uses.

In conclusion, data are little red bricks.²⁰ They can be piled up one over another to improve existing constructions; combined with a set of wheels they will turn into cars; they can generate several new exciting games whose limit only lies in the fantasy of the kid playing with them. Posit, however, that multiple children decide to join and make their own bricks available to create a more intriguing construction. They will surely need rules. In designing those rules, what kind of construction should be selected as a goal? More overtly, if data can drive cumulative, combinatorial, and generative innovation alike, which one should be the objective of policies for innovation commons? In this paper, I prudently approach this interrogative: before advancing with the analysis, the next Section 3.3 identifies a safe baseline for political and regulatory action.

3.3 Data Access, a Common Objective

Policymakers engaged in the regulation of the data economy face a fundamental question: would the pursuit of one kind of innovation hinder the evolution of another? Would initiatives supporting cumulative innovation, for instance, affect the evolution of combinatorial or generative innovation? In providing an answer to such a quest, the bottom line is that any institutional response to data-driven

¹⁹ The use of extremely high amounts of data subject to relatively low levels of project-specific reorganizations and sorting bears the risk of dragging on – and possibly reinforcing – existing biases in society. See Pigliucci (2009).

²⁰ This assumption shall be read as one of the many metaphors used in discussing data. The underlying implication is not that the economic function of data is assimilated to the function of bricks; more simply, the accent is posed on the influence of the relational and regulatory environment on the results stemming from the shared use of data. For a more extensive discussion of the economic function of data, see Section 5.

innovation commons should have an overall positive effect on society as a whole. In other terms, a policy should not provoke more damage than benefit. The uncertainty inherent in the regulation of new technologies risks being harmful in the long run (Anderlini et al., 2013). Analytical tools are needed to reduce uncertainty.

At a basic level, there is one policy goal encouraging all three kinds of innovation. Fostering data access and sharing is a fundamental enabler of data-driven innovation. Innovation builds on existent information. While single data points, taken alone, carry little information, data sharing is undermined by the risk of communicating information.²¹ Sharing can be hindered as a consequence of what Kenneth Arrow described as the “information paradox”: a potential buyer of information cannot assess the value of the transaction before they receive the information itself, but if the seller were to reveal the content of the information to the buyer before concluding the contract, there would be no incentives for the buyer to proceed with the transaction, as he would already possess the information (Arrow, 1962, p. 19). Sometimes, however, information is spontaneously shared by the players in the economy. That is possible if they share a common goal. Allen and Potts studied information commons in the early process of collective pooling of information (Allen & Potts, 2015; Potts, 2018). When uncertainty is higher, and the possible innovation trajectories are almost boundless, information is extremely valuable.²² According to their framework, as soon as innovation becomes established the need for cooperation is bound to decrease. Uncertainty over the innovation trajectory lessens. Competition begins to operate. Incentives to solve the innovation paradox shrink, eventually leading to a reduction in the amount of information exchanged.

Competitive dynamics carve innovation into one of the three above-mentioned shapes. As such, they also affect cooperation. Incentives for data sharing are linked to the competitive and cooperative relations existent in the digital economy: a possible framework for the analysis of such relations is presented in the next Section 4. By introducing the concept of the business ecosystem, the ambition of this paper is to offer the reader a pair of glasses to more clearly distinguish the need for institutional intervention in this extremely dynamic sector of the economy. Mostly, these lenses permit us to discern which innovation do data enable, and when. Commons could thus be managed to favor, respectively, cumulative, combinatorial, or generative innovation. A context-dependent intervention, I argue, necessitates tools to understand the context. The concept of ecosystem is instrumental to root up the relational dynamics of the data economy.

4 Ecosystems, Loci of Innovation

The development of business ecosystems is the organizational backbone of digital-based innovation. Ecosystems enable what Benkler defines as commons-based peer production (Benkler, 2002), a third mode of production alternative to markets and

²¹ In this sense, talking about data access (by firms that might not necessarily derive from the data the same information as the data holder) necessarily involves a discussion around the cession of information (by economic players worried that data sharing might pass sensitive information to competitors).

²² Regarding the relation between uncertainty and technological innovation, also see Dequech (2004).

firms especially frequent in the digital economy.²³ Peer production can be described as “a process by which many individuals, whose actions are coordinated neither by managers nor by price signals in the market, contribute to a joint effort that effectively produces a unit of information or culture” (Benkler, 2003, p. 1256). Commons-based peer production, in the context of the digital economy, depends on the aggregation of independent firms which autonomously “scour their information environment in search of opportunities to be creative in small or large increments” (Benkler, 2002, p. 376). Such exploration is based on data sharing: by accessing new data firms retrieve new information, reduce their level of uncertainty, and undertake innovation ventures. Note that aggregation implies a certain degree of cooperation. Cooperative relations are nurtured to better respond to continual competitive threats. In the digital sector, innovation is the engine of competition (OECD, 2022).²⁴ Cooperation, in this challenging arena, becomes a competitive instrument (Teece, 1992). Cooperative data-driven innovation is faster and more apt to respond to ever-evolving threats (Petit & Teece, 2020). The complex net of cooperative relations that rises around major players’ technologies constitutes the structure of business ecosystems.

The picture of technology-intensive machines voraciously analyzing exceptionally wide datasets recalls sci-fi imaginaries rather than organic biological ecosystems. Still, the term ecosystem is part of the academic jargon of multiple fields interested in the data economy: from business, management, and innovation studies to computer sciences, it is rapidly spreading to the legal and economic literature. In particular, competition law scholars are advancing the idea that ecosystems allow to grasp and systematize the multi-dimensional nature of competition in the digital sector (Jacobides & Lianos, 2021; Petit & Teece, 2020; Robertson, 2021). More than that, the term made it to legal texts (Digital Markets Act, Digital Services Act) and Court decisions (Google LLC and Alphabet, Inc v European Commission, 2022). In this paper, business ecosystems are defined as comprising firms that collectively offer value to customers, independently setting their business strategy but strongly connected one with another. Independence and interdependence are both necessary but insufficient conditions. Economic players independently designing their products are part of markets²⁵; full interdependence between outputs is achieved in

²³ Benkler uses the concept of commons-based peer production to analyze “the phenomenon of large- and medium-scale collaborations among individuals that are organized without markets or managerial hierarchies” which “is emerging everywhere in the information and cultural production system” (Benkler, 2002, p. 375). While the subject of his research is limited to decentralized modes of information production (such as, e.g., free software), his framework can easily be adopted for the study of information production structures in business ecosystems at large, whether or not the latter are governed by a leader. For this purpose, however, the behavior that Benkler attributes to individuals is assigned to firms.

²⁴ The idea that innovation constitutes a relevant dimension of competition is not new: already in 1962 the economist Richard Nelson observed that “increasingly the focus is on competition through new products rather than on direct price competition” (Nelson, 1962, p. 4). In the digital sector, the phenomenon is simply accentuated.

²⁵ Here, and in the remainder of this paper, I refer to “products” to indicate the output of players in digital ecosystems, implying the much longer notation “products/services/solutions.”.

hierarchical organizations such as firms or conglomerates. Ecosystems, like commons-based peer production systems, stand in between “hierarchies” and “markets” (Benkler, 2003; Gawer, 2014; Jacobides et al., 2018).

Although it is possible to identify examples of business ecosystems that originated as early as the 1920s, the managed business ecosystem as an organizational form is connected to the computer industry of the 1960s. James Moore identifies two major shifts that played a pivotal role in its affirmation (Moore, 2006). The first was the development of the family of computers IBM System/360.²⁶ Thanks to a modularized architecture, IBM was able to offer several variations of the same product able to accommodate the needs of different market segments, without the need to develop and maintain multiple product lines (Liu, 2016). The modularized architecture allowed for the development of complementary markets for specific parts of the computer (Moore, 2006). Modules made it possible to launch an extremely complex product on the market, at the same time assuring that it could easily evolve to accommodate changes in demand. The second paradigmatic shift identified by Moore as generative of the business ecosystem as a managed form of business organization was operated in the same period by HP. While IBM established a new technical paradigm, based on modularized interoperable architecture, HP laid the foundation for a new cultural paradigm. The company’s internal organization was grounded on collaboration. Small groups of engineers would cooperate on specific projects and flexibly re-arrange themselves at their conclusion. The organization was based on open and loose groups, among which information flew relatively freely (Burgelman et al., 2017). The collaboration of autonomous individuals promotes creativity (Benkler, 2003).

Modularized architecture and collaborative culture are, indeed, distinct features of digital ecosystems. Chiefly, digital ecosystems permit the extension of those paradigms beyond the borders of the firm. Complex innovations are developed as combinations of distinct modules. Specific product design choices help firms to generate product families and lead to systematic, quick innovation through the use of common assets (Gawer, 2014). Innovation is boosted by specialization. In each module, knowledge is accumulated. As such, new firms can access the market by providing new complementary solutions that can be integrated into the main product. At the same time, module recombination allows for multiple possibilities for combinatorial innovation. It is by recombining the components that Apple can accommodate the needs of all the market segments covered by its personal computers offer. The collaborative culture that made HP successful in a highly technological sector is frequently adopted by firms belonging to the same ecosystem. Unless they are fully complementary, companies in the ecosystem alternate competitive and cooperative relations based on time, market, and functions.²⁷ They are said to “coopete” (Brandenburger & Nalebuff, 1998).

²⁶ On this point, Moore builds on the work of Baldwin and Clark (2000).

²⁷ Functions are marketing, sourcing, operations, research, and development.

Autonomous and asynchronous innovation can be conducted by multiple firms using the same resource. At the same time, access to resources may provide a competitive advantage. Data pooling and sharing are influenced by inter- and intra-ecosystem cooperative dynamics. On the bones of business ecosystems, data ecosystems take shape. An interesting strand of the literature on the regulation of data is devoted to the identification of technical solutions capable of fostering open data ecosystems or explicitly facilitating data reuse by supporting collaborative networks.²⁸ The object of these studies is constituted by spontaneous or privately managed data-sharing platforms. It is, however, important to note that data and business ecosystems are not overlapping. Data ecosystems enable data commons; business ecosystems to innovation commons. By focusing on the latter, the article intends to flag the potential rather than the actual data commons: the ones that could happen, if regulation is successful. Moreover, the focus is kept exclusively on the data commons that foster innovation. Indeed, the structure of business ecosystems determines the architecture of innovation.²⁹ Regulation on data access affects the information available to ecosystem members. Thus, they influence the trajectory of innovation. How it happens are hard to predict due to the complexity of ecosystems' structures. The next part (4.1) attempts to analyze the distribution of information in digital ecosystems.

4.1 Information-Bound Innovation Trajectories

Digital ecosystems are complex systems (Briscoe, 2010). The number of publications dedicated to the understanding of the digital economy might make this assertion sound trivial. But puzzled researchers of the digital era might feel reassured recalling the definition of complexity provided by complexity science: the complexity of a system is related to the amount of information necessary to describe it (New England Complex Systems Institute, n.d.). The description of digital ecosystems requires more information than the description of traditional markets. The concept can be better understood through an example. Consider a hypothetical “grandparent test”³⁰ on the environment of operations of a hotel and the environment of operations of its closer digital equivalent, Airbnb. A traditional hotel purchases toiletry sets, breakfast products, and cleaning services from its suppliers, and provides a room for the night to its clients. The operations of Airbnb involve a significantly wider number of players (Fig. 1). Additionally, note that the relationship among them is more varied compared to the traditional market. Contracts and purchases are no longer the main way of interaction. Players that are not in direct communication can nonetheless be highly interdependent (Shaughnessy, 2019). The amount of information needed to describe Airbnb activities is significant.

²⁸ See, for instance, Immonen et al. (2014) and Oliveira et al. (2019).

²⁹ Lessig, 2002.

³⁰ A “grandparent test” consists in explaining a technical topic without recurring to specialized jargon, in a way that would make it understandable to someone not familiar with the technicalities (Winsor, 2019).

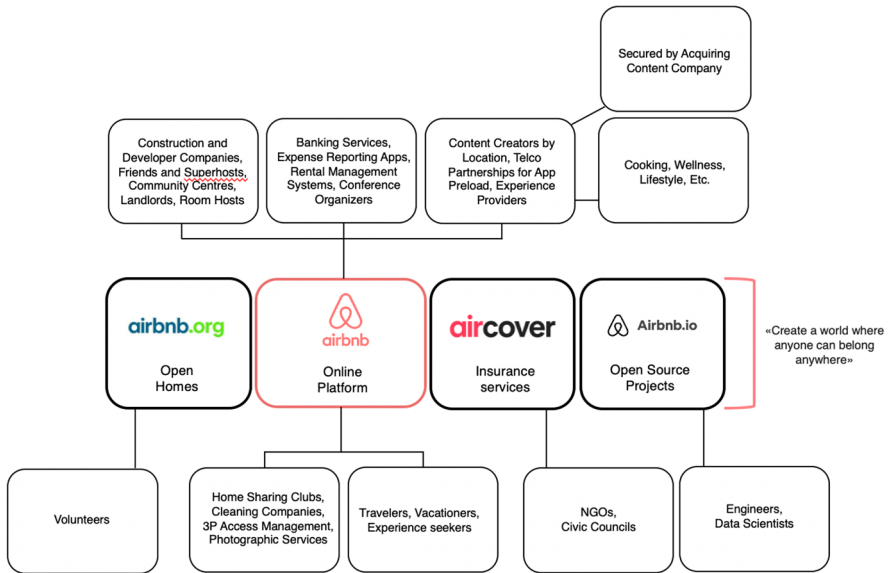


Fig. 1 The Airbnb Ecosystem. Source: own elaboration based on Shaughnessy, 2019

The main characteristic of complex systems is that small changes in one of the parameters can produce large changes in the aggregated behavior of the system (Petit & Schrepel, 2023). This can be easily understood by referring to a peculiar example of complex systems: the atmosphere. In 1962, the meteorologist Edward Lorenz observed that a butterfly’s flap in Brazil could cause the formation of a tornado in Texas.³¹ In the context of business ecosystems, the butterfly effect explains the unpredictability of innovation trajectories. The course of technology is influenced by what Arthur referred to as “small historical events” (Arthur, 1983).³² Apparently, random choices in the early stages of development of a new technology become cemented in the technological structure of the economy. Arthur states that “micro-events become magnified by positive feedbacks; their cumulation decides the outcome and forms the causality” (Arthur, 1983). The anticlockwise hands in the 1433 clock displayed in Florence’s cathedral testify that casualty ultimately is entrenched in conventions.³³ “History becomes destiny” (Arthur, 1983, p. 16).

³¹ To be precise, in 1962 Lorenz observed that a flap of a seagull’s wing could change the course of weather forever. In 1972 he organized a conference titled “Could a butterfly’s flap in Brazil cause a tornado in Texas?” The meteorological phenomenon he discovered was named the “butterfly effect,” however, I take the official date of discovery of the effect as 1962, as the animal flapping wings is the only modification to the scientific finding that occurred in 1972.

³² Note that Arthur’s studies presupposes increasing returns to technology, that is decreasing cost of supply. This typically holds in the digital economy.

³³ The “anticlockwise convention” disappeared around 1550 (Arthur, 1983, p. 15).

Strong path dependencies undermine the applicability of the neoclassical “rational agent” assumption.³⁴ Participants in the system are extremely bounded agents: their decisions are heavily dependent on their starting point, and the effects of such decisions are determined by a net of interdependencies of which they are likely not fully aware. Ecosystem complexity is made manageable through modularity (Baldwin, 2007; Moore, 2006). An organizational structure is modular when it is composed of elements, i.e., modules, that independently perform distinctive functions (JK Gershenson et al., 2003; Simon, 1962). It tends to emerge in large systems characterized by a high number of interdependencies (Simon, 1962; DL Parnas, 1972; Ethiraj & Levinthal, 2004). The separation into modules allows for the creation of sub-systems. Participants in each sub-system are closely connected to one another and loosely related to participants in other sub-systems. Within modules, the unit of analysis is limited to a reduced amount of interactions: hence, complexity is reduced. Conventions and standards dominate exchanges among the different modules.

Modularity is the key to agile innovation. If technological paradigms continuously shift, how can firms build resources and capabilities that sustain competitive advantage? Modularity allows parallel work to proceed independently. Chiefly, multiple components of a complex product can be innovated at the same time. The case of IBM/360 illustrates the advantages of a modular product design for innovation. First, autonomous innovation takes place within components. New, updated data entry units could be released anytime, as long as they respected the measures of the console. Research and development are independently carried on by each component. More generally, players producing alternative modules compete with each other. In dynamic environments, competition is based on innovation (OECD, 2022). A modular structure, by simplifying complexity, reduces the information necessary to adduce incremental improvements. Moreover, reverse engineering and imitation are made easier. Cumulative knowledge and joint problem-solving within modules are incentivized (Pil & Cohen, 2006). Thus, we could expect higher rates of incremental innovation.

Second, combinatory innovation takes place by exchanging and replacing lower-performance modules with higher-performance modules. In the IBM/360, exploitation of combinatory innovation gave rise to multiple models, each of whom was better suited for different typologies of users. The relative ease with whom it was possible to recombine the essential elements of the computer gave birth to a flourishing market of non-original substitutes. Peripheral products could be attached to the System/360 processor thanks to its standard interface. Third-party suppliers and manufacturers quickly entered the new market. Consequently, competition increased. A modular architecture facilitates the development of new markets, encouraging value creation.³⁵

³⁴ Neoclassical economists were aware that participants of the economy do not always behave in rational ways. Nonetheless, the adoption of a presumption of full rationality was justified by the so-called “as-if” justification: as players would be rational the majority of the time, it is safe to build theories “as-if” they act rationally (Friedman, 1953).

³⁵ On this point, see Coase (1960).

Lastly, the modularized architecture of digital ecosystems enables generative innovation. The early production of the IBM/360 can be considered a case of a closed ecosystem. When ecosystems are open, that is to say when the membership can be acquired by any third party capable of annexing their product to the system's complex offering, the boundaries become more malleable (Um et al., 2013). Innovation can advance by harnessing the distributed creativity of heterogeneous players (Yoo et al., 2012). The generativity of open ecosystems "comes from the variety of plug-ins of different kinds," whereas closed ecosystems can only rely on a variety of modules of the same kind (Um et al., 2013; Yoo et al., 2012). The generative capability of the Android ecosystem, for instance, is given by the virtually never-ending diversity of the third-party applications that can run on it. An Android smartphone can turn into a training device, a music player, or even a metal detector. The openness of the system benefits from product-agnostic modules, such as Google Maps APIs, which can be integrated into a multitude of different products. Open systems count more members; more members translate into increased complexity. For this reason, open ecosystems often proliferate around the figure of a leader: hierarchy is a powerful way to manage complexity (Simon, 1962).³⁶ Undoubtedly, leaders detain a substantial advantage over the other members: they can increase variations in their products and raise the overall flexibility of the system without incurring high transaction costs (Um et al., 2013). By designing the architecture, they can easily steer innovation trajectories. Their control over innovation is, however, not absolute. Generative innovation relies on exchanges and unforeseeable contacts, and increases with the flow of information within and outside the ecosystem.

The relative position of economic players, the openness of connection between modules, and the overall architecture of the system affect the characteristics of innovation in the data economy. The level of competition ecosystem's members are most sensible to, together with the perceived rivalry of data, determines the incentives to share data. Mostly, remember that the economic players aggregated in an ecosystem are independent. Ultimately, their competitive strategy determines what kind of innovation they want to pursue. As their competitive strategy depends on which information they have, and considering that data embed such information, access to data determines the resulting innovation. For this reason, the next Section 5 concentrates on the possible roles that data can assume in ecosystems. Different (perceived) data functions determine the willingness to access data commons and the strategic choice of the participants in the economy. The resulting innovation will be qualitatively different. Finally, when policies for innovation commons endorse a specific data function, they implicitly support specific qualitative features of innovation.

³⁶ This this is, of course, not always the case. Numerous Open Data movements originate spontaneously, and it is not rare for members of data commons to engage in participative forms of administration. However, Big Tech's ecosystems are the ones that mostly catalyze the interest of regulators, preoccupied with the excessive powers retained by those leaders.

5 The Dual Role of Data

The context in which innovation commons take shape comprises both the external forces fostering or hindering collaboration and an individual calculation of the costs and benefits of sharing. Incentives to cooperate, thus engaging in data commons, and compete, i.e., excluding opponents from the commons, depend on the perceived value of data. The latter is, in turn, linked to players' perceived economic function of data. The modular architecture of digital ecosystems permits the identification of two different economic functions for data. Each function facilitates a specific kind of innovation. The next paragraphs focus on the perceptions that data points are ecosystem resources shared according to players' incentives to cooperate or compete.

Section 5.1 examines two analogies that describe different declinations given by the literature to data as resources. In particular, Sub-section 5.1.1 focuses on the analogy of data as commodities; Sub-section 5.1.2 presents data as common pool resources. Access to resources enables cumulative and combinatorial innovation. But data is also an infrastructure for digital ecosystems: this is the subject of Section 5.2. The relational nature of data, and the embedded generative potential, serve the economic players to organize the world. The infrastructural role of data enables generative innovation, which has a greater chance of having disruptive effects on society. As illustrated in Section 4, the ecosystem-mediated relationship among the firms participating in an innovation common determines the characteristics of the resulting innovation. As a consequence, different functions for data can prevail at different levels of the ecosystem. Policies aiming at fostering innovation commons are expected to appropriately select the narrative they adopt towards data and match it with the role of the firms that are expected to partake in the commons and the characteristics of the desired innovation.

5.1 Data as a Resource

Data points are the immaterial but fundamental inputs of the data economy. Two analogies represent the relational implications of data as an economic resource: when data is equated to commodities (Sub-section 5.1.1), a light is shed on its competitive consumption; when data is linked to common pool resources (Sub-section 5.1.2), attention is brought to the importance of cooperation.

5.1.1 Commodities

The metaphor that had the most powerful grip on public opinion is certainly one that associates data with a commodity. "Data is the new oil" is a sentence that rapidly surged to the position of workplace litany, when not a company mantra, to the point that it is considered by many a tired cliché (Gilbert, 2021). In highlighting the fundamental role of data in fueling companies' growth, the metaphor pairs a descriptive claim with a normative one. Neglecting data management becomes the equivalent, for a company, of forgetting to fill the tank of the car. Data is freed by its technical

aura and normalized, becoming part of firms' daily operations as raw materials, with their supply taking a central role in the development of the business strategy. Oil is seldom substituted by gold, highlighting the value that the resource has for businesses. The other facet of the same medal is the analogy equating data to a currency that consumers spend inadvertently. This metaphor provides a tentative explanation for the rapid rise of Big Techs, firms whose business model does not usually involve direct payment by the consumer. The mystery of how those firms could produce value while offering a free service is thus quickly solved: users do not pay with money but with data. Data is likened to a currency.³⁷ It provides companies with a resource whose value they are unaware of, and Big Tech companies able to extract data from users can resell it by making an immense profit. An illustration published on the cover of *The Economist* in 2017 perfectly represents the metaphor. Big Techs are drawn in the guise of oil platforms, a wordplay based on their status as online platforms. The implicit admonition contained in the analogy warns against the free and unconscious transfer of valuable goods that accompanies many digital actions.

The success of the data is the new oil, gold, or currency metaphors can be due to their ability to provide a characterization of data that reflects and explains their behavior as observed by individual users and professionals in their daily experiences. However, its deconstruction reveals that it moves from a series of implicit assumptions, transferring to data a set of non-trivial economic properties. Data is a scarce resource that has to be extracted, possesses commercial value and is fungible. Moreover, the amount of data at their disposal determines businesses' competitive strength: no market player will be willing to make its data available to other firms (Graef, 2016). As a matter of fact, due to the high fungibility of data, companies operating in the data economy can quickly expand their activities in different markets. This represents a considerable obstacle to data sharing. Firms considering its data a commodity are not likely to make it accessible for re-use. The risk they incur is a loss of competitive advantage.

The economic function of data is (or is considered to be)³⁸ the one of commodities when firms are subject to competitive threats. In this regard, it is important to observe the multidimensionality of competition in the digital ecosystems. Static, price-based competition can take place horizontally among complementors offering substitutable products; vertical intra-ecosystem competition refers to value captured through joint collaboration; innovation-based competition takes place between different ecosystems that offer comparable value added to customers through the provision of multiple products. The ecosystem theory reaffirms that the assessment of firms' competitive advantage cannot overlook the analysis of the aggregate level. In any case, the extraction and accumulation of data having the function of a commodity support cumulative evolution within the boundaries of the module—be it the single firm or the ecosystem vis-à-vis competing ecosystems.

³⁷ Currencies are specific cases of commodities. On this topic, see Clements and Fry (2008)

³⁸ In this regard, remember that self-perception is as relevant as reality. Members of the ecosystem define their strategies based on their bounded rationality.

5.1.2 Common Pool Resources

Multiple scholarly contributions on the data economy equated data to a common pool resource.³⁹ Data is non-rival: consumption by one actor does not prevent re-use by others. In addition, it has limited excludability: it is common practice to limit data access through the adoption of technical barriers. Data can be kept a secret. Technical solutions meant to control third-party access are commonly adopted by companies operating in the data economy. And, if it is undeniable that cybersecurity attacks can undermine the efficacy of such solutions, it is pairwise true that so far the large data holders such as Google, Meta, and others “do not seem to suffer from a vast copying and leaking of their huge amount of collected data” (Kerber & Schweitzer, 2017). Goods characterized by non-rivalry and limited excludability are considered to be impure public goods (Leach, 2003). Thus, the most appropriate framework for the analysis would thus be the one of common pool resource (CPR) elaborated by the Nobel laureate academic Elinor Ostrom, who in 2003 co-authored a seminal paper studying information as a CPR (Hess & Ostrom, 2003). Innovation is possible when data governance can overcome collective action problems associated with data access.⁴⁰

The CPR perspective focuses on the social need for firms’ cooperation. Data that holds little to no value for the firm that held their access are often left abandoned in data lakes or data swamps. If such data was to be shared, new firms could use the same data points in a different domain. Thus, combinatorial evolution could surge.

5.2 Data as an Infrastructure

The last analogy equates data to an infrastructure. In the same way of roads and electric systems, data comprises “a backbone for much of modern social and economic activity” (Ruhaak, 2020). Considering the multitude of downstream innovations made possible by the use (and re-use) of data, its wide availability will foster technological progress and increase social welfare. Data can be considered a multi-purpose resource (OECD, 2015). The value of data in contexts different than the one in which it is originated may be difficult to assess, given that the value of information is context-dependent, but it is most likely positive (OECD, 2015). The infrastructural angle raises attention to the transformative effects that data-driven innovation entails on the economy and society. Relatedly, the emergence of such characterization of data is more recent: it moves from the recognition that the disruptive changes brought by the quick technological advances in the data economy affected a multiplicity of actors, many of whom in unforeseeable ways. As such, the analogy sheds light on the multiplicity of technological trajectories typical of

³⁹ This assumption is particularly prominent in (but not limited to) sector-specific research. See, for instance, (Dulong de Rosnay & Le Crosnier, 2012; Erik de Man, 2006; Fia, 2021; Reichman & Uhlir, 2003; Šestáková & Plichtová, 2019).

⁴⁰ The very same framing of “innovation commons” adopted in Section 2 derives from the treatment of data as CPRs.

the digital economy. The value of an infrastructure varies depending on the activity that it enables. When data is linked to infrastructures, then, the focus is not on their use value nor their commercial value; the policy discourse is centered around their potential value (Ducuing, 2020). Brett Frischmann proposes a three-step test to define an intangible as an infrastructural resource: it should be non-rivalrous in consumption to some appreciable extent; social demand for it shall be driven primarily by downstream product activities that require the resource as an input; it shall serve as input for a wide range of goods, be they commercial or non-commercial. Infrastructures are “used by many different users, with the usage evolving over time, as may the type of users” (Frischmann, 2012). The combination of the three criteria results in the definition of intellectual infrastructure as “non-rival inputs for a wide variety of outputs” (Frischmann, 2012). The OECD underlines that such a definition perfectly describes the nature of data (OECD, 2015). Indeed, the value from data is created subsequently to their transfer and in relation to reuse.

The economic function of data as an infrastructure potentially enables the flourishing of generative innovation. Infrastructures are a way in which interconnected systems can be conceptualized (Henfridsson et al., 2013). In digital ecosystems, data connect the modules with the rest of the ecosystem. But data also transmits elements of the social context in which they have been co-created. As Star and Ruhleder write, “an infrastructure occurs when the tension between local and global is resolved” (1996). The data to which players in the ecosystem have access determines the technological landscape that they can explore; at the same time, the distribution of data in the ecosystem defines the connections with the other economic players. Gray renames the data infrastructure “data worlds”: his invite is to not consider data exclusively as resources, but to investigate how political, social, and cultural values emerge from data infrastructures (Gray, 2017).⁴¹ Data worlds provide the horizon of intelligibility; the information to which each participant in the ecosystem has access defines its ability to move in the world. Ultimately, they decide the direction of innovation of the ecosystem and, as such, the well-being of society.

Data worlds, according to Gray, offer transnational coordination (Gray, 2017). Data as infrastructures, indeed, assist the governance of the ecosystem. They are an instrument for the ecosystem leader in its quest to maintain alignment of complementors’ interests. Ecosystem leaders influence the architecture of the ecosystem through the design of their products. Usually, this includes a certain influence on the definition of standards and rules for interoperability. This directly affects the allocation of data, shaping the modules of the ecosystem; in turn, size and relations between modules strongly affect innovation trajectories. The starker the influence of the leader over the ecosystem’s infrastructure, the more power will it have over the orchestration of data

⁴¹ The locution “data world” originated in the context of research on open data; it can nonetheless easily be transferred to open ecosystems. Note that the expression “data infrastructures” is used by Gray in the sense of “socio-technical arrangements which underpin the production of data: consisting of relational ecologies of software components, data standards, methods, techniques, committees, researchers, instruments and other things”. The relational components makes the concept overlap with our definition of data as an infrastructure.

resources. The higher chances of successful generative innovation materialize when data infrastructures are open, participative, and dynamic.

The same data serves the function of resource and infrastructure. The perceptions of the players, together with contextual use-related factors, determine the economic players' attitude towards cooperation through data commons. The same metaphors guide regulators' interpretation of the complex and dynamic digital economy. It helps them understand which function does data hold. If data is perceived as a commodity, a CPR, or an infrastructure, the legislative focus will be drawn to different ecosystem levels. The kind of innovation encouraged by the regulatory intervention will thus differ. Metaphors are abstract guides for the interpretation of reality but entail material effects. The three above-mentioned metaphors for data address a need from a public policy perspective. Paraphrasing the OECD, they "provide a framework that can guide policymakers in identifying when data warrant their attention" (OECD, 2015, p. 178). The next Section 6 attempts to sketch the consequences stemming from the adoption of one metaphor other than another; further research is needed to disentangle the implicit assumptions driving policymakers in the design of innovation commons.

6 Policy Implications

Data enables innovation. How it happens is mediated by business relationships in ecosystems. The analysis presented in this paper invites us to adopt a more granular view of the data economy. Cooperation through data commons has different costs and foreseen benefits depending on whether data are treated by ecosystems' members as resources or if they represent an infrastructure. Taking an ecosystem perspective assists in identifying areas and modalities of intervention able to leverage existing incentives. Different options of data governance emerge as possible. Policies aiming to foster innovation commons shall take into account the structure of the business ecosystem they intend to address to (1) be effective, and (2) prevent unexpected long-time effects on the ecosystem governance, affecting multiple actors (Cennamo, 2021).

Different regulatory solutions may foster innovation commons in digital ecosystems. When data has the role of a commodity, the starting point of the legislative discussion lies in the recognition that data hold commercial value. Therefore, the regulator adopting this point of view is likely to increase transparency in the market, so that consumers are fully aware of the economic value of the data they are creating. Relatedly, the legal framework is charged with the task of facilitating the emergence of a healthy and well-functioning market for data. Although a certain degree of concentration among data providers appears as unavoidable, directly stemming from the characteristics of the commodity extracted and supplied, the role of legal institutions driven by the willingness to foster commons for commodity data is to ensure access to data to all the companies operating in downstream markets. Those companies will be enabled to use them to fuel the provision of new and better services, generating growth and innovation. If the grip that the analogy holds among managers and the public is reflected in the policymaking, we can expect the

legislation to focus on the commercial value of data, providing incentives for their exchange in huge quantities, considering the fungibility they have across different markets and ultimately ensuring that all the companies transforming them in goods and services have access to this essential commodity.

The GDPR contains a curious case of the data-as-commodity metaphor. Art.20 of the Regulation establishes the right to “data portability,” giving the data subject the right to receive the personal data concerning her “in a structured, commonly used and machine-readable format” and to transmit those data to another controller. This particular article appears to pursue a different goal to the rest of the GDPR: while the remainder of the regulation builds the foundation of a fundamental right to data protection, art. 20 seems to be guided by the desire to push growth and competitiveness. One may go as far as arguing that, by facilitating the transfer of data from and to competing controllers, data portability could assist incremental innovation. Each player could rely on the same resources and independently pursue their innovation strategy. However, the GDPR does not contain any indication that supports the trade of such a precious commodity.⁴² The result is an almost forgotten, never enforced, and arguably ineffective provision.

The data-as-CPR analogy designs a clear priority for the regulator that adopts it: to foster data sharing by increasing control over data. The attention is drawn to issues of data undersupply due to collective action problems determining an inefficient use (and re-use) of data. The regulator will aim at facilitating data production and reuse by correcting economic agents’ failure to spontaneously negotiate the optimal level of the good. The assignment of clear and defined control over data may be considered an unavoidable step. A specific declination of this view is offered by Birch when he refers to the *assetization* of data (Birch et al., 2020). The author problematizes innovation as being increasingly driven by the pursuit of rents, and proposes laws that protect (personal) data subjects with time-limited property rights.⁴³

The data-as-CPR analogy appears to have guided the European Commission in the drafting of the Data Governance Act, adopted in May 2022. The objective of the Regulation is the stimulation of innovation through the establishment of a clear framework for data reuse and sharing across specific sectors. Among the many provisions, several regard the so-called “data-altruism”—which is, the voluntary disclosure of personal data by data subjects. For example, Art. 17 regulates public

⁴² Indeed, the European Commission has clearly indicated in the EU Horizontal Provisions on Cross-Border Data Flows and Protection of Personal Data and Privacy in the Digital Trade Title of EU Trade Agreements that the protection of personal data is fundamental right, non-negotiable in the context of trade agreements. See COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL Data Protection as a Pillar of Citizens’ Empowerment and the EU’s Approach to the Digital Transition—Two Years of Application of the General Data Protection Regulation (2020).

⁴³ For an example, see the analysis of the regulatory framework for the access to interconnected vehicles’ data by Kerber & Moeller (2019). According to them “[...] There is a broad consensus that the crucial challenge for competition on the markets for repair and maintenance services in the ecosystem of connected driving is the exclusive control of the OEMs (i.e. car manufacturers) of the access to in-vehicle data and the connected car” (p. 9). An appropriate response to the challenges of digitalization in the automotive industry, thus, requires the granting of control to independent economic players.

registers of recognized data altruism organizations. Data is considered to be a necessary input, and the regulator has the duty to overcome the organizational and technical obstacles that impede innovation commons. However, access to data is provided only to certain pre-determined categories of economic players. As such, the resulting innovation can only be cumulative or, at best, combinatorial.

Lastly, when regulators recognize data as the infrastructure of the ecosystem, they will design institutions to make it available for use in a non-discriminatory way. They will have to address a problem connected with the general-purpose nature of infrastructures: as the value that will be produced through infrastructures cannot be known *ex-ante*, public policies should ensure that they are sufficiently produced (OECD, 2015). All the interested third parties could then profit from the resource, unlocking a wide and unpredictable range of downstream activities. Regulation towards infrastructure should be particularly mindful: data, in this declination, govern and enable the relationships among members and design the space in which they operate. The data infrastructure contributes to shaping ecosystems' incentives, facilitating cooperation, and making generative innovation possible. Ecosystem leaders influence such infrastructure by governing the ecosystem and designing its architecture. Standards-setting and interoperability multiply the relations among the ecosystem members and govern their contacts with external (competing) ecosystems. Regulation underpinned by the infrastructure metaphor is likely to promote new standards and mandate interoperability. In the context of *ex-post* enforcement, considering data's role as an infrastructure leads to measuring Big Tech's impact (also) against the relational impact that it holds.

The recognition of the infrastructural function of data is quite recent. However, it is already possible to uncover instances in which the European regulator, more or less consciously, has looked at the market through the lenses of this metaphor. It is the case of the access-to-account rule (XS2A) contained in the Open Banking Directive (PSD2). The rule mandates incumbents (usually traditional banks) to disclose information on users' accounts to third-party providers (prior authorization of the users themselves). The rationale behind it is explained by the legislator's intention of facilitating the entrance of new agents.⁴⁴ By improving the level playing field for payment service providers, the XS2A rule does not only improve conditions for the entrance of banks' direct competitors, but it promotes the flourishing of payment initiation services and account information services. Consumers could thus benefit from an infinite range of new products: new apps for managing expenses complement the banks' product, new banks can more easily secure consumers or completely new services can arise. *Ex-post*, it appears that the latter case has been the most favored by the Directive. The PSD2 enabled, in particular, the blooming of "PayTech" companies (Polasik et al., 2020): a plurality of niche-targeted and varied services that offer customers previously inexistent tools. In this sense, the XS2A rule seems to have favored the emergence of generative, intra-module, innovation.

Through the XS2A, the European regulator intends to foster innovation by making data available not only to operators in downstream markets for which it

⁴⁴ Directive (EU) 2015/2366, Art. 67.

Table 1 Characterizations of data as commodities, public goods, and infrastructure

	Data as commodities	Data as public goods	Data as infrastructure
Source of value	Commercial value	Use value	Potential value from future applications
Underlying assumptions	<ul style="list-style-type: none"> - Users are data producers, from whom companies <i>extract</i> data; - Data needs to be <i>refined</i> before sale; - Data are <i>fungible</i>; - Data are an <i>essential asset</i> 	<ul style="list-style-type: none"> Pure public goods; - Data are <i>non-excludable</i>; - Data are <i>non-rival</i> Impure public goods: - Data are rival but non-excludable; - Data are non-rival but excludable 	<ul style="list-style-type: none"> - Data are <i>non-rival</i>; - Data are <i>inputs</i> for downstream applications; - Data enable a <i>wide and unpredictable</i> array of goods/services
Objectives of the legislation	<ul style="list-style-type: none"> - Promote transparency and fair compensation of the data producers; - Facilitate the establishment of a well-functioning market for data; - Ensure access to data by all the companies operating in downstream markets 	<ul style="list-style-type: none"> - Solve issues of underproduction and/or undersupply; - Enable coordination of collective actions problems; - Assign clear control over data 	<ul style="list-style-type: none"> - Promote non-discriminatory third-party access to data; - Promote interoperability and standardization

represents a necessary raw material but also to newcomers who can use such a resource in ways that the directive is unable to foresee or restrict. This is consistent with the narrative that equates data to infrastructure and advocates that institutions should make sure that it is made available for use in a non-discriminatory manner. This way, all the interested third parties could profit from the resource, unlocking a wide and unpredictable range of downstream activities. However, it needs to be noted that no unique standard was defined for payment service providers' APIs. The European Banking Authority was designated to draft the regulatory technical standards indicated in the legislative act, which were subsequently approved by the Commission. The definition of standards is a fundamental step for the success of any legislative intervention which aims to foster data infrastructures: the inappropriate design of APIs would have undermined the favorable outcome of the directive (Borgogno, 2019).

Standards play, indeed, a major role in defining the architecture of modularized systems. They influence the thickness of the transaction points, hence incentives to entry. The introduction of an alternative standard, determining a different organization of modules, is a major push for intra-ecosystem competition. The introduction of disruptive innovation is made possible by architectural shifts. As such, standard-setting intended to foster innovation and competition shall necessarily move from the definition of which kind of innovation and competition represents their goal. Incremental, combinatorial, or generative innovation? Horizontal intra-ecosystem competition, vertical intra-ecosystem competition, or inter-ecosystem competition?

Further research is needed to better delineate the policy implications stemming from the adoption of a metaphor other than another. Table 1 summarizes the findings enucleated in this section.

But the examples presented above merely scratch the surface of the complex legal analysis needed to establish which function of data is recognized behind a legislative act. Additionally, although the regulation of the digital economy is a vibrant and rapidly evolving branch of the law, the majority of the legislative interventions have been enacted too recently to draw more than initial considerations on their effects. In this article, I offered a theoretical framework that awaits further testing. The next Section 7 offers some preliminary conclusions.

7 Conclusions

Data connects digital firms in a complex net. Its understanding is challenging: it requires taking into account the multiple horizons of intelligibility that coexist in the production (and fruition) of a single product. However, such disentanglement represents a necessary step to fostering cooperative innovation. The exchange of (data) resources, together with the smooth coordination of independently achieved technological progress, promises to advance society's well-being in the form of generative innovation. The promotion of data-driven innovation is with good reason a goal of European policies for the digital sector. But the regulation of digital markets shall include, as a preliminary step of the legislative intervention, the outline of the kind of innovation that is intended to achieve. By altering the distribution of data in

the ecosystem, regulation may modify the incentives to compete and cooperate. As such, a context-dependent approach shall be favored. Different dimensions of competition (horizontal inter-ecosystem, vertical inter-ecosystem, intra-ecosystem) may be dependent on different kinds of innovation (cumulative, combinatorial, generative). Further research is needed to examine the inevitable trade-offs among them, and the balancing actions available for government intervention. Ultimately, the regulation of the data economy should rest on the acknowledgment that the future trajectory of innovation depends upon the lenses through which the complexity of data worlds is cognized. Setting the rules means having a voice in the narrative guiding the development of technology, a narrative that, in conclusion, will be chorally co-created by the many participants of digital ecosystems.

Funding Open access funding provided by European University Institute - Fiesole within the CRUI-CARE Agreement. The author receives a grant from the Italian Ministry of Foreign Affairs and International Cooperation.

Data Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Informed Consent The author has the consent of the Institution where the work was carried out. The author was the sole contributor to the submitted paper.

Competing Interest The author declares no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). Competition and innovation: An Inverted-U relationship. *The Quarterly Journal of Economics*, 120(2), 701–728.
- Allen, D., & Potts, J. (2015). The innovation commons why it exists, what it does, who it benefits, and how. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2617141>
- Allen, D., & Potts, J. (2016). How innovation commons contribute to discovering and developing new technologies. *International Journal of the Commons*, 10(2).
- Anderlini, L., Felli, L., Immordino, G., & Riboni, A. (2013). Legal institutions, innovation, and growth. *International Economic Review*, 54(3), 937–956.
- Anderson, C. (2008, June 23). *The end of theory: The data deluge makes the scientific method obsolete*. Wired. <https://www.wired.com/2008/06/pb-theory/>
- Arrow, K. J. (1962). Economic welfare and the allocation of resources for invention. In National Bureau of Economic Research, *The Rate and Direction of Inventive Activity* (pp. 609–626). Princeton University Press.

- Arthur, B. (1983). On competing technologies and historical small events: The dynamics of choice under increasing returns. *The Economic Journal*, 99(394), 116–131.
- Arthur, W. B. (2014). *The nature of technology: What it is and how it evolves*. Free Press.
- Baldwin, C. Y. (2007). Where do transactions come from? Modularity, transactions, and the boundaries of firms. *Industrial and Corporate Change*, 17(1), 155–195.
- Baldwin, C. Y., & Clark, K. B. (2000). *Design rules: The power of modularity*. The MIT Press.
- Bar-Yam, Y. (n.d.). *Significant points in the study of complex systems*. Retrieved April 12, 2022, from <https://necsi.edu/significant-points>
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623.
- Benkler, Y. (2002). Coase's Penguin, or, Linux and 'The Nature of the Firm.' *The Yale Law Journal*, 112(3), 369–446.
- Benkler, Y. (2003). Freedom in the commons: Towards a political economy of information. *Duke Law Journal*, 52.
- Birch, K., Chiappetta, M., & Artyushina, A. (2020). The problem of innovation in technoscientific capitalism: Data rentiership and the policy implications of turning personal digital data into a private asset. *Policy Studies*, 41(5), 468–487.
- Black, M. (1955). Metaphor. *Proceedings of the Aristotelian Society, New Series*, 5, 273–294.
- Borgogno, O. (2019). Data sharing and interoperability: Fostering innovation and competition through APIs. *Computer Law & Security Review*, 35.
- Borrás, S. (2003). *The innovation policy of the European Union: From government to governance*. Edward Elgar Publishing.
- Brandenburger, A., & Nalebuff, B. (1998). *Co-opetition*. Doubleday.
- Bresciani, S., Ciampi, F., Meli, F., & Ferraris, A. (2021). Using big data for co-innovation processes: Mapping the field of data-driven innovation, proposing theoretical developments and providing a research agenda. *International Journal of Information Management*, 60.
- Briscoe, G. (2010). Complex adaptive digital EcoSystems. *Proceedings of the International Conference on Management of Emergent Digital EcoSystems - MEDES '10*, 39.
- Brooks, H. (1994). The relationship between science and technology. *Research Policy*, 23(5), 477–486.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Amodei, D. (2020). Language models are few-shot learners. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, & H. Lin (Eds.), *Advances in neural information processing systems* (Vol. 33, pp. 1877–1901). Curran Associates Inc. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfc94967418bfb8ac142f64a-Paper.pdf
- Buchholtz, S., Bukowski, M., & Sniegocki, A. (2014). Big and open data in Europe: A growth engine or a missed opportunity. *Warsaw Institute for Economic Studies Report Commissioned by DemosEUROPA*, 10, 116.
- Burgelman, R. A., McKinney, W., & Meza, P. E. (2017). *Becoming Hewlett Packard: Why strategic leadership matters*. Oxford University Press.
- Cennamo, C. (2021). Competing in digital markets: A platform-based perspective. *Academy of Management Perspectives*, 35(2), 265–291.
- Chomsky, N., & Chomsky, I. P. & P. of L. (Emeritus) N. (2006). *Language and mind*. Cambridge University Press.
- Clements, K. W., & Fry, R. (2008). Commodity currencies and currency commodities. *Resources Policy*, 33(2), 55–73.
- Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics*, 1–44.
- Communication from the Commission to the European Parliament and the Council. (2020). *Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition—Two years of application of the General Data Protection Regulation*. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0264>
- Crawford, K., Miltner, K., & Gray, M. L. (2014). Critiquing big data: Politics, ethics, epistemology. *International Journal of Communication*, 8, 1663.
- Dalton, C. M., Taylor, L., & Thatcher (alphabetical), J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1).

- De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65(3), 122–135.
- Dequech, D. (2004). Uncertainty: Individuals, institutions and technology. *Cambridge Journal of Economics*, 28(3), 365–378.
- Ding, Y., & Stirling, K. (2016). Data-driven discovery: A new era of exploiting the literature and data. *Journal of Data and Information Science*, 1(4), 1–9.
- Ducuing, C. (2020). Data as infrastructure? A study of data sharing legal regimes. *Competition and Regulation in Network Industries*, 21(2), 124–142.
- Dulong de Rosnay, M., & Le Crosnier, H. (2012). *An introduction to the digital commons: From common-pool resources to community governance*. <https://biogov.uclouvain.be/iasc/index.php?page=fullpapers.%20ffhshs00736920>
- Erik de Man, W. H. (2006). Understanding SDI; complexity and institutionalization. *International Journal of Geographical Information Science*, 20(3), 329–343.
- Ethiraj, S. K., & Levinthal, D. (2004). Modularity and innovation in complex systems. *Management Science*, 50(2), 159–173. LEM Papers Series, Article 2003/15.
- Fia, T. (2021). An alternative to data ownership: Managing access to non-personal data through the commons. *Global Jurist*, 21(1), 181–210.
- Fleming, L. (2001). Recombinant uncertainty in technological search. *Management Science*, 47(1), 117–132.
- Friedman, M. (1953). *The methodology of positive economics: Vol. Essays In Positive Economics* (pp. 3–16; 30–43). Univ. of Chicago Press.
- Frischmann, B. M. (2012). *Infrastructure: The social value of shared resources*. Oxford University Press.
- Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 43(7), 1239–1249.
- Generative AI: A game-changer society needs to be ready for. (2023, January 9). World economic forum. <https://www.weforum.org/agenda/2023/01/davos23-generative-ai-a-game-changer-industries-and-society-code-developers/>
- Gershenson, J. K., Prasad, J., & Zhang, Y. (2003). Product modularity: Definitions and benefits. *Journal of Engineering Design*, 14(3), 295.
- Gilbert, R. J. (2006). Competition and innovation. *Journal of Industrial Organization Education*, 1(1), 1–23.
- Gilbert, S. (2021). *Good data: An optimist's guide to our digital future*. Welbeck.
- Google LLC and Alphabet, Inc v European commission, Case T-604/18 (GC 14 September 2022). <https://eurlex.europa.eu/legal-content/en/TXT/?uri=CELEX:62018TJ0604>
- Graef, I. (2016). *EU competition law, data protection and online platforms: Data as essential facility*. Kluwer Law International B.V.
- Graham, M. (2012, March 9). *Big data and the end of theory?* The Guardian. <https://www.theguardian.com/news/datablog/2012/mar/09/big-data-theory>
- Gray, J. (2017). Redistributing data worlds: Open data, data infrastructures and democracy. *Open Data*, 10.
- Griliches, Z. (1992). The search for R&D spillovers. *The Scandinavian Journal of Economics*, 94, S29–S47.
- Guo, J., Cui, L., Sun, S. L., & Zou, B. (2022). How to innovate continuously? Conceptualizing generative capability. *Journal of Innovation & Knowledge*, 7(2), 100177.
- Harper, D. A. (2020). 'New combinations' in Schumpeter's economics: The lineage of a concept. *History of Economics Review*, 75(1), 22–30.
- Henfridsson, O., Bygstad, B., Norwegian School of IT, & University of Oslo. (2013). The generative mechanisms of digital infrastructure evolution. *MIS Quarterly*, 37(3), 907–931.
- Hess, C., & Ostrom, E. (2003). Ideas, artifacts, and facilities: Information as a common-pool resource. *Law and Contemporary Problems*, 66, 111.
- Hylving, L., & Schultze, U. (2013). *Evolving the modular layered architecture in digital innovation: The case of the car's instrument cluster*. ICIS 2013 Proceedings. <https://aisel.aisnet.org/icis2013/proceedings/OrganizationIS/13>
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2), 2053951716674238.
- Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an open data based business ecosystem. *IEEE Access*, 2, 88–103.

- Jacobides, M. G., Cennamo, C., & Gawer, A. (2018). Towards a theory of ecosystems. *Strategic Management Journal*, 39(8), 2255–2276.
- Jacobides, M. G., & Lianos, I. (2021). Ecosystems and competition law in theory and practice. *Industrial and Corporate Change*, 30(5), 1199–1229.
- Jones, B. F., & Summers, L. H. (2020). *A calculation of the social returns to innovation*. NBER. Working Paper No. 27863.
- Kent, C. (November 30, 2021). The Watch That Made Everything Now. *Wired*. Retrieved February 2, 2023, from <https://www.wired.com/story/pulsar-digital-watch-design-time-history/>
- Kerber, W., & Moeller, D. (2019). Access to data in connected cars and the recent reform of the motor vehicle type approval regulation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3406021>
- Kerber, W., & Schweitzer, H. (2017). Interoperability in the digital economy. *JIPITEC*, 8(1).
- Leach, J. (Ed.). (2003). Impure public goods. In *A Course in Public Economics* (pp. 187–199). Cambridge University Press.
- Lessig, L. (2002). The architecture of innovation. *Duke Law Journal*, 51(6), 1783.
- Lundqvist, B. (2018). Competition and data pools. *Journal of European Consumer and Market Law*, 146–154.
- Luo, J. (2023). Data-driven innovation: What is it. *IEEE Transactions on Engineering Management*, 70(2), 784–790.
- Macdonald, S. (1998). *Information for innovation: Managing change from an information perspective*. OUP Oxford.
- Mansfield, E., Rapoport, J., Romeo, A., Wagner, S., & Beardsley, G. (1977). Social and private rates of return from industrial innovations. *The Quarterly Journal of Economics*, 91(2), 221–240.
- Mazzocchi, F. (2015). Could big data be the end of theory in science? *EMBO Reports*, 16(10), 1250–1255.
- Minevich, M. (January 29, 2023). *The generative AI revolution is creating the next phase of autonomous enterprise*. Forbes. Retrieved February 2, 2023, from <https://www.forbes.com/sites/markminevich/2023/01/29/the-generative-ai-revolution-is-creating-the-next-phase-of-autonomous-enterprise/>
- MIT Technology Review Insight & Infosys Cobalt. (2021). *Capitalizing on the data economy*. Available at <https://www.technologyreview.com/2021/11/16/1040036/capitalizing-on-the-data-economy/>
- Moore, F. (2006). Business ecosystems and the view from the firm. *The Antitrust Bulletin*, 51(1), 31–75.
- Murmann, J. P., & Frenken, K. (2006). Toward a systematic framework for research on dominant designs, technological innovations, and industrial change. *Research Policy*, 35(7), 925–952.
- Nelson, R. (1962). Introduction to “the rate and direction of inventive activity: Economic and social factors”. In *The Rate and Direction of Inventive Activity: Economic and Social Factors* (pp. 1–16). Princeton University Press.
- OECD. (1992). *Proposed guidelines for collecting and interpreting technological innovation data*. Paris: OECD. Oslo manual, OCDE/GD (92)26.
- OECD. (2015). Data-driven innovation: Big data for growth and well-being. https://read.oecd-ilibrary.org/scienceand-technology/data-driven-innovation_9789264229358-en
- OECD. (2017, April 26). *Big data: Bringing competition policy to the digital era*. Executive Summary. DAF/COMP/M(2016)2/ANN4/FINAL. [https://one.oecd.org/document/DAF/COMP/M\(2016\)2/ANN4/FINAL/en/pdf](https://one.oecd.org/document/DAF/COMP/M(2016)2/ANN4/FINAL/en/pdf)
- OECD. (2022). *OECD handbook on competition policy in the digital age*. <https://www.oecd.org/daf/competition-policy-in-the-digital-age/>
- Oliveira, M. I. D. S., Barros Lima, G. D. F., & Farias Lóscio, B. (2019). Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems*, 61(2), 589–630.
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Ostrom, E. (2005). *Understanding institutional diversity*. Princeton University Press.
- Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. *Communications of the ACM*, 15(12), 1053.
- Petit, N., & Schrepel, T. (2023). Complexity-minded antitrust. *Journal of Evolutionary Economics*, 33(2), 541–570.

- Petit, N., & Teece, D. J. (2020). Taking ecosystems competition seriously in the digital economy: A (Preliminary) dynamic competition/capabilities perspective. *OECD DAF/COMP/WD(2020)90*.
- Pigliucci, M. (2009). The end of theory in science? *EMBO Reports*, 10(6), 534.
- Pil, F. K., & Cohen, S. K. (2006). Modularity: Implications for imitation, innovation, and sustained advantage. *Academy of Management Review*, 31(4), 995–1011.
- Polasik, M., Huterska, A., Iftikhar, R., & Mikula, Š. (2020). The impact of Payment Services Directive 2 on the Paytech sector development in Europe. *Journal of Economic Behavior & Organization*, 178, 385–401.
- Potts, J. (2018). Governing the innovation commons. *Journal of Institutional Economics*, 14(6), 1025–1047.
- Rawls, J. (1999). *A theory of justice* (Rev. ed). Belknap Press of Harvard University Press.
- Reichman, J. H., & Uhlir, P. F. (2003). A contractually reconstructed research commons for scientific data in a highly protectionist intellectual property environment. *Law and Contemporary Problems*, 66, 149.
- Ricœur, P. (2003). *The rule of metaphor the creation of meaning in language*. Routledge Classics.
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169.
- Robertson, V. H. S. E. (2021). Antitrust market definition for digital ecosystems. *Concurrences*, 2, 3. SSRN Scholarly Paper No. 3844551.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102.
- Ruhaak, A. (2020, May 28). *Data commons & data trust*. Medium. <https://medium.com/@anoukruhaak/data-commons-data-trust-63ac64c1c0c2>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1), 205395171882054.
- Scholz, T. M. (2017). *Big data in organizations and the role of human resource management*. Peter Lang D.
- Schön, D. A. (1993). Generative metaphor: A perspective on problem-setting in social policy. In A. Oliveira, M. I., Barros Lima, G. de F., & Farias Lóscio, B. (2019). Investigations into Data Ecosystems: A systematic mapping study. *Knowledge and Information Systems*, 61(2), 589–630. Ortony (Ed.), *Metaphor and Thought* (2nd ed., pp. 137–163). Cambridge University Press.
- Schumpeter, J. A. (1926). *The theory of economic development*. Transaction Publishers.
- Šestáková, A., & Plichtová, J. (2019). Contemporary commons: Sharing and managing common-pool resources in the 21st century. *Human Affairs*, 29(1), 74–86.
- Sharon, T. (2021). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, 23(1), 45–57.
- Shaughnessy, H. (2019, August 18). *Business ecosystems. The sudden importance of ecosystems*. Medium. <https://medium.com/@haydnshaughnessy/business-ecosystems-c5a2f4ad86c>
- Shiller, R. J. (2019). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society*, 106.
- Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70(1), 65–94.
- Star, S. L., & Ruhleder, K. (1996). Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1), 111–134.
- Teece, D. J. (1992). Competition, cooperation, and innovation. *Journal of Economic Behavior and Organization*, 18, 1.
- Tirole, J., & Rendall, S. (2017). *Economics for the common good*. Princeton University Press.
- Um, S., Yoo, Y., Wattal, S., Kulathinal, R., & Zhang, B. (2013). *The architecture of generativity in a digital ecosystem: A network biology perspective*. Thirty Fourth International Conference on Information Systems. <https://core.ac.uk/download/pdf/301361204.pdf>
- UNECE. (2020). *RECOMMENDATION No. 24, second edition, adopted by the United Nations Centre for Trade Facilitation and Electronic Business (UN/CEFACT)*. https://unece.org/DAM/cefact/recommendations/rec24/rec24_ecetrd258e.pdf
- Van Den Hoven, J., & Rooksby, E. (2008). Distributive justice and the value of information: A (Broadly) Rawlsian approach. In J. van den Hoven & J. Weckert (Eds.), *Information Technology and Moral Philosophy* (pp. 376–396). Cambridge University Press.

- Vedder, B. (2002). On the meaning of metaphor in Gadamer's hermeneutics. *Research in Phenomenology*, 32, 196–209.
- Verendel, V. (2023). Tracking AI in climate innovation with Patent Data. *Nature Climate Change*, 13(1), 40–47.
- West, J. (2009). Policy challenges of open, cumulative, and user innovation. *Washington University Journal of Law & Policy*, 30(1), 17–42.
- Winsor, J. (December 13, 2019). *The grandma test: Are you explaining what You do in a way that makes sense?* Forbes. Retrieved April 12, 2022, from <https://www.forbes.com/sites/johnwinsor/2019/12/23/the-grandma-test-are-you-explaining-what-you-do-in-a-way-that-makes-sense/>
- Winters, J. (2020). *Is the cultural evolution of technology cumulative or combinatorial?* SocArXiv Papers, aypnx. <https://doi.org/10.31235/osf.io/aypnx>
- Yildiz, D. M. (2022, November 11). *Here's what combinatorial innovation mean.* EUPHORIA. <https://medium.com/sensible-biohacking-transhumanism/combinatorial-innovation-16e6cefd6163>
- Yoo, Y., Boland, R. J., Lyytinen, K., & Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization Science*, 23(5), 1398–1408.
- Zech, H. (2016). *Data as a tradeable commodity* (pp. 51–79). Intersentia. A. De Franceschi (Ed.).
- Zittrain, J. L. (2005). The generative internet. *Harvard Law Review*, 119(7), 1975–2040.