Social welfare gains from innovation commons:

Theory, evidence, and policy implications

Jason Potts, Dietmar Harhoff, Andrew Torrance, and Eric von Hippel

September, 2021

Abstract

Innovation commons – which we define as repositories of freely-accessible, "open source" innovation-related information and data - are a very significant resource for innovating and innovation-adopting firms and individuals: Availability of free data and information reduces the innovation-specific private or open investment required to make the next innovative advance. Despite the clear social welfare value of innovation commons under many conditions, academic innovation research and innovation policymaking have to date focused almost entirely on enhancing private incentives to innovate by enabling innovators to keep some types of innovation-related information at least temporarily *apart from* the commons, via intellectual property rights.

In this paper, our focus is squarely on innovation commons theory, evidence, and policy implications. We first discuss the varying nature of and contents of innovation commons extant today. We summarize what is known about their functioning, their scale, the value they provide to innovators and to general social welfare, and the mechanisms by which this is accomplished. Perhaps somewhat counterintuitively, and with the important exception of major digital platform firms, we find that many who develop innovation-related information at private cost have private economic *incentives* to contribute their information to innovation commons for free access by free riders.

We conclude with a discussion of the value of more general support for innovation commons, and how this could be provided by increased private and public investment in innovation commons "engineering", and by specific forms of innovation policymaking to increase social welfare via enhancement of innovation commons.

Social welfare gains from innovation commons:

Theory, evidence, and policy implications

1. Introduction and overview

The conception of knowledge as a common pool resource is not part of modern innovation policy. Instead, innovation policy is based on economics of industrial innovation that is centered on the externality in private investment in new knowledge (Nelson 1959, Arrow 1962, Romer 1990). As is well known, policy has addressed this externality via strengthened property rights in ideas (Gallini and Scotchmer 2002, Stiglitz 2014), plus public R&D support and innovation supporting institutions (Romano 1989, Nelson 1993, Jones and Williams 1998, Martin and Scott 2000, Davidson and Potts 2016, Edler and Fagerberg 2017, Bloom *et al.* 2019).[1]

The value and functioning of commons containing innovation-related data and information have long been overlooked in public policy analysis of shared economic resources in preference for market-based solutions. However, across a wide variety of circumstances it has been both theorized and documented that information and knowledge commons can be highly efficient economic institutions (Ostrom 1990, Raymond 1999, Benkler 2006, Torrance 2010, 2016, 2017, Frischmann *et al.* 2014, Gambardella and von Hippel 2019). Furthermore, the falling costs, diffusion and capability-enhancing affordances of new digital technologies are making the commons a steadily more valuable institution for supporting innovation (Allen and Potts 2016, Potts 2018, 2019). Via this paper we hope to begin the process of moving consideration of innovation commons into its proper, important place in the innovation policy field.

We define innovation commons as repositories of freely-accessible, “open source” innovation-related information. Contents range from aggregations of raw data to collections of finished innovation designs. As we will see, the data and information contents of innovation commons can be a very significant resource for innovating and innovation-adopting firms and individuals. The basic welfare-enhancing factor associated with innovation commons is that they reduce the innovation-specific private or open investment required from innovators in their quests to make the next innovative advance.

[1] For theoretical and empirical counterargument against this approach to innovation policy, see Boldrin and Levine (2008), Moser (2012), Kealey and Ricketts (2014).

The general argument we will make is that *if and as* private incentives to generate valuable innovation-related information can be sustained – which we will argue is often the case – social welfare will be enhanced when innovation-related data and information are transferred from private ownership to an innovation commons. As noted earlier and as is generally understood, this is in part because general access to valuable innovation-related data and information will lower innovation development costs. An important additional argument we will make, novel to our knowledge, is analogous to Hayek's (1945) observations regarding the importance of local information and market opportunities that only local purchasers and sellers can see. Local actions driven by local information, Hayek said, make markets better aggregators of overall demand and determiners of prices than central planners can be. Central planners simply cannot have access to the full demand-related information held by all potential purchasers and sellers of an item. The analogy to our case of the innovation commons is that a firm or other entity possessing data and information that it does *not* allow others to fully inspect, can only be aware of *some* potentially-valuable innovations related to it. Allowing or compelling free and full inspection by all can bring potential applications to light that: (1) are *not* visible to an exclusive information possessor, and (2) that *will* be visible to some others if they were allowed to engage in detailed inspection with the different lenses they bring to that task.

The remainder of this paper is organized as follows: In section 2 we define and describe innovation commons. In section 3 we examine social welfare benefits from open access. In section 4 we examine the motivations for voluntary contributions to the innovation commons. In section 5 we consider the incentives to privatize innovation related data and information. In section 6 we propose a set of policy approaches to support the innovation commons. Section 7 concludes by discussing how these proposals might be applied.

2. Defining and describing innovation commons

Recall that we define innovation commons as pools of innovation-related information that can be accessed by anyone for any purpose without payment and without limit. Our definition's restriction of commons' content to information sets us apart from the historical use of the term within economics, which has very generally referred to physical resources that are not privately owned. Examples are air, public waters, fish in the sea, and so forth. Physical resource commons are of course inherently rival (Hardin 1968). When "tragedy of the commons" conditions occur or are anticipated in the case of commons consisting of physical, rival resources like fish in the sea or fresh water, governance is commonly resorted to, implemented by governmental regulations and/or by grass-roots forms of regulation, monitoring, and enforcement devised by resource users themselves (Ostrom 1990). Unlike physical resources, of course, data and information such as innovation designs are non-rival goods (Jones and Tonetti 2020). For this reason, as we will later see, governance requirements for innovation commons center on incenting information holders to place information *into* the commons, rather than on the need to manage withdrawals from the commons.

There are many types of information commons containing innovation-related data and information already widespread in the world. To grasp the nature and extent of the phenomenon, it is useful to position information commons along a "formalization" axis. At one end we see freely-accessible information that is completely unorganized. At the other, we see highly curated collections of free, open-source data and information designed for convenient search.

To illustrate an entirely non-formalized innovation commons, consider data that exist in in nature independent of human action. For example, data regarding everything from the state of the weather to the behaviors of plants exist independent of human creation or curation. Those data are in an information commons by our definition, because anyone and everyone is free to access them and add them to their own private stores without permission from anyone: the data are a non-rival resource that no one "owns".

As a second broad category of informal commons, consider that humans and organisms more generally generate a great deal of data and information as they go about their daily lives. Much of this is ignored or "thrown away" by its creators, but nonetheless persists in the environment. For example, the fossilized footprints of early dinosaurs are still available for study today on public, freely-accessible terrain. In a similar fashion, millions of out-of-copyright photos and text files reside on the internet today, often in an entirely unorganized way and often entirely forgotten by their creators. Much of this data and information – indeed, vast troves – are freely accessible to and fully inspectable by any and all interested searchers via internet searches.

We call commons consisting of either type of data or information described above as "midden heap commons." Our analogy, of course, is to the midden heaps of physical domestic waste – seashells, broken tools and other discarded items – thrown away by human communities in prehistory, and sought today by archeologists as valuable inputs to their research.

As an example of value gained from mining human-generated information freely lying about in midden heaps commons, consider the work of the open-source, investigative community named Bellingcat.com (Higgins 2021). This volunteer group very successfully searches publicly available information ranging from casually-taken photos posted by individuals on a whim, to openly posted official documents, to reveal important political matters that state actors or others are trying their best to hide. For example, Bellingcat volunteers pieced together photos of antiaircraft missile units driving on roads. Many of these had been casually taken by citizens and posted on the web – e.g., "look at what just rolled through my village." Bellingcat project participants then combined that information with other publicly-posted materials like publicly-posted official rosters of individuals manning Russian antiaircraft missile units in the Ukraine. By finding, organizing and analyzing relevant free information lying about in midden heap commons, Bellingcat participants were ultimately able to convincingly document the identity of the specific individuals responsible for the downing of Malaysian Airlines Flight 17 in 2014.[2]

The value of information freely available in midden heap commons is steadily increasing as methods for rapidly searching vast amounts of information freely posted on the web are becoming steadily cheaper and more effective. For example, it is today possible to use AI methods to search the entire textual and image contents of the internet for specific content, and to do this at rapidly decreasing times and costs (von Hippel and Kaulartz 2021).

At the other end of the 'formalization' axis are highly organized commons, often devoted to the collection, curation, and free diffusion of very specific types of information. As an example, consider PubChem. Funded by the US National Institutes of Health, PubChem "..is the world's largest collection of freely accessible chemical information. Anyone can search this information commons by chemical name, molecular formula, structure, and other identifiers. Under each entry, they can find chemical and physical properties, biological activities, safety and toxicity information, patents, literature citations and more."[3] Millions query this free information commons each year.

Sources providing information to the PubChem commons include academic researchers, government agencies, chemical vendors, and journal publishers. Economic incentives motivate these individuals and firms to *not* keep these data private, but to rather contribute them "for free"

[2] https://www.youtube.com/watch?v=mozxTk3Brqw
[3] https://pubchem.ncbi.nlm.nih.gov/

to the PubChem commons. Motivations to contribute include requirements made by research funders upon the scientists they fund to publicly reveal their findings. Also, journals are increasingly requiring that findings and data generated by researchers must be placed in an appropriate commons as a condition of publication. Still further researchers themselves, to the extent that they are motivated by a wish to make their work useful to and easily discoverable by others, have strong personal and reputational incentives to invest the effort required to have their findings properly formulated for inclusion in a commons. As one academic researcher told us with respect to PubChem: "PubChem is where colleagues look for new findings. If your work is not on posted on that site, you might as well not do it."

Finally, of course, in this hybrid world there are hybrid institutions that have some but not all characteristics of a commons. A prominent example is patent systems. Via purposeful legislative design, complete descriptions of inventions are provided by inventors in exchange for temporary monopoly rights over the inventions claimed. As soon as a patent is issued, all that descriptive material enters the innovation commons. However, the rights to *use* the claimed invention as opposed to learning about it are restricted for a period of time. After expiration of a patent, both the rights to learn from and the rights to use the patented information become available to all as a part of an innovation commons.

3. Social welfare benefits from open access

As was mentioned at the start of this paper, the basic welfare-enhancing factor associated with innovation commons is that they reduce the innovation-specific private or open investment required from innovators in their quests to make the next innovative advance. To this basic argument we add another. Research into problem-solving strongly supports the conclusion that *many* searchers, each equipped with different resources, tools, and interests, can discover more economic opportunities within a data set than can a more restricted set of searchers. To understand why this is so, recall Hayek's observations regarding the importance of local information and opportunities that only local participants in a market can see or discover. This advantage, he argued, makes markets a better informed, higher-quality aggregator of information that collectively determines overall demand and pricing than central planners can be. The analogy to our case of the innovation commons is that when many potential innovators, each possessing unique local information regarding their own insights, resources, and interests, are allowed to inspect a given data base, they will collectively discover or create additional opportunities related to it or inspired by its contents than any individual information owner can do.

This logic echoes that offered by Eric Raymond (1999) regarding a phenomenon he observed in open source software development communities and dubbed "Linus's Law." In the case of traditional, firm-based proprietary software development projects, discovering and repairing subtle code errors or bugs was known to be very arduous and costly (Brooks 1979). However, in open source projects, Raymond observed that debugging was much quicker and cheaper. The reason was, he argued, that the code was opened to detailed inspection by a large community of software developers and users – something not done in traditional, software producer development processes. Under these conditions, Raymond said, "given a large enough beta tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone. Or, less formally, 'given enough eyeballs, all bugs are shallow.'" He explains: "More users find more bugs because adding more users adds more ways of stressing the program. . . . Each approaches the task of bug characterization with a slightly different perceptual set and analytical toolkit, a different angle on the problem. So adding more beta-testers . . . increases the probability that someone's toolkit will be matched to the problem in such a way that the bug is shallow to [easily understood and solveable by] *that person*." (1999, pp. 41–44).

Other researchers have also argued that opening of information to many, and combining information from many, will enable more valuable solutions than can a single solver in exclusive possession of that same data (Raymond 1999; Benkler 2006; Jeppesen and Lakhani 2010; Frey *et al.* 2011, Eghbal 2020). In service to this expected benefit, practical methods have been developed to invite the attention of many to a specific problem by "crowdsourcing" it (Howe 2006). Crowdsourcing is defined as "the act of outsourcing a task to a 'crowd,' rather than to a designated 'agent' ... in the form of an open call" (Afuah and Tucci 2012, 355; Howe 2006). The crowdsourced innovation or problem-solving "task" may range from very general ("Come fix software bugs with us") to very specific ("We need a solution to this specific problem"). Crowdsourcing calls provide a way to get contributions from many whose potentially valuable contribution to solving a problem is not known in advance to a project's sponsors. Individuals self-nominate based on their self-assessment of the potential value of their contributions.

Posting data and information to a freely-accessible commons is in a sense an inversion of normal crowdsourcing practice. Instead of bringing solutions to a seeker-stated problem, people come to an innovation commons to search for inputs to a solution to problems *they* have defined. And/or, they come to be stimulated to create a novel need-solution pair based upon the new information they encounter in the commons (von Hippel and von Krogh 2016).

4. Motivations to volunteer contributions to the innovation commons

The conflict between the social welfare case for placing information into a commons and incentives to generate that information at private cost will be immediately apparent to our readers. In addition, since innovation commons by definition do not require users to pay for access to or use of commons contents, there is also the question as to how to motivate information providers to expend additional private funds to curate and actively contribute their information for the benefit of free-riders.

Interestingly, it is often the case today that this long-assumed conflict does not exist: those who generate information at private cost also have private profit motivations for contributing it to a commons. In the remainder of this section, we explain three broadly-applicable situations where those that make contributions to the innovation commons at private expense receive private rewards from doing so – thus avoiding the market failure just described. In the case of each, under some commonly-encountered conditions, deployment of the strategy by private actors can provide innovators with *greater* economic returns than can traditional strategies centered upon gaining monopoly profits via keeping innovation-related innovation private. When this is so, strategy users are likely to enhance social welfare by opening "more" of their data and information to inspection and potential application by any and all – *and* doing so without inhibiting the generation or discovery of new knowledge by private actors deploying the strategies.

4.1 Open design of common inputs

Firms that purchase an input of proprietary design from a supplier understand that there is no competitive advantage to be gained from such a purchase – rivals can purchase the identical input. They also realize that they and their rivals share a common penalty in making the purchase – they all pay a premium above the marginal cost of production to obtain input units containing the supplier's protected, proprietary design. Accordingly and increasingly, purchasers elect to *collaboratively* develop a substitute, open source design for proprietary inputs. In the case of software and information goods, these can be directly downloaded and used by design contributors and free riders alike. In the case of physical goods, these may be produced by input manufacturers as a "white label good" at a price closer to marginal cost (Gambardella and von Hippel 2019). The decision to make their collaborative designs open to free riders instead of protecting them as proprietary club goods is driven both by a desire to avoid the costs that would be required to prevent free riding, and also by positive benefits associated with granting access to free-riders (Gambardella and von Hippel 2019, von Hippel and von Krogh 2003, Harhoff *et al*. 2003, Nagle 2018). In their theoretical model, Gambardella and von Hippel show that an open strategy equilibrium arises endogenously in the case of software and hardware inputs purchased by multiple firms.

An illustrative example is the Open Compute Project. This very successful and carefully designed open-source hardware design project was launched in 2011 by Facebook, to create open substitutes for proprietary computer server hardware designs it had previously purchased from suppliers. Rather than using the technology only internally, or creating a club good open only to firms actually contributing to a design, Facebook created an open-source collaboration to develop the designs. These were then freely revealed to all in an information commons. Hardware incorporating the free designs developed by Open Compute Project participants are available from a number of producers as white label non-proprietary products (Dignan 2015). The Open Compute website claims these customer-developed, open-source designs are 38% more energy efficient and 24% less expensive to run than proprietary hardware substitutes available for purchase (Open Compute Project 2018).

The likely general high importance of an "open inputs" strategy based upon single firm or collaborative development of input designs by downstream firms can readily be seen. After all, every firm resides in a supply chain, and buys inputs from upstream suppliers to incorporate into the products, processes, and services that they use or sell. Widespread adoption of the open inputs strategy we describe can therefore produce a general shift towards open design sharing without public policy interventions.

4.2 Contributions to the commons by "bottleneck-owning firms"

The second of the three strategies we will describe involves "bottleneck-owning" firms contributing designs they have developed at private cost to the innovation commons (Teece 1986, 2006). The strategy that these firms successfully execute involves: (1) establishing tight proprietary control – via intellectual property rights or other means – over a focal innovation or asset and/or over one or more complements essential to design purchasers – a "bottleneck" – and; (2) attempting to make all other complements needed by the customer either free or as low-profit as feasible while insuring a supply (Jacobides *et al*. 2006; Baldwin 2015). The strategic goal is to enable firms to capture a greater portion of the potentially available customer surplus with reduced investment. In addition, producers may strategically reveal designs into the commons to disrupt rents being obtained by rivals.

A well-known example of this strategy is IBM's pioneering decision to provide significant developer support to the open-source software operating system Linux. It's business-enhancing goal was to weaken the position of Microsoft, a supplier of a popular proprietary software operating system. It also hoped to gain profits from increased sales of the complementary computer hardware products it sold – and that Microsoft did not sell. Casadessus and Llanes (2011) and Alexy *et al.* (2018) have modeled this strategy, and distinguish conditions under which the supplier firm is best off keeping all its products proprietary, making them free, or adopting a "mixed-source" strategy in which some products are kept proprietary and others are made free. Selective openness can also reshape markets by weakening competitors, particularly in highly rivalrous environments (Jacobides *et al.* 2018, Teece 2018, Almudi *et al.* 2020, Giarratana and Mariani 2013, Harhoff 1996).

Further examples of this strategy are seen when platform firms like Google and Facebook solicit users to post User Generated Content on their sites for general diffusion. They capture rents from this freely-provided content via their bottleneck control of their proprietary platforms. Pursuant to a strategy to create and strengthen their control over bottlenecks, a firm may actually invest to protect the open status of contributed knowledge. For example, bottleneck holder firms sometimes actually purchase intellectual property rights that may arise over time to potentially threaten the free status of a complement, with the specific intention to also contribute those rights to the commons (Alexy *et al*. 2013, 2017).

4.3 Free citizen innovation

A third major source of voluntary contributions to innovation commons involves "free citizen innovations" developed within the household sector (von Hippel 2017). Household innovation was not anticipated in Schumpeterian, producer-centered theory. Nevertheless, ten nationally-representative surveys conducted to date have all documented that the household sector in fact *is* an important source of innovation investment and development. Across only the ten nations studied to date, it has been found that tens of millions of consumer-innovators spend tens of billions of dollars annually on developing new product innovations for themselves. Many of these prove to be of general value, and later become the underlying design source for commercially-produced and sold products. The same surveys also find that only a very small fraction of household sector innovators attempt to protect the innovations they develop via any form of intellectual property rights. Instead, householders very generally release their unprotected innovations into the innovation commons – sometimes to formalized commons and sometimes to midden heap commons. They also very generally pronounce themselves willing to have peers and producers free ride on their innovation investments and designs (von Hippel 2017).

How can individual consumers justify investing in the development of free innovations when no one pays them for either their labor or for their freely revealed innovation designs? The answer is that they are *self-rewarded.* When they personally use their own innovations, they are self-rewarded by benefits they derive from that use (von Hippel 1988, 2005, 2017). When they benefit from such things as the fun and learning of developing their innovations, or the good feelings that come from altruism, they are also self-rewarded (Raasch and von Hippel 2013). Due to its self-rewarding nature, free innovation does not require compensated transactions to reward innovating consumers for the time and money they invest to develop their innovations.

It has also been found that, about 90% of the time, citizen innovators have no incentive in investing to protect their innovations from free riders – because they themselves have no intention of commercializing them or gaining competitive advantage over others via exclusive use (de Jong *et al.* 2015). However, and in contrast to the two strategies described earlier, these innovators may have no incentive to invest their private resources to further the diffusion of knowledge of their innovations to potential free riders. For this reason, there may be a case for public investments to offset this specific type of market failure by supporting the diffusion of freely-revealed innovations developed by citizens. We will return to this matter later when we discuss public policymaking useful for supporting innovation commons.

4.4 Government generation of data and information for innovation commons

Historically, governments have invested in many types of basic research and development using public funds. These publicly-funded research activities are carried out in a mix of governmental, university, and corporate laboratories. Learnings developed are usually but not always freely revealed into the global information commons by academic publication and other means (Montgomery et al. 2020).

Government investments in research and innovation are commonly justified as necessary complements to the innovation efforts of private, profit-seeking firms. Sometimes the justification is that the investment required to pioneer a new field is simply too large for private firms to take on. For example, the collection and analysis of global weather data involved massive investments in space satellite systems and massive data analysis capabilities. Today, these data are very generally placed into information commons, and companies utilize those data to create profitable commercial applications (Coyle et al 2020). The GPS satellite system is another notable example, initially developed for military use, and today providing the data needed for a myriad of location-based applications. Another justification often used for government investment is that development of a new area of innovation "pre-commercial" research with a path to profitability that is simply too long to be economically justified by profit-seeking entities given their internal discount rates. For example, decades of basic research investment by governments into genetics and related fields has created an intellectual platform upon which private companies can today develop important and profitable new medical treatments (Nelson 1959).

5. Incentives to privatize innovation-related data and information

Taken together, the first three strategies described in section 4 that privately reward contributions to the innovation commons are applicable to an unknown but likely very significant fraction of all innovation designs. First, consider the wide applicability of the "open-source inputs" strategy we described. Every firm resides in a supply chain and buys inputs from upstream suppliers to incorporate into the products, processes, and services that they use or sell. If each level of a supply chain elects to enhance profits by substituting open-source designs for proprietary input designs, there is potential for a major general shift towards design openness by producer firms from this strategy alone. Second, the bottleneck strategies we described are widely applicable and widely applied today. Platforms like Facebook and Google that host and profit from freely-contributed user-generated content are prominent examples. Third and finally, household sector innovation has proven to be of substantial scale, and 90% of the designs consumers create are not protected, and so potentially discoverable by others via midden-heap searches at a minimum.

However, these strategies were not intentionally designed to be comprehensive. In this section, we consider circumstances under which data and information collectors or creators see a private benefit in keeping information they generate with private funds *out of* an innovation commons. As we will see, there is a significant likelihood that, in spite of the general trend towards openness described in the previous section, the value of privately-held data is increasing these days. In section 6 we consider how innovation policy can be improved to improve the balance between private data and information and the innovation commons.

5.1 Innovation-related information kept private via intellectual property rights

Innovation-related data and information kept private via intellectual property rights are those covered by patent or copyright grants or by trade secrecy law. When these mechanisms are used, information owners are making a conscious decision to protect their innovation-related information from inspection and/or use by others.

The fraction of valuable information held by private parties and protected by patent may be relatively small. The only portion protected is that which is formally claimed as an invention, and the duration of protection offered to innovators is for a period of 20 + / - years only. However, information protected by patents is by construction relatively recent, and so likely to be an unusually valuable fraction of extant information for follow-on innovators.

The fraction of information relevant to innovation that can be protected by copyright is likely restricted to software. Software has been deemed by US courts to be a form of "writings" and so coverable by copyright. However, this protection is not very strong given the intentions of follow-on innovators. As with all writings, only the 'expression' is protected by copyright not underlying ideas. Thus, in the case of software, only specific code is protected by copyright, not the possibly novel function that code provides. Others therefore are free to study copyrighted code's functions and are free to write a different version able to perform similarly if they wish to do this.

Trade secrecy is likely the most frequently-used way to protect innovation-related information. The quantity and value of trade secrets is unclear, because they are often unquantified and unrecorded. Trade secrets can be, for example, production methods held in the form of uncodified "knowhow" held in the minds of factory workers and passed on to new employees by observation of the behaviors of their more experienced colleagues.

Often, trade secrets have to do with very specific process practices of specific firms. For example, "This is the special and advantageous way we heat-treat the saw blades we manufacture and sell". Or "This is how we have learned to operate X specific process machine in our silicon foundry to achieve a higher yield than our rivals." When application-specific, keeping these data private as trade secrets is will reduce social welfare only within a narrow range. The only ones interested are likely to be direct rivals whose processes would also would benefit from that secret, application-specific information. However, when the value of that information is or could be made more general – "this is how our trade-secret method for motivating our salespeople could enhance the effectiveness of *any* salesperson" – then there is a larger social welfare loss from keeping that information secret via the affordances of trade secrecy law.

5.2 The increasing extent of privately-harvested, generally-useful "digital exhaust"

Digital technologies have dramatically lowered the cost of creating, and searching, sharing and analyzing innovation-relevant data (OECD 2019). The digitized data that economies produce as both a product of and as a near-costless by-product of ordinary activities (sometimes called digital exhaust) is growing at an extraordinary rate. This is due to economic activity becoming increasingly digital or interacting with digital infrastructure (e.g. payments, platforms). It is also a consequence of more economic goods producing digital data streams (e.g. phones, vehicles, houses, etc.) due to ever falling costs of hardware and ever improving technologies (including software) for processing, storing and sharing data (Brynjolfsson and McElheran 2017).

There are many different types of data: market and price data, transaction data, industrial data, environmental data, personal data, scientific data, machine generated data. The economic importance of this data is observable in: (1) Intangibles; (2) Prices: digital causes prices to fall (often to zero) for many economic costs – e.g. search, replication, transport, tracking, and verification; (3) Technology: in investments and advances in digital technology and innovation. (4) Economic organization: growth of digital firms and creation of new types of markets and jobs, and growth in platform-based business models. (5) Data: rise of data (and software) as an economic resource or factor of production.

Today, as a result of large-scale collection of many types of data by huge, internet platform firms like Google, Facebook, and Amazon, the likely welfare loss associated with keeping data sets of potential general value private has become much more salient. For example, Facebook's Instagram application, in the ordinary course of carrying out its primary business function of offering to serve as a repository for personal photographs, has private access to, at essentially no extra cost, the huge repository of images it stores. Similarly Google, as a result of offering a search service to millions of users, "owns" a huge data base of information regarding what is searched for, connectable many specific attributes of people engaging in those searches. These data can be used for many generally-useful purposes, only some of which will be understood by and potentially profitable to Facebook or Google themselves. Many university researchers and also firms – mostly non-rival with respect to the businesses of the firms controlling the data – could gain value from also having access to those data and information for their own purposes. Indeed, they are severely disadvantaged by not having it. For example, development and improvement of new general purpose digital technologies such as artificial intelligence require access to huge troves of training data (Cockburn *et al.* 2018).

When firms that collect such data in the course of their ordinary activities keep access to these data private, as they routinely do, a great deal of innovation opportunities can be lost, because "many eyes" are prevented from inspecting that data, as was discussed earlier. Again following the logic of Hayek, these same data bases could also have great value to others – who would predictably discover additional valuable opportunities by using their own lenses of local information – *if* the firms collecting this information placed it into an innovation commons.

Despite the social welfare advantages that inclusion of digital data in the commons could bring, there are clear reasons why private holders of major and generally valuable data collections would fiercely resist such transfers (Acemoglu *et al.* 2020). Indeed, business models of large platforms like Facebook and Google are premised on the advantages that exclusive access to private hoards of innovation-relevant data such as transaction-related data and preference-related data can bring to those firms. Offering selective access at a price to others via some sort of market run by those firms is not a viable alternative solution. Because data-owning firms have no way of knowing what opportunities others will find if given access, they have no way to appropriately price access (Arora and Gambardella 2010, Bergemann *et al.* 2018). In practice, this tends to result in a default solution – not desirable from the perspective of social welfare – of allowing very little access and only under tightly controlled conditions.

It is becoming increasingly recognized that innovation policy needs to adapt to address these constraints and resulting constrained access to new innovation new opportunities (Planes-Satorra and Paunov 2019).

6. Policy approaches to support the innovation commons

It has long been known that, since data and information are non-rival goods, once they have been created, social welfare is enhanced if all have access to them. And, of course, the countervailing argument is also well known. If all have free access to information that some have created at private cost and also can freely use it – can free ride – the private incentive to invest in generating new information will drop to zero, and social welfare will suffer from its absence.

In this section, we focus on the question: what should innovation policymaking do to draw increasing amounts of information into the commons without reducing incentives on the part of private parties to create or collect such information? This focus is relatively novel in the innovation policymaking literature. As we noted earlier, modern industrial innovation policy is not geared to support the commons, nor does the economic theory it is based on presume that critical or valuable innovation resources are generally found in the commons. Instead, stated broadly, modern innovation policy is designed to affect the strategic investment decisions of innovating firms in an industrial market context. It seeks to create high-powered incentives for private investment in innovation inputs and outputs (working through courts, regulatory agencies, and the tax system). It also seeks to secure demand for new products that embody new technologies and help create markets around the world, and to supply complementary public goods and innovation infrastructure and institutions (Foray *et al.* 2009, Edler and Fagerberg 2017). Present-day innovation policy, in other words, is designed to support private investment in new knowledge and capital, through property rights or through supporting subsidies and transfers, or public supply of innovation relevant inputs. But it not designed to support common pool resources.

In Arrow's (1962) model of innovation as production of an information good the properties of non-convexity, non-rivalry and uncertainty predict market failure (Shell 1966, Romer 1986). This is the basis in economic theory of modern industrial innovation policy. But today, as we argued earlier, it will be valuable to enhance innovation policy via inclusion of common pool resources (Ostrom 1990, Hess and Ostrom 2006). Society's innovation problem is not just how to best use known scientific and technical information, but also how to put to use the full resources of data, information and knowledge, including sticky, tacit knowledge resources disbursed among many separate individuals (von Hippel 1994, Lüthje *et al.* 2005, Potts 2019). Innovation commons are a general class of mechanism to pool and coordinate the use of distributed local information and data.

In what follows, we consider three complementary policy types that can increase the level of contributions to the innovation commons. In the cases of the first two, it is clear that the incentives for privately-funded data generation and/or collection are not affected. In the third approach, as we will discuss, this is not entirely clear.

6.1 Public support for innovation commons infrastructure

As we saw in section 4, firms and individuals that generate generally-valuable information at private cost often have private incentives to contribute that information to an innovation commons – and so will do this without added regulatory or legislative inducements. However, in many cases, those with private information they are willing to contribute "for free" to a commons are deterred from doing this by the high costs of interacting with commons that are very poorly designed. Accordingly, we propose that governments support both research related to developing well-designed commons, and invest in creating and operating such commons as well. In effect, we are suggesting that governments develop policies that view commons, including data commons, as valuable public infrastructure like roads or bridges (Frischmann 2013, Lane 2020, Coyle *et al.* 2020).

As an example of a very important type of information where private holders are fully willing to share – but are today frustrated by poor commons design – consider the discovery by medical clinicians of new, medically-valuable "off-label" uses for FDA-approved pharmaceuticals. The context for this example is that the FDA approves new drugs for specific "labeled" applications. However, medical clinicians are authorized to also use these drugs for any additional "off-label" purposes that they may themselves discover or hear about and think useful. This is a very valuable authorization because, after drugs have entered the market and are in clinical use, thousands of new off-label applications are created or serendipitously observed by clinicians annually. Research shows that many could offer widespread medical benefit if the information were widely diffused. Indeed, over half of the most valuable drug applications have their roots in field discoveries by practicing clinicians (von Hippel, Demonaco and de Jong 2017).

Today, however, very few valuable clinical case observations are publicly reported upon by clinician discoverers – even though clinicians state themselves willing to provide that information "for free". The reason is that the form of commons available to support diffusion is not designed to diffuse many case reports. Instead, specialized medical journals – the primary "legitimating" diffusion channel - are designed to accept and diffuse findings from carefully-designed medical studies, not off-label usage case reports. The chance of a case report actually being accepted and published if submitted is very small – and the personal time cost to a physician of even making the attempt is high: It takes many hours to draft and submit a paper to a journal for review. Further, *non*-academic clinicians – that is, most clinicians - who may have a generally-valuable off-label application to diffuse can expect little or no professional advantage even if they succeed in publishing an academic article.

Clearly, the existing commons mechanism in the case of off-label discoveries of new applications for FDA-approved drugs frustrates the general diffusion of this valuable form of innovative discovery. Research shows that non-academic clinicians are on average willing to devote ~ 15 minutes of their own time to contribute a case report to a commons (von Hippel, Demonaco and de Jong 2017). A well-engineered commons clearly could be developed to meet this input cost constraint, and could be operated at minimal expense and to great public benefit.

There are doubtless huge numbers of other types of data and information available for contribution to the commons that are similarly stranded in private hands today for similar reasons: holders of generally valuable information are willing to give away their valuable, innovation-related private knowledge for free, but do not want to incur significant private costs required to successfully diffuse it to a commons. As a major additional example, consider that there are tens of millions of household sector innovators spending tens of billions of dollars annually developing innovations for their own use. 90% of these report that they are willing to provide their innovation designs for free. But, as of today, there is no general commons structure in place to support general diffusion of household innovator commons (von Hippel 2017).

The cost of both design and operating support for even very important commons can be quite low, and is clearly a justifiable form of public expenditure in many instances. Recall again the example of PubChem. PubChem was designed and is being operated a public expense by NIH. The cost management and curation of this important innovation commons is quite low, given that it handles millions of queries per year. In 2005, the annual PubChem budget was $3 million, and its staff numbered only 13 individuals. This low cost is possible because the site is well designed: Contributors to the commons have private incentives to generate, format, and contribute chemical information to the commons. These contributors also have high incentives to make sure the information is of high quality, because their names and reputations are linked to their contributions.

Each situation involving free information stranded in place due to poor commons infrastructure could have a specific solution engineered to reduce costs of contributing that information to a data commons. Or, and preferably, research into commons structure and functioning could be devoted to creating general well-designed structures for innovation commons, along with open standards and general tools to enable the operation of commons at a low cost. We suggest that public policy should be put into place to support these efforts.

6.2: Government investment in creating generally valuable data

Recall from section 4.4 that, historically, governments have invested in many types of basic research and development using public funds. Learnings developed are usually but not always freely revealed into the global information commons by academic publication and other means (Montgomery et al. 2020). These government investments in research and innovation are commonly justified as necessary complements to the innovation efforts of private, profit-seeking firms. Sometimes the justification is that the investment required to pioneer a new field is simply too large for private firms to take on. In other cases the justification for government investment is that development of a new area of innovation "pre-commercial" research with a path to profitability that is simply too long to be economically justified by profit-seeking entities given their internal discount rates. For example, decades of investment by governments into basic research on genetics and related fields has created the intellectual platform upon which private companies can today develop important and profitable new medical treatments (Nelson 1959).

Today, our author team proposes that there is clear public justification for investing in additional forms of generally-valuable data such as behavioral data and other types of data *that are not necessarily in themselves innovative, but that nonetheless serve as an essential platform for research and application development by many*. Examples of such data are assemblages of transaction data associated with consumer and producer purchases, assemblages of specific search topics posed by individuals to Google and other search engines, and so forth. These data collections have been found to be very useful for research on such topics as user preferences, and that in turn can lead to commercially-valuable information and innovations enabling, for example, firms to better understand and serve those preferences.

Even without such policy support, user need for access to generally-valuable data and information collected by private firms but not shared by them is sometimes considered so important to innovative advance that researchers may gather together to create a public equivalent for themselves. For example, a Stanford professor, Fei-Fei Li, initiated the creation of a public data set of images called ImageNet[4] for use by any and all engaged in artificial intelligence research – or in any other form of research where this data set could be useful. The images in ImageNet were scraped from the web and then labeled by volunteers via crowdsourcing. Both the images and the labels are made freely available in a commons. Li was motivated to create ImageNet because it seemed very important to her and to many fellow researchers that all researchers have access to this important type of database, rather than just researchers working for firms like Google and Facebook.

As a second example, in Europe, a group called the Open Search Foundation has proposed a plan to create a common internet index that can underpin many European search engines. Stefan Voigt, the group's founder, explains why he thinks open search is important for public welfare: "The algorithms and huge data pools of the major search engines are now commercially-operated "black boxes". We can't look inside. … Just as road maps and land

[4] https://en.wikipedia.org/wiki/ImageNet

survey information are now public goods and accessible free of charge to all people, the internet must also be mapped openly."[5]

In net, we propose that policies be set into place to systematically determine when government investments in creating data and information for the commons is socially justifiable *beyond* the categories of scientific research traditionally focused upon. In effect, and by way of analogy to the concept of General Purpose Technologies (GPT) – we are proposing research to identify promising areas for government investment in generating General Purpose Information (GPI) and placing it into information commons. This idea has not been previously considered to our knowledge.

The importance of systematic analysis is clear. Today, absent research on this topic, whether a generally-useful database is developed publicly or privately has been largely a matter of happenstance. For example, both highly detailed world weather data from satellites and GPS capabilities were initially developed with large investments of public funds - both initially justified by their value to the US military. Later, the immense general and economic importance of freely-revealing both types of data was recognized. Had these same capabilities been developed first by private firms, it is likely that much general benefit would have been lost, as private owners would predictably have sought to parcel out access to their GPI data selectively in order optimize their private benefits.

Of course, sometimes data bases containing GPI data and information already exist as the private property of private firms. If the data is not shared by the private owner, governments may still elect to invest in a free alternative for the sake of public benefit. And, in such cases, a policy issue will predictably emerge: governments may be deemed by owners of those private data bases to be cases of "government competing with private business" – a practice generally frowned upon in economies centered on privately-funded business activities. Indeed, in the case of PubChem, a scholarly society made precisely this claim. The American Chemical Society was making a business of collecting chemical information published in its journals and charging researchers for access to those data. The public PubChem data base was much more extensive than that owned by the American Chemical Society, and NIH decided to continue with its free,

5 https://opensearchfoundation.org/en/open-search-foundation-home/

larger-scale data base service in the name of public welfare. The Society petitioned congress to shut down Pub Chem – but was ignored.[6]

6.3 Compel sharing of some data collected at private cost

Where government has paid for the generation or collection of valuable innovation-related data and information as in the cases described above, it has clear standing to insist that the findings from this research be published promptly, and that related data and methods be placed into a commons. When the data in question has been collected or generated using private funds, however, the matter is not so straightforward. Recall from section 5.2 the increasingly important case of privately-held digital data collected in massive amounts by internet platform firms like Google, Facebook, and Amazon. In that section, we argued that innovation would predictably be enhanced and innovation development costs lowered if that information were placed into innovation commons – given that this could be done without reducing the incentive of private firms to collect those data.

In the previous section, we noted the option that governments could elect to simply recreate GPI data at public expense and place it into innovation commons for free access by all. This approach is in our view likely to have clear data quality advantages. From the start, public data collection and curation could be designed and conducted with an eye towards general utility and general ease of use. In contrast, data collected by firms for their own purposes and then shared is likely to involve firm-specific data collection modes and data categorizations that are not optimal for many follow-on uses.

However, despite these likely advantages, governments may sometimes prefer policies that induce or compel holders of GPI data to share their holdings openly with the public – even when firms do not find it economically attractive to do so voluntarily. How this goal can be actually accomplished in practice without reducing the private incentive to collect such data by private firms must be determined via research and experiment. However, it may be useful to illustrate the possibilities of this general approach, and so we outline one hypothetical example next.

First, we should note that massive private digital data collections of general GPI interest are today held by only a very few major firms such as Facebook, Google, and Alibaba. As

[66] See https://osc.universityofcalifornia.edu/2005/05/american-chemical-society-calls-on-congress-to-shut-down-nihs-pubchem/

Hartmann and Henkel (2020) have documented in the case of AI research and development activities, these superior collections of privately-held data can give these few firms major innovation development advantages over both corporate rivals and over academic researchers as well. We propose that the hypothetical example we develop next be applied to these few major platform firms, and not to all private entities individually collecting small amounts of GPI data.[7]

Suppose next that these few private holders of major collections of digital GPI data were required to place into an innovation commons only the "raw" data they collect, e.g. order data, image data, search query data, etc.. Additionally, suppose that the affected firms were *not* required to share the uses they make of these data, nor the analyses they perform related to them. In such a case, all individuals and firms accessing the commons would have an opportunity to conduct their own analyses for their own research or business purposes on the raw data shared – but would not be able to build upon the analyses privately invested in by the data provider or other users of this commons. We think this requirement for data openness would not adversely affect most or all firms' willingness to collect the raw data, simply because collection and transaction-level use of those data are essential for the conduct of their own businesses. We also think it would not adversely affect the incentives of the collecting firm to invest in using those data in-house for further analyses and related innovation: those analyses, under our general proposal, need not be made public.

In effect, we are proposing that private incentives will be adequate to preserve firms willingness to invest in collecting data they presently collect, even if public sharing is made mandatory. However, if it turns out that in practice additional private incentive increases are needed, these could be designed in. For example, one might consider a policy allowing firms to delay posting of their raw data to the commons for a few days or weeks in order to have an initial, exclusive chance to explore and exploit their value. Indeed, giving firms a short period of exclusivity might give them a socially-beneficial incentive to move more quickly to find and exploit new innovative possibilities.

---

[7] The set of relevant firms can be defined using criteria equivalent or similar to those proposed in the Digital Markets Act of the European Commission. See European Commission (2020). The proposed legislation defines in its Article 3 "gatekeepers" as online platforms that exceed a number of size thresholds, e.g., have more than 45 million active monthly end users or more than 10,000 active yearly business users. See Cabral et al. (2021) for an assessment of the proposal from an antitrust economics viewpoint.

On the surface, this proposal to require private entities to contribute raw data they collect into the commons seems attractive with respect to leveling the playing field between firms presently having very different private data endowments. However, it is important to note that much work would be required to insure that such a measure would work well in practice. Consider that there will predictably there will be very difficult problems with respect to both protecting rights to privacy under this proposal, and also insuring that users of the commons can make sense of data that are contributed. With respect to the first point, consider that much of the value that can be gleaned from "raw" digital exhaust lies in uncovering and exploring patterns visible across data that is linked at the level of the individual: Why do people who buy books on X tend to be in A income brackets, in B political parties, and have an educational level of C? But enabling all users to link data at the level of individuals will almost certainly undo policies aimed at protecting individuals' privacy. Individual identities may be hidden with respect to specific data elements shared but, given enough linked data, data analysts can surely uncover specific identities. For example: "The only person on street X who has recently purchased solar panels and also ordered a how-to-do it book on panel installation is, with high likelihood, Joe Smith." (Note that these privacy problems already exist – our proposal does not introduce additional ones.)

Second, consider that it may be easy for private providers of raw data to an innovation commons to make that data unusable by or misleading to others. What is needed is the equivalent of the clarity and completeness requirements placed upon those filing for a patent. In order to have their claims upheld, inventors must describe their inventions so clearly that it can be successfully reduced to practice by anyone of ordinary skill in the art.[8]

For enforcement purposes, we next propose bringing raw data held monopolistically clearly within the realm of antitrust consideration. When a corporation is found guilty of violating antitrust law, several remedies are common. For especially egregious monopolistic behavior, criminal penalties may be assessed or civil penalties may be amplified by such provisions treble damages. Future antitrust violations may be prevented by courts granting injunctions against specific courses of action. Consent decrees mandating particular behavior may also be granted by courts. The special context of the innovation commons may suggest new antitrust remedies. For example, companies that gather and hoard large amounts of data for their

[8] *In re Wands*, 858 F.2d 731 (Fed. Cir. 1988).

own private purposes might avoid harsh legal penalties if they were to release at least some of the data they collect into the innovation commons. A unique safeguard on this remedy could be the fact that companies are not the only parties with an accurate record of data collected; users who participate in generating that data in the first place will, in principle, also know it.

Clearly, there is a need for research to develop and test policymaking with respect to why, when, and how to require open access to privately-collected data. However, it is also clear that these are early days. In this regard, it may be useful to note that there are many specific instances of data-revealing requirements already extant in modern economies, and some of these may also point the way to additional, more general approaches to the problem. For example, medical patients in the US today have legal rights to access their personal medical records in many jurisdictions – but not to "own" them. (Medical records consist of a mix of patient-supplied data and clinician-supplied data such as diagnostic opinions). Patients also have the right to freely share their personal medical data in a commons if they wish, where it is typically anonymized for personal privacy reasons.[9]

And of course, as was mentioned earlier, governments already have in place a mechanism that *does* mandate transfer of innovation-related information to the commons – after a lag. We refer of course to the intellectual property rights system. Both patent laws and copyright laws are intended to induce the eventual contribution of knowledge to the innovation commons by offering a temporary period of exclusivity for information protected by the grant of intellectual property rights. That is, as a result of purposeful legislative design, patented and copyrighted materials are first protected as privately-owned information via the granting of intellectual property rights, but then, after a prescribed lag, *automatically* become part of innovation commons. In this case, government has the standing to insist on contributions to the commons in these cases because both sides have willingly entered into a contract to do this.

Raw data commons will clearly be a valuable resource for innovation in any nation. The challenge lies in the specific institutional design to create sufficiently powerful incentives for private contributions and to address privacy concerns. There are a number of strategic pathways that modern innovation policy might take, and research and field experimentation could usefully explore and develop them.

[9] Health Insurance Portability and Accountability Act of 1996, 1996 Public Law 104-191 (United States).

7. Discussion

We have argued that great economic and social value can accrue if academics and government policymakers both study and support the transfer of innovation-related data and information to information commons. The digitization of much information, the sheer scale of the information collected, and the rapid advance of methods for analyzing it are strong trends that are collectively changing the structure of the innovation system towards using more pooled data, being more platform based, and making better use of distributed locally-known information (von Hippel 2017, Potts 2019). Innovation economics has not fully caught up with this in theory or policy. Instead, as we noted at the start of this paper, innovation economics and industrial innovation policy are today still designed primarily to address problems of market failure and knowledge externalities associated with private investment (Martin and Scott 2000, Fagerberg 2017, Bloom *et al* 2019). We hope that this paper has made it very clear that there is much need for more commons-related research and practice development to improve our understanding and policies and practices – and that this work will be very interesting and likely difficult as well.

7.1 The goals of innovation policy in a digital economy

Industrial innovation policy was and is about many things, but ultimately it sought to incentivize private investment in useful knowledge. It sought to provide public goods and market infrastructure to incentivize that investment. Innovation commons-supporting policy is also about many things, but as we have argued here, it is ultimately about the design, protection and support of open data institutions. The overarching goal should be to place information and data into innovation commons a new general purpose innovation resource. This goal is in stark contrast to the goal of a privatized and closed information and data flow in innovation policies designed to support Schumpeterian industrial dynamics or the modern resurgence of mission-oriented industrial/innovation policy.

In practice, this means designing mechanisms that enable data, knowledge and information generated in economic production to move quickly and cleanly into the commons, where it can be reused for further innovation. This flow of data from normal economic activity through to innovation and back again is becoming a major driver of the innovation engine of a modern economy. The purpose of innovation policy research in a digital economy is to design the institutional support mechanism for this cycle. The policy challenge is that of ensuring this new resource is effectively developed and widely used.

7.2 The need to make data privacy and data sharing compatible

Data privacy and data access to support innovation are two very important goals that are often in serious conflict (Acquisti *et al* 2016). In various jurisdictions, data privacy is given the standing of a fundamental right. Examples including the General Data Protection Regulation (GDPR) is a privacy regulation enacted May 2018, affecting anyone processing the data of EU residents. In the USA, the State of California enacted the Consumer Privacy Act (CPA) which puts limits on the use of personal information. While effective protection of personal data is one of the cornerstones of a digital society, regulatory policy faces the major challenge of, on the one hand, effectively protecting people's legitimate interests in private information, and, on the other hand, designing data protection law in such a way that it promotes innovation in the digital world and does not impede welfare-enhancing aspects of a data strategy.

There is a plethora of technical means that may help to alleviate privacy problems such as anonymization, pseudonymization, and aggregation. However, as was noted earlier, much of the value that can be gleaned from "raw" digital exhaust lies in uncovering and exploring patterns visible across data that is linked at the level of the individual. But enabling data linking at the level of individuals will almost certainly undo policies aimed at protecting individuals' privacy. Individual identities may be hidden with respect to specific data elements shared but, given enough linked data, data analysts can surely uncover specific identities.

Attempts to protect individual privacy despite this reality may in practice result in the design of innovation "commons" that are in practice unusable or nearly so. Regulations imposed on research use of US Census data illustrates this problem. First, one must *apply* for access to detailed census data rather than simply having free access. "The Census Bureau considers proposals from qualified researchers in social science disciplines consistent with the subject matter of the surveys and censuses collected by the Census Bureau. Proposals may be submitted at any time. Proposed projects must: Provide benefits to Census Bureau programs; Demonstrate scientific merit; Require non–public data; Be feasible given the data; Pose no risk of disclosure. Further, researcher access to the data can only be obtained during physical visits to restricted rooms – and data cannot be removed from those rooms."[10] All in all, daunting requirements for modern research methods in our view, and as a result a very valuable data set is likely seriously underused.

The OECD (2015, 2019), in proposing an industrially targeted digital economy policy agenda, emphasizes 'data access' as a prime concern for governments, along with support and development of data markets (Planes-Satorra and Paunov 2019). Yet the OECD's policy models, such as strengthened intellectual property to support data markets, increases the costs of distributed innovation, and its recommended increased privacy regulations favor a permissioned access regime rather than unstructured exploration with open access. Renegotiations with digital platforms for greater tax revenue in return for social license to data might benefit government's revenue, but society could be better off with a different deal, one that encourages contribution of collected data and information into the commons. In other words, and in our view, we need to redesign modern innovation policy to adapt it to a digital economy in such a way that opportunities and resources for innovation are both maximally developed and also that these resources are available to the greatest number.

One policy pathway that has been proposed is data trusts, (Coyle et al 2020). These are in effect a consumer side cooperative or union to pool data that is then governed by trustees with a fiduciary responsibility to those who generated the data, but also acting under regulatory guidelines to use data for the public good. Trustees can negotiate on behalf of data providers to sell data into data markets, but they also have the option of choosing to place data into data commons where it is judged safe and worthwhile to do so. Data trusts may thus be a desirable hybrid of raw data markets and commons. Should public policy decide in favor of data trust solutions, then our proposal is to make extensive use of the commons option in order to reduce cost of access for innovators as much as possible.

In net, effective protection of personal data is one of the cornerstones of a digital society. However, regulatory policy faces the major challenge of, on the one hand, effectively protecting people's legitimate interests in private information, and, on the other hand, designing data protection law in such a way that it promotes innovation in the digital world and does not impede welfare-enhancing aspects of a data strategy. How to best make progress on both of these two goals simultaneously is not yet clear (see also Federal Government 2021: 16 f., OECD 2018: 28)

[10] See https://www.census.gov/programs-surveys/ces/data/restricted-use-data/apply-for-access.html

7.3 Conclusion

The path from an industrial to a digital economy is changing the innovation system in many ways. Platform business models, digital networks, and distributed toolkits are increasingly dominating economic production (Gawer 2009, 2014). Technologies such as search engines and deep learning algorithms are becoming steadily more deeply embedded in businesses (Cockburn *et al.* 2018, Goldfarb and Tucker 2019). Because of these shifts, information ranging from innovation designs to raw data are becoming an ever more important economic resource for both production and for innovation (Jones and Tonetti 2020, Veldkamp and Chung 2019, Farboodi and Veldkamp 2021). In consequence, innovation policy premised on industrial models of production, which ignore data as a resource in general and its institutional governance in particular, needs to be updated (Guellec 2018, Nambisan *et al* 2019).

Data and information are non-standard economic goods. The economic properties of data (non-rivalry, zero or low marginal cost, increasing returns, costly excludability, joint production) suggest the relative efficiency of common property regimes, i.e. data pools, as an efficient institutional structure to organize this economic resource. We predict that the importance of open data as a resource for innovation will continue to grow, driven by rapid expansions in the amount of data generated (consider as one example the increasing impact of the Internet of Things) and also the quite spectacular and rapid improvements in the capability of digital search and data processing tools.

Innovation policy also has strategic international considerations (Potts 2016), and these also extend to data commons. Because data pools have increasing returns to scale, pooling data resources increases value in the same way that joining local markets into a larger common market facilitates economic gains from trade and specialization. Reciprocal data sharing into a common pool can be usefully added to trade negotiations or grandfathered into existing agreements. And just as individuals can 'pay with data' for platform services (e.g. email, social networking, search), nations can provide 'aid' support with data by providing access to data commons, or by supporting infrastructure and toolkits to facilitate digital production and innovation.

All in all, we think that there is an strong case for further research and development on the utility and functioning of innovation commons, and an urgent need for policymaking to enhance support for both the design and operation of innovation commons.

References

Acemoglu D, Makhdoumi A, Malekian A, Ozdaglar A (2020) Too much data: Prices and inefficiencies in data markets. *Am. Econ. J.: Microeconomics* (forthcoming)

Acquisti A, Taylor C, Wagman L (2016) The economics of privacy. *J. Econ. Lit.* 54(2):442–92.

Afuah A, Tucci CL (2012) Crowdsourcing as a solution to distant search. *Acad. Man. Rev.* 37(3):355–375.

Alexy O, West J, Klapper H, Reitzig M (2018) Surrendering control to gain advantage: Reconciling openness and the resource-based view of the firm. *Strat. Man. J.* 39(6):1704–1727.

Alexy O, Henkel J, Wallin MW (2013) From closed to open: Job role changes, individual predispositions, and the adoption of commercial open source software development. *Res. Pol.*, 42(8):1325–1340.

Allen D, Potts J (2016) How innovation commons contribute to discovering and developing new technologies. *International Journal of the Commons* 10(2).

Almudi I, Fatas-Villafranca F, Fernández-Márquez CM, Potts J, Vazquez FJ (2020) Absorptive capacity in a two-sector neo-Schumpeterian model: a new role for innovation policy. *Ind. & Corp. Change* 29(2):507–531.

Arora A, Gambardella A (2010) Markets for technology. In B Hall and N. Rosenberg (eds) *Handbook of Economics of Innovation* (Elsevier, Amsterdam)

Arrow K (1962) Economic welfare and the allocation of resources for innovation. In R. Nelson (ed.) *The Rate and Direction of Inventive Activity* (Princeton University Press: Princeton)

Baldwin C (2018) Bottlenecks, modules, and dynamic architectural capabilities. in D. Teece (ed). *Oxford Handbook of Dynamic Capabilities*. (Oxford University Press, Oxford)

Baldwin C, von Hippel E (2011) Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organ. Sci.* 22(6):1399–1417.

Benkler Y (2006) *The Wealth of Networks* (Yale University Press: New Haven)

Bergemann D, Bonatti A, Smolin A (2018) The design and price of information. *American Economic Review* 108(1):1–48,

Bloom N, Van Reenen J, Williams H (2019) A toolkit of policies to promote innovation. *J. Econ. Pers.* 33(3):163–184.

Boldrin M, Levine D (2008) Perfectly competitive innovation. Journal of Monetary Economics 55(3):435–453.

Brynjolfsson E, McElheran K (2017) The rapid adoption of data-driven decision making. *Am. Econ. Rev.*, 106(5):133–139.

Brooks, F. (1979) *The Mythical Man-Month* (Addison-Wesley, New York).

Cabral, L, Haucap, J, Parker, G, Petropoulos, G, Valletti, T and Van Alstyne, M (2021). The EU Digital Markets Act: A Report from a Panel of Economic Experts. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3783436

Casadesus-Masanell R, Llanes G (2011) Mixed source. *Management Science* 57(7):1212–1230.

Cockburn I, Henderson R, Stern S (2018) The impact of artificial intelligence on innovation. NBER working paper 24449.

Coyle D, Diepeveen, Wdowin J, Tennison J, Kay L (2020) *The Value of Data.* (Bennett Institute for Public Policy, Cambridge)

Davidson S, Potts J (2016) A New Institutional Approach to innovation policy. *Australian Economic Review* 49(2):200–7.

de Jong J, von Hippel E, Gault F, Kuusisto J, Raasch C (2015) Market failure in the diffusion of consumer-developed innovations: Patterns in Finland. *Res. Pol.* 44(10):1856–1865.

Dignan L (2015) Do server vendors have an answer for the white box trend? ZDNet (March 4), http://www.zdnet.com/article/do-server-vendors-have-an-answer-for-the-white-box-trend/.

Edler J, Fagerberg J (2017) Innovation policy: what, why, and how. *Oxford Review of Economic Policy*, 33(1):2–23.

Eghbal N (2020) *Working in Public: The Making and Maintenance of Open Source Software*. (Stripe Press, San Francisco)

European Commission (2020). Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act). https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en

Farboodi M, Veldkamp L (2021) A growth model of the data economy. NBER working paper 28427.

Fagerberg J (2017) Innovation policy: Rationales, lessons and challenges *J. Econ. Surveys* 31(2):497–512.

Foray D, David P, Hall B (2009) Smart Specialization. The concept. *Knowledge Economists Policy Brief* #9. http://ec.europa.eu/invest-in-research/pdf/download_en/kfg_policy_brief_no9.pdf?11111

Frey K, Lüthje C, Haag S (2011) Whom should firms attract to open innovation platforms? The role of knowledge diversity and motivation. *Long Range Planning* 44(5-6):397–420.

Frischmann B (2013) *Infrastructure: The Social Value of Shared Resources*. (Oxford University Press, New York)

Frischmann B, Madison M, Strandburg K (eds) (2014) *Governing Knowledge Commons* (Oxford University Press: New York)

Gallini N, Scotchmer S (2002) Intellectual property: When is it the best incentive system? *Innovation Policy and the Economy* 2:51–77.

Gambardella A, von Hippel E (2019) Open sourcing as a profit-maximizing strategy for downstream firms. *Strategy Science* 4(1):41–57.

Gans J, Stern S (2010) Is there a market for ideas? *Ind. & Corp. Change* 19(3):805–837.

Gawer A (2009) *Platforms, Markets and Innovation* (Edward Elgar, Cheltenham)

Gawer A (2014) Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy* 43(7):1239–1249.

Giarratana M, Mariani M (2013) The relationship between knowledge sourcing and fear of imitation. *Strategic Management J.* 35(8):1144–1163

Goldfarb A, Tucker C (2019) Digital economics. *J. Econ. Lit.* 57(1):3–43.

Guellec P (2018) Innovation policies in a digital age. *OECD Science, Technology and Policy Papers* #59.

Hardin G (1968) The tragedy of the commons. *Science* 162(3859):1243–1248.

Harhoff D (1996) Strategic spillovers and incentives for research and development. *Management Science*, 42(6):907–925.

Harhoff D, Henkel J, von Hippel E (2003) Profiting from voluntary information spillovers: How users benefit from free revealing their innovations, *Research Policy*, 32:1753–1769.

Hartmann, P and Henkel, J (2020). The rise of corporate science in AI: Data as a strategic resource. *Academy of Management Discoveries,* Vol. 6, No. 3, 359–381. Online only https://doi.org/10.5465/amd.2019.0043

Hayek F (1945) The use of knowledge in society. *American Economic Review* 35(4):519–530.

Hess C, Ostrom E (eds) (2006) *Understanding Knowledge as a Commons* (Edward Elgar: Cheltenham)

Higgins E (2021) *We Are Bellingcat.* (Bloomsbury, London)

Howe, J. 2006. The rise of crowdsourcing. *Wired*, June 1 14(6):1–4. Accessed January 29, 2016. http://www.wired.com/2006/06/crowds/

Jacobides M, Cennamo C, Gawer A (2018) Toward a theory of ecosystems. *Strategic Management Journal* 39(8):2255–2276.

Jacobides M, Knudsen T, Augier M (2006) Benefiting from innovation: Value creation, value appropriation and the role of industry architectures. *Research Policy* 35:1200–1221.

Jeppesen LB, Lakhani KR (2010) Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* 21(5):1016–1033.

Jones C, Tonetti C (2020) Nonrivalry and the economics of data. *Am. Econ. Rev*. 110(9):2819–2858.

Jones C, Williams J (1998) Measuring the social return to R&D. *Quarterly Journal of Economics* 113(4):1119–1135.

Kealey T, Ricketts M (2014) Modelling science as a contribution good. *Research Policy* 43(6):1014–1024.

Lane J (2020) *Democratizing our Data* (MIT Press: Cambridge, MA)

Lüthje C, Herstatt C, von Hippel E (2005) User-innovators and 'local' information: The case of mountain biking. *Res. Pol.* 34(6):951–965.

Martin S, Scott J (2000) The nature of innovation market failure and the design of public support for private innovation. *Research Policy* 29(4-5):437–447.

Montgomery L, Hartley J, Neylon C, Gillies M, Gray E, Herman-Pillath C, Chun-Kai H, Leach J, Potts J, Ren X, Skinner K, Sugimoto C, Wilson K (2020) *Open Knowledge Institutions* (MIT Press. Cambridge, MA)

Moser P (2012) Innovation without patents: Evidence from World's fairs. *J. Law & Econ.* 55(1):43–74.

N.D. Cal. 2017 hiQ Labs, Inc. v. LinkedIn Corp. (273 F. Supp. 3d 1099, 1103

Accessed August 27, 2021 https://casetext.com/case/hiq-labs-inc-v-linkedin-corp-1

Nagle F (2018) Learning by contributing: Gaining competitive advantage through contribution to public goods. *Organ. Sci.* 29(4):569–587.

Nambisan S, Wright M, Feldman M (2019) The digital transformation of innovation and entrepreneurship. *Res. Pol.* 48(8):1037–1073.

Nelson R (1959) The simple economics of basic scientific research. *J. Pol. Econ.* 67(3):297–306.

Nelson R (ed.) (1993) *National Innovation Systems: A Comparative Study*. (Oxford University Press, Oxford)

OECD (2015) *Data Driven Innovation* (OECD Publishing: Paris)

OECD/Eurostat (2018), *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using*

Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities, OECD Publishing, Paris/Eurostat, Luxembourg.

OECD (2019) *Digital Innovation: Seizing policy opportunities* (OECD Publishing, Paris)

Ostrom E (1990) *Governing the Commons*. (Cambridge University Press, Cambridge)

Planes-Satorra S, Paunov C (2019) The digital innovation policy landscape in 2019. OECD Science, Technology and Industry Policy Papers.

Potts J (2016) Innovation policy in a global economy. *J. Entrep. & Pub. Pol.* 5(3):308–324.

Potts J (2018) Governing the innovation commons. *J. Instit. Econ.* 14(6):1025–1047.

Potts J (2019) *Innovation Commons: The Origin of Economic Growth* (Oxford University Press, Oxford)

Raasch C, von Hippel E (2013) Innovation process benefits: The journey as reward. *MIT Sloan Management Review*, September 2013, pp.33–39.

Raymond E (1999) *The Cathedral and the Bazaar* (O'Reilly Media, Cambridge, MA).

Romano R (1989) Aspects of R&D subsidization. *Q. J. Econ.* 104(4):863–873.

Romer P (1986) Increasing returns and long run growth. *J. Pol. Econ.* 94:1003–1026.

Romer P (1990) Endogenous technological change. *J. Pol. Econ.* 98:S78–S102.

Shell K (1966) Toward a theory of inventive activity and capital accumulation. *American Economic Review* 56(1/2):62–68.

Stiglitz J (2014) Intellectual property rights, the pool of knowledge, and innovation. NBER working paper 20014.

Teece D (1986) Profiting from technological innovation: implications for integration, collaboration, licensing, and public policy. *Res. Pol.* 15:285–305.

Teece D (2006) Reflections on profiting from innovation. *Res. Pol.* 35(8):1131–1146.

Teece D (2018) Profiting from innovation in the digital economy: Enabling technologies, standards, and licensing models in the wireless world. *Res. Pol.* 47(8):1367–1387.

Torrance, AW (2010) Synthesizing Law for Synthetic Biology. *Minnesota Journal of Law, Science & Technology*, 11(2):629-65.

Torrance, AW (2016) A Natural Experiment on Innovation Without Patents In D Harhoff and K Lakhani (eds) *Revolutionizing Innovation: Users, Communities, and Open Innovation* (MIT Press: Cambridge, MA).

Torrance, AW (2017) Better to Give than to Receive: An Uncommon Commons in Synthetic Biology In KJ Strandburg, BM Frischmann, and MJ. Madison (eds) *Governing Medical Knowledge Commons* (Cambridge University Press: Cambridge, UK).

Veldkamp L, Chung C (2019) Data and the aggregate economy. Working paper presented at Annual Meeting Plenary, Society for Economic Dynamics (forthcoming *J. Econ. Lit*.)

Voigt S (https://opensearchfoundation.org/en/open-search-foundation-home/).

von Hippel E (1988) *The Sources of Innovation*. (Oxford University Press, New York)

von Hippel E (1994) Sticky information and the locus of problem solving: Implications for innovation. *Manag. Sci.* 40(4):429–439.

von Hippel E (2005) *Democratizing Innovation: The Evolving Phenomenon of User Innovation* (MIT Press, Cambridge, MA).

von Hippel E (2017) *Free Innovation* (MIT Press, Cambridge, MA).

von Hippel E, Kaulartz S (2021) Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web. *Res. Pol.* 50(8):104–156.

von Hippel E, von Krogh G (2003) Open source software and the "private-collective" innovation model: Issues for organization science. *Organ. Sci.* 14(2):208–223.

Von Hippel E., von Krogh G (2016) Crossroads—Identifying viable "need–solution pairs": Problem solving without problem formulation. *Organ. Sci.* 27(1):207–221.

von Hippel E, Demonaco H, de Jong J (2017) Market failure in the diffusion of clinician-developed innovations: The case of off-label drug discoveries. *Science and Public Policy* 44(1):121–131. doi: 10.1093/scipol/scw042