

RESEARCH ARTICLE

Sustaining Scholarly Infrastructures through Collective Action: The Lessons that Olson can Teach us

Cameron Neylon

Centre for Culture and Technology, Curtin University, AU

cn@cameronneylon.net

The infrastructures that underpin scholarship and research, including repositories, curation systems, aggregators, indexes and standards, are public goods. Finding sustainability models to support them is a challenge due to free-loading, where someone who does not contribute to the support of an infrastructure nonetheless gains the benefit of it. The work of Mancur Olson (1965) suggests that there are only three ways to address this for large groups: compelling all potential users, often through some form of taxation, to support the infrastructure; providing non-collective (club) goods to contributors that are created as a side-effect of providing the collective good; or implementing mechanisms that lower the effective number of participants in the negotiation (oligopoly).

In this paper, I use Olson's framework to analyse existing scholarly infrastructures and proposals for the sustainability of new infrastructures. This approach provides some important insights. First, it illustrates that the problems of sustainability are not merely ones of finance but of political economy, which means that focusing purely on financial sustainability in the absence of considering governance principles and community is the wrong approach. The second key insight this approach yields is that the size of the community supported by an infrastructure is a critical parameter. Sustainability models will need to change over the life cycle of an infrastructure with the growth (or decline) of the community. In both cases, identifying patterns for success and creating templates for governance and sustainability could be of significant value. Overall, this analysis demonstrates a need to consider how communities, platforms, and finances interact and suggests that a political economic analysis has real value.

Keywords: infrastructure; financing; governance; collective action; scholarship

Introduction

The scholarly community depends on a large set of shared infrastructures. Some of these infrastructures (specific data repositories such as the Protein Data Bank or shared information resources such as Crossref) have clear institutional forms. At the other extreme are concepts as diffuse as 'the scholarly publishing system'. Elinor Ostrom would include these as 'institutions' under the definition of 'the prescriptions that humans use to organize all forms of structured interactions...' (Ostrom 2005, 3). They are also 'infrastructures' in terms of the set of characteristics described by Star and Ruhleder (1996)—embeddedness, transparency, reach beyond a single site, and 'taken-for-grantedness'—although they may not fit the dictionary definition used by Bilder et al: 'the basic physical and organisational structures and facilities (e.g. buildings, roads, power supplies) needed for the operation of a society or enterprise' (2015). In practice, as Star and Ruhleder note, the term *infrastructure* is difficult to define adequately, a problem that also contributes to the challenge of securing funding. The shared nature of these resources means that their funding is a collective action problem, and collective action problems are difficult to solve.

This article does not, therefore, seek to definitively solve the problem of defining infrastructure. Rather, its focus is on how we can sustain shared platform systems that support scholarly communities through the collection, storage, and transmission of shared resources. These resources are primarily archives, databases,

and information systems that service more than one research community. Such resources can include physical materials, but the primary focus of this article is on digital information infrastructures. These infrastructures frequently have formal institutional structures: they are often incorporated not-for-profit organisations (Crossref, ORCID, Cambridge Crystallographic Data Centre), but also include projects that are not formal organisations but have a definite identity (Europe Pubmed Central, Humanities Commons), or units within larger organisations (Genbank).

The broad question addressed in this paper is how infrastructures of this type can be sustained. More specifically, what lessons from classical economics and political economy can help guide us to sets of models or patterns that will work? Sometimes, it seems that discussions of infrastructure sustainability only occur when there is a crisis. How can we use our existing knowledge of political economics to ensure that one of Star and Ruhleder's defining characteristics of infrastructures—that they only 'become ... visible upon breakdown' (1996)—is avoided?

The General Provisioning Problem for Collective Goods and Its Relation to Scholarly Infrastructures

A fundamental political and economic challenge for groups in general, including scholarly communities, is the provision of 'collective goods' or 'general utility.' These terms refer to goods that are non-rivalrous – they can be infinitely shared – and non-excludable – it is difficult, if not impossible, to stop someone from using them. Classical economics tells us that there is a provisioning problem for such goods: rational individual actors will never contribute because, whether they do or not, they can still benefit.

Scholarly infrastructures, such as repositories that provide access to data, articles, and code, are close to the ideal of public goods. Resources are necessary to build infrastructures, to create the objects that they house, and to maintain them over time. The public-good nature of an infrastructure lies in the combination of its content, access to that content, and the metadata that supports its discovery and use.

In *The Logic of Collective Action*, Mancur Olson (1965) developed a model outlining how such collective goods can be provided, the circumstances under which they will not be provided, and possible solutions to the provision dilemma. In particular, the book discusses how group size has a profound influence on the provision of public goods, noting that optimal provision is possible in two cases: one, for sufficiently small groups, and two, for larger groups, as a by-product of providing non-collective goods to contributors to collective goods. Olson describes three different size ranges of groups: those that are small enough to reach an agreement to provide the collective benefit; those that are too large to do so; and those that lie in the middle, where the collective good may be provided but at a level which is below optimal.

Olson's description of the groups that can and cannot provide collective goods maps closely onto scholarly infrastructures. Small disciplinary communities frequently develop local infrastructures out of existing resources. Large-scale infrastructures are also provided through the collaboration of small communities. However these are small groups made up of large organisations. These might be groups of funders (such as those that fund Europe Pubmed Central) or of national governments (as is the case for physical infrastructures that are generally formed as inter-governmental organisations, such as the European Organization for Nuclear Research [CERN]). The transition from small to large group size is challenging and 'medium'-sized infrastructures often struggle to survive by moving from grant to grant, and, in some cases, shifting to a subscription model.

In the case of digital infrastructures, a collective good (such as an online article or dataset) can be made excludable (i.e. converted into a 'club good') by placing an authentication barrier around it to restrict access to subscribers (as is the case for online subscription journals and databases). Buchanan (1965) probed how club goods and club size relate. His core finding was that sustainable clubs reach an equilibrium where their size depends on congestion in access to the collective goods produced (i.e. the extent to which it is non-rivalrous) and the value those goods provide. With digital resources, congestion is low – although not zero – and the club can therefore grow large. This, in turn, creates a challenge. Digital resources are not natively excludable; a technical barrier has to be put in place. As the group size rises, the likelihood of 'leakage' – referred to as 'sharing' or 'piracy' (Lawson 2017), depending on your perspective – increases. Thus, resources are expended on strengthening excludability, which leads to both economic and political costs—as seen, for example, in the open access debate.

Solutions to the Provisioning Problem

If our political goal is to provide large-scale access to resources – to make goods more public-like through the provision of shared infrastructures – then we need to develop a political economics of this 'public-making.' Olson describes three routes to creating sustainable infrastructures that provide collective goods:

Compulsion: The collective good is provided through a mechanism that requires contributions from the whole community. Closed union shops, where all workers in a given company are required to be members of a union, are an example that Olson discusses in detail (1965: 66). Taxation is another example. Examples in the scholarly infrastructure space include overhead and indirect costs taken by institutions and top-slicing of funder budgets to provide infrastructures and services.

By-product: The collective good is provided through a mechanism that provides additional club-like and/or private goods to contributors. Olson (1965, 153) discusses the example of insurance schemes only available to members of mutual benefit societies (club goods) that also lobby on behalf of a broader community of interest that includes non-members (a collective good). In the research enterprise, publishers have to join Crossref to be able to assign Crossref DOIs to outputs (a club good). As a by-product, the whole community has access to an interoperable and (largely) open metadata set with defined schemas and access points (a collective good).

Reducing effective group size through oligopoly: Oligopolies exist where a small number of players (but more than two) control a space. The example Olson gives is of trade associations formed for lobbying. These might have many members but the majority of funds was generally provided by a small number of companies (1965: 146). There are far too many research funders globally for them to (easily) agree upon a mechanism for contributing to any shared scheme. However, because a relatively small set of funders fund a substantial proportion of biomedical research, they have in some cases been able to agree to fund data infrastructures such as Europe Pubmed Central (Europe PMC). Other funders may contribute, but, as Olson predicts, many smaller ones will free-ride. Compulsion and oligopoly can be difficult to distinguish because the identification can depend on perspective. Europe PMC is a good example of this overlap. The oligopoly mechanism is important for funders to solve their collective action problem. However, from the perspective of individual researchers, the decision of those funders to directly allocate funds from the general research pool to Europe PMC is a form of taxation.

The difficult truth that Olson articulates is that there is no mechanism that will lead directly to a large community supporting the provision of a large-scale public-good infrastructure. Any successful sustainability model will depend on some mixture of these three approaches for resourcing. There are interesting models that quantitatively address these collective action problems. These include Assurance Contracts, discussed by Crow (2013) as a means of addressing the size issues. Assurance Contracts operate in a manner similar to many crowd-funding approaches, where a project only proceeds if sufficient contributions are made. These contracts can be seen as ways for a community to implement compulsion on itself. Olson notes, for instance, that it was not uncommon for the majority of a group of workers to vote, even in private ballots, to form a closed union shop (1965: 85). This parallels an Assurance Contract.

If our challenge in delivering on the openness and transparency agenda is how to support the conversion of successful medium-scale club-like infrastructures into open systems that provide collective goods, then we need to solve the political and economic problems of transitioning from the club state to a model that successfully provides a mix of these models.

Case Studies of Scholarly Infrastructures

In this section, I examine a series of examples of scholarly infrastructures to give an outline of how Olson's model can be applied for the research community. These case studies draw in part on statements made by representatives of data infrastructures for a session at the 2016 SciDataCon meeting, which was part of the work of the Organisation for Economic Co-operation and Development (OECD) Expert Group on Data Infrastructure Sustainability, as well as some additional examples drawn from the community. The examples have been selected to show some historical parallels, as well as variations in the models developed and attributes of the collective goods being provisioned.

Cambridge Crystallographic Data Centre

The Cambridge Crystallographic Data Centre (CCDC, <https://www.ccdc.cam.ac.uk/>) is a longstanding not-for-profit organisation that manages small molecule crystallographic data (Groom et al. 2016). The CCDC was founded over fifty years ago and is the dominant resource of its type in the chemical sciences. The CCDC has offered free online access to individual datasets for some time, but access to various discovery and analytical systems requires a subscription (Bruno and Ward 2016a, 2016b). Access to these systems is a non-collective benefit of paying the subscription and a by-product of providing the collective public-like good of a centralised data resource.

As political and academic pressure for greater access has increased, the Centre has opened up resources in a controlled way, but it retains key benefits for subscribers. A central argument for the CCDC

subscription/services model is that commercial interests provide a substantial amount of revenue. In turn, the restrictions that protect this revenue create some opportunity costs, difficult to quantify, for the system as a whole.

Two options exist: one, shifting to a model for subscription (non-collective) benefits other than access to data resources; two, shifting to an oligopoly model in which a consortium of interested industrial and public funders underwrite the costs of provision. While both of these are potentially plausible in principle, the major challenges are the risks of managing a transition to a new model that cannot be guaranteed to work.

The Arabidopsis Information Resource/Phoenix Bioinformatics

The Arabidopsis Information Resource (TAIR, <https://www.arabidopsis.org/>) provides plant genomic data resources to a wide community of biologists. In its original form as a US National Science Foundation (NSF)-funded project, it offered these resources free online. However, when the NSF progressively reduced and then removed funding, the resource was obliged to find a new model (Reiser et al. 2016). It shifted from being a free resource to a limited-access model with subscriptions for heavy users. In some senses, it has travelled the opposite direction as CCDC to arrive at a similar point.

The current TAIR model is for metered access. Any IP address can access a number of pages free each month (currently 75). Heavy users will encounter a subscription barrier. As with CCDC, subscribers gain additional non-collective benefits: in this case, unmetered access (Huala et al. 2016a). In common with CCDC, TAIR charges private corporations at a higher rate than academic users, who are subsidized alongside unaffiliated small scale users. There is also a community-building aspect because those who subscribe know they are contributing to the resource. TAIR is a longstanding resource that provides labour-intensive curation and management of highly structured data. The availability of this data underpins the ability of plant scientists to be able to manipulate and study plant genomes and functionality; therefore, the database provides a collective good.

The TAIR team (Huala et al. 2016b; Reiser et al. 2016) describe a high acceptance of this new model by the target community of core users. TAIR has many characteristics that align with its ability to fund a collective good, even though it is relatively large. Its success is likely attributable to the fact that TAIR has strong support from a relatively homogeneous community of disciplinary researchers. In addition, the crisis created by the loss of NSF funding created excellent conditions for the community to come together and solve an urgent collective problem that was now internal.

TAIR seeks to balance the need for revenue generation, and therefore non-collective benefits to members (i.e. database access), with the provision of public goods by allowing limited access to light users. The potential opportunity cost is difficult to quantify. TAIR notes that actual usage has not reduced substantially since subscriptions were introduced. Nonetheless, the model means that TAIR is more focused on member provisions and growing usage and membership than on developing the aspects of the resource that benefit the public good: 'A pay for access model provides a business incentive to increase usage. Result: data curation, discovery and reuse are priorities, as they drive up revenue' (Huala et al. 2016a). The focus here is on membership provision, again emphasising the (relatively small) community and identifying the means of showing the importance of their collective action in funding the resource.

Both CCDC and TAIR operate in ways that are club-like in the sense of Buchanan's club economics models. In both cases, the challenge is not identifying plausible models for making the resources more like public goods than they currently are. Rather, it is underwriting an attempted transition from one set of non-collective benefits to another that cannot be guaranteed to work in practice.

Worldwide Protein Data Bank

The Worldwide Protein Data Bank (wwPDB, <https://www.wwpdb.org/>) is the global host for structural data on biomolecules. Built from the beginning as an open resource (Berman 2008; Berman et al. 2012), its history is substantially different to that of the CCDC. The reasons for this difference are both historical and cultural, and beyond the scope of this paper, but the history and expectations of the community are important factors (Byrd et al. 2016).

The original Protein Data Bank (PDB) founded in 1971 was funded by US Federal Agencies for several decades. It experienced funding uncertainty in the early 2000s, which was part of the motivation for the formation of an international consortium in 2003. This consortium is made up of members that each seek funding from regional sources. Each member (the US, Europe, and Japan), is supported by a small set of funders which support a local data centre (Byrd et al. 2016). This is effectively an oligopoly model for each region. The regional approach is also a form of oligopoly which provides some resilience if a single region were to fail.

The wwPDB is such a widely used and important resource that it is too crucial to be allowed to fail at this point. A small number of resources reach this level of importance, but the wwPDB appears to be unique in its level of significance and its model of sustainability through a network of regional grants. Other data resources in this class are funded through unusual direct arrangements (such as Genbank, Pubmed Central, and other critical resources housed at the National Centre for Biotechnology Information [NCBI] within the US National Library of Medicine) or through intergovernmental agreements, as is the case for the biological resources supported at the European Bioinformatics Institute (EBI).

The models for all these critical resources can be characterised as a mixture of compulsory taxation models (the resources are top-sliced from funder budgets such as NCBI) or oligopoly models (such as the intergovernmental agreement which funds the EBI). It is worth noting that the intergovernmental oligopoly model is also the main model for large-scale physical scholarly infrastructures, particularly in Europe, including CERN, the Institut Laue-Langevin, and European Synchrotron Radiation Facility. While not the focus of the current article, this similarity suggests patterns of success may be consistent across very different infrastructures.

Europe Pubmed Central

Europe Pubmed Central (Europe PMC, <https://europepmc.org/>), supported by European funders and managed by the European Bioinformatics Institute (EMBL-EBI), is an open-access database that provides a central resource for bibliographic information and full-text scholarly articles on biomedical research (Europe Pubmed Central 2015). It started as a collaboration between UK funders, with the Wellcome Trust and Medical Research Council as main drivers. It now counts 27 funders as members of the funding consortium (Europe Pubmed Central 2016).

Europe PMC is an example of an oligopoly model in the way it solves the funders' collective action problem. The database started with the unilateral action of a small set of funders and has grown over time, but the Wellcome Trust remains the lead partner, and contributions are based on the size of the contributing funders. A five-year grant was awarded to the EMBL-EBI to fund Europe PMC's continuation through to 2020. As noted above, the oligopoly element here applies to the funders themselves. The collective good that the funders gain is access to and confidence in a platform that delivers on their mission, which most could not fund independently. However, from the perspective of the research community, this is a taxation model because the funding for Europe PMC is top-sliced from the general research funding pool.

It is difficult to see how Europe PMC could transition to a different sustainability model. It could conceivably expand the membership base to include institutions or publishers in some form, but the benefits of this are unclear. For their provision of the public good, contributing funders gain some non-collective benefits in the form of a designated repository that they can require their grantees to use (Europe Pubmed Central n.d.). However, the main goal of funding the database is to shift the research enterprise towards an open-access model, a clear case of a small group deciding to directly fund a collective good.

Crossref and ORCID

Crossref (<http://crossref.org>) provides a collective public-like good in the form of freely accessible bibliographic metadata and its supporting infrastructure. Financially, Crossref is supported by member publishers who gain the ability to assign Crossref Digital Object Identifiers (DOIs) to work that they publish and access to a set of services that aid in metadata creation, discovery, and validation of content (e.g. for plagiarism through the Similarity Check service).

Crossref was started essentially by fiat by a small number of publishers. In practice, between five and nine publishers dominate the market of scholarly publishing (Larivière et al. 2015). The setup and early funding of Crossref, while more complex than a simple agreement (Crossref 2009), was successful largely because a group representing a substantial proportion of scholarly publishing could come together. This is a classic oligopoly.

Crossref has offered non-collective membership benefits since its inception, including the ability to assign DOIs and the traffic that results from referral. A 2004 report conducted by Greenhouse Associates for an unnamed publisher noted the direct financial benefits arising for publisher members engaging with the Crossref system (Greenhouse Associates 2004). These benefits can be framed as a by-product of producing the public good of a shared metadata system.

Today, membership of Crossref is effectively compulsory for any serious publisher of scholarly articles in STEM subjects and increasingly so for publishers of articles in humanities and social sciences, as well as

scholarly book publishers. It is just part of the cost of doing business for a publisher—a compulsory, tax-like, part of the system.

It is helpful to contrast Crossref with the growth of ORCID (<http://orcid.org>; Haak et al. 2012), the provider of identifiers for contributors to scholarly work. Publishers were a core group of ORCID's early advocates, alongside a few funders. Publishers also provided the majority of its initial financing before funders started to contribute over the next several years. While a small group of publishers dominate the market, the group of important funders is larger. Consistent with Olson's predictions, publishers agreed on collective action to support ORCID before funders would. ORCID's slower growth, compared to Crossref is a consequence of the necessity for coordinating a large and more heterogeneous community.

Universities and research institutions have been very slow to engage with ORCID despite potentially reaping the greatest benefits from its usage. However, the most important benefits will only arise when there is substantial adoption by such institutions. Here we see the classic adoption problem described by Olson: there are too many universities and research institutions, even if we restrict ourselves to the traditional Euro-American centres of scholarship, to easily solve the collective action problem. Some progress has been made over the past few years; however, this has been primarily due to national coordination actions (Paglione 2016). Coordination at the level of tens of countries poses less of a collective action problem than across thousands of institutions. An important lesson from comparing the ORCID experience to that of Crossref is that where it is necessary to coordinate a large and more heterogeneous community an effective strategy is to shift that problem of coordination to a level where the number of players is small enough to solve the collective action problem.

AddGene

Addgene (<https://www.addgene.org/>) is an interesting counterpoint to the other resources here because it provides access to a physical resource. Addgene is a not-for-profit organisation founded in 2004 to provide access to plasmids, DNA materials used across many areas of biology. Addgene does not charge researchers to archive, store, and quality-assure plasmids (Kamens 2015). To obtain a sample of a specific plasmid, researchers pay a fee (Addgene n.d.).

The Addgene model works because it couples a collective benefit (central collection, archiving, and quality assurance of valuable materials) with a non-collective benefit (access to a physical rivalrous material, the plasmid itself). For users, paying the relatively modest fee to obtain a plasmid is well justified because Addgene provides quality assurance. While these materials circulate widely between labs, often without financial charge, the cost of lost time if the fragile materials degrade in transit can easily outweigh the upfront cost of obtaining quality-assured material from Addgene. Avoiding the risk of lost time, at the cost of an upfront fee, is a non-collective benefit. This is Olson's by-product model.

By centralising the availability and management of plasmids, Addgene creates a central information resource that is more valuable than the sum of its parts. The quality assurance, archiving, and even distribution processes are much cheaper when managed at scale, and Addgene is effective at exploiting that scale to create value for those paying for the resource.

The Politics of Compulsion and the Need for Shared Governance

Of the three approaches to sustainability, it is generally the by-product, or membership, model that funders expect infrastructures to pursue as they grow. Membership models can work in cases where the creation of club goods attracts members. The traditional model is to make membership a condition for access to the resource. Training experiences or access to valued meetings are possible alternatives that offer a route to sustaining open-access resources. Beyond the scholarly community, this alternative parallels the model of Patreon (<https://www.patreon.com/>), an online service where members pay to support artists and creators to produce freely available materials such as music, web comics, and online writing. Contributors get exclusive or early access to some materials, exclusive access to the artist (or more generally to specialist expertise) in online discussions, or a say in setting the artist's priorities.

Scholarly societies could play an important role in supporting such models. Both scholarly societies and academic libraries can also play an important role in reducing the effective number of actors in a negotiation. Libraries already often play this role by being the key point of contact for funding collective actions, including TAIR and CCDC, as well as other collective efforts, such as arXiv, the physics preprint server; the Inter-university Consortium for Political and Social Research (ICPSR); and the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3). A coordinated consortium with both the authority and credibility to negotiate on behalf of communities, has significant value if it reduces the number of effective actors to the point where collective action becomes possible.

In the scholarly infrastructures space, the compulsion/taxation and oligopoly approaches are very similar in practice because top-slicing by a group of funders amounts to a tax on overall research funds. As noted for Europe PMC, applying oligopoly models to solve the collective action problem among funders leads to a taxation system from the perspective of researchers. Some membership models, such as Crossref, also approach the level of compulsion. While compulsion is rare in scholarly communities, it is common in professional communities such as medicine, law, and some areas of engineering. Schemes offering professional certification (including the validation of degree programs by scholarly societies) blur this boundary as well.

The word 'compulsion' is pejorative, but there are many activities within the work of researchers that are compulsory. Gaining a doctorate, publishing at some level, and having access to the scholarly literature in some form are all effectively compulsory. These forms of compulsion (or call them 'social expectations,' if you prefer) are considered acceptable because they fit within a known and understood system of rules. Systems of taxations are acceptable, according to Adam Smith (1776), where there is proportionality, predictability, convenience, and efficiency. Today, we would probably add representation in governance and sustainability. Commitment to these principles requires us to build shared institutions, in the broad sense that Ostrom (2005, 3) describes. Much of political economics is bound up in trying to justify post-hoc the provision of institutions such as governments, courts, and the law by inventing concepts such as 'the social contract.' Our advantage as a scholarly community, or communities, is that we can explicitly develop and agree upon principles of operation as a way of reducing the costs of creating institutions.

A Common Set of Principles for Foundational Infrastructures

Building institutions is hard. It takes resources. Ostrom identifies this as a 'second order provisioning problem' (1991, 41). That is, while institutions can help enable a community to solve the collective action problem, this solution simply relocates the collective action problem to the problem of creating institutions, which are themselves collective goods. To reduce the costs of institution building, it makes sense to build templates: sets of agreed-upon principles under which such institutions and systems should operate. If our communities can sign up to a set of principles up front, then building institutions and infrastructures that reflect those principles should become a lot easier.

To address this issue, a draft set of Infrastructure Principles were developed (Bilder et al. 2015) to provoke a conversation about the governance and management of infrastructures within the scholarly community. These principles rest on three pillars: transparency and community governance; financial sustainability, efficiency, and commitment to community needs; and mechanisms to protect integrity, as well as manage and mitigate the risk of failures. They draw from observations of successes in providing foundational infrastructures and seek to generalise from these observations. The focus is on building *trustworthy* institutions.

The principles also map quite closely on to Smith's (1776) four principles for sound taxation. The commitment to representation is more modern, but a core part of building stable community and trust. The concept of enabling a community to 'fork' a project (i.e. make an independent working copy of it) by committing the project to use Open Source and Open Data, while modern in its approach, could be framed as an expression of the principles of efficiency and predictability. The intent of the Infrastructure Principles is to ensure that the community, or club, has a series of mechanisms that provide consistency and predictability. The first line is community governance and control; the second is financial stability. Forkability ensures there is a reliable path for dealing with a worst-case scenario in which an infrastructure is co-opted or 'goes rogue.'

Smith's principles and the Infrastructure Principles of Bilder et al. are examples of pattern languages in the sense advanced by Bollier and Helfrich (Bollier and Helfrich 2015; Helfrich 2015). Pattern languages are sets of templates that are highly generalisable, but need to be applied in specific contexts. They are less specific and rigid than fixed templates, but more contextualised than broad statements of practice. The Principles also speak to a broader goal of creating a good scholarly culture that supports infrastructures. Ultimately, we will need a broader understanding of how institutions and culture need to evolve as their communities grow (Hartley and Potts 2014; Wilson et al. 2014). In the shorter term, identifying patterns and understanding where they can be applied offers significant promise of making progress.

The link between these two perspectives, the classical economics and the more cultural community focus, can be found in the research programs developed out of the work of Ostrom. Ostrom's work (1991), which examined the political economy of communities managing collective resources, identified a set of patterns that defined successful governance institutions. Her work focused on common-pool resources—goods that

are rivalrous and non-excludable such as forests and fisheries—and showed that neither the state nor the market can successfully manage this category of collective goods. In contrast, communities can, at the appropriate scale, successfully manage these goods and how governance institutions can evolve that sustain them in the long term. While Ostrom worked towards the end of her life on extending these patterns of success to digital knowledge commons, this work remains underdeveloped.

Such patterns (or a future refinement or replacement for them) can serve scholarly communities in two ways. First, they can be used to set out the minimum requirements for governance and sustainability necessary before funders are willing to directly fund infrastructures (the oligopoly or tax mechanisms). Second, they provide a template for a developing club (either one built in a community sufficiently small to bootstrap its funding or one that has found a by-product model) to demonstrate its ability to make the transition from a local *club* to a community-wide *infrastructure*.

Predictions and Practical Consequences

While the case studies are helpful, the above remains an abstract argument. What are its practical consequences for actually sustaining scholarly infrastructures? First, we can make a prediction to be tested: All sustainable scholarly infrastructures providing collective (public-like) goods to the research community will be funded by either one of the three models (taxation, by-product, oligopoly) identified by Olson or some combination of them.

Second, we can look at stable longstanding infrastructures (Crossref, Protein Data Bank, NCBI, arXiv) and note that, in most cases, governance arrangements are an accident of history and were not explicitly planned. Crises of financial sustainability (or challenges of expansion) for these organisations are often coupled with or lead to a crisis in governance and/or community trust. Changes are therefore often made to governance in response to a specific crisis.

Where there is governance planning, it frequently adopts a best-practice model that looks for successful examples to draw from. It is not often based on worst-case scenario planning. This is a problem. We can learn as much from failures of sustainability and their relationship to governance arrangements as from successes. The strength of Olson's, Buchanan's, and Ostrom's work is as much in showing us where collective action is difficult or impossible as in identifying how it can be successful.

We can look towards using these economic models to investigate what forms of funding and organisational structures are likely to work for the future. An initial suggestion is that membership-fee-based sustainability models will only work where a non-collective benefit is provided. In turn, this means identifying where such non-collective benefits are a viable model. A deeper understanding of which non-collective benefits are appropriate will be valuable and will help address the assumption that membership and subscription systems can only be tied to content access. It will also be important to identify where non-collective benefits are not a viable model and to avoid forcing the model on organisations for which it cannot work.

With this deeper understanding of how models work, and in particular how they relate to the scale of communities, it should be possible to create pathways to sustainability. These pathways would link the growth of infrastructures from club-like services, through community adoption, to the creation of public goods. They would be connected with the appropriate sustainability models and financial instruments to support them and a graduated set of governance requirements appropriate to the stage of growth and development.

Previous work on sustaining infrastructures has mostly focused on the examination of specific financial models (Bastow and Leonelli 2010; Ember and Hammisch 2013). Crow's (2013) report for Knowledge Exchange offers the major exception here, and its practical development of the options that Assurance Contracts may offer is an important complement to this paper. While the issue of financial sustainability is key, and the political discussion of the balance of resource allocation between infrastructure platforms and research projects is important, my claim here is that most studies have not addressed two important dimensions in sufficient detail.

First, infrastructures need to be seen as both sustaining and being sustained by the communities that they serve. *Political economy* needs to be addressed, not simply financial issues. Second, the size of the community and the scale of the infrastructure are critical factors for defining what sustainability models can work, and sustainability models must therefore change throughout the growth and development of an infrastructure.

Buchanan, Olson, Ostrom, and the many scholars who have developed their work have shown that simple modelling and analysis could help us understand and make sense of successes and failures in tackling collective action. Olson and Ostrom in particular worked from case studies and illuminated them with models

and theory. All three scholars showed that there are key contributing factors for a community seeking to provide collective goods. Among these, perhaps the most important is that the size of the community is a key determinant of success.

Above all, the key is to learn from our experiences, as well as from the theory of economics and governance, to identify the patterns and templates that produce resilient organisations and infrastructures whose trustworthiness earns the trust of their communities through both good times and bad. The good news is that the work of Olson and others, even though it is nearly 50 years old, is a valuable and largely untapped resource to draw on.

Competing Interests

The author was a contributor to the *Infrastructure Principles* (Bilder et al. 2015) and a member of the OECD Expert Group on Funding of Data Infrastructures. He is a board member and President (2016–17) of FORCE11 and a contributor to work of Knowledge Unlatched Research, all of which can be seen as infrastructures that would benefit from shared funding arrangements.

References

- Addgene. n.d. 'Addgene: How to Order.' Addgene. Accessed March 2017. <https://www.addgene.org/ordering/>. Archived at: <https://perma.cc/2YZ4-8AC8>.
- Bastow, Ruth, and Sabina Leonelli. 2010. 'Sustainable Digital Infrastructure.' *EMBO Reports* 11(10): 730–34. DOI: <https://doi.org/10.1038/embor.2010.145>
- Berman, Helen M. 2008. 'The Protein Data Bank: A Historical Perspective.' *Acta Crystallographica Section A: Foundations of Crystallography* 64(1): 88–95. DOI: <https://doi.org/10.1107/S0108767307035623>
- Berman, Helen M., Gerard J. Kleywegt, Haruki Nakamura, and John L. Markley. 2012. 'The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future.' *Structure* 20(3): 391–96. DOI: <https://doi.org/10.1016/j.str.2012.01.010>
- Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. 2015. 'Principles for Open Scholarly Infrastructures-V1.' Figshare. DOI: <http://doi.org/10.6084/M9.FIGSHARE.1314859>
- Bollier, David, and Silke Helfrich. 2015. 'Overture.' In: *Patterns of Commoning*, edited by David Bollier and Silke Helfrich, 1–12. Amherst, MA: Levellers Press.
- Bruno, Ian, and Suzanna Ward. 2016a. 'Sustaining Access to Research Data through Value-Added Services and Software.' In *SciDataCon*. Denver. <http://www.scidatacon.org/2016/sessions/45/paper/148/>.
- Bruno, Ian, and Suzanna Ward. 2016b. 'Witness Statement – CCDC.' *SciDataCon Session: Witness Statements on Repository Business Models*. http://wiki.codata.org/w/images/wiki.codata.org/8/8d/SciDataCon-45-Business_Models-CCDC_statement.pdf.
- Buchanan, James M. 1965. 'An Economic Theory of Clubs.' *Economica* 32(125): 1–14. DOI: <https://doi.org/10.2307/2552442>
- Byrd, R. Andrew, Helen M. Berman, Stephen K. Burley, Gerard J. Kleywegt, John L. Markley, Haruki Nakamura, and Sameer Velankar. 2016. 'Economics and Impact of the Protein Data Bank (PDB) Archive.' In: *SciDataCon*. Denver. <http://www.scidatacon.org/2016/sessions/45/paper/91/>.
- Crossref. 2009. 'The Formation of Crossref: A Short History.' <https://www.crossref.org/pdfs/CrossRef10Years.pdf>.
- Crow, Raym. 2013. 'Sustainability of Open Access Services. Phase 3: The Collective Provision of Open Access Resources.' *Knowledge Exchange*. <http://repository.jisc.ac.uk/6206/1/Sustainability%2BOA%2BService%2Bphase%2B3.pdf>.
- Ember, Carol, and Robert Hammisch. 2013. 'Sustaining Domain Repositories for Digital Data: A White Paper.' ICPSR. http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf.
- Europe Pubmed Central. n.d. 'Joining Europe PMC.' Europe Pubmed Central. Accessed March 2017. <https://europepmc.org/Joining>. Archived at: <https://perma.cc/4Z53-U9QX>.
- Europe Pubmed Central. 2016. 'Europe PMC Annual Report 2015.' Hinxton: European Bioinformatics Institute. https://europepmc.org/docs/Europe_PMC_Annual_Report_2015.pdf.
- Europe Pubmed Central, Europe. 2015. 'Europe PMC: A Full-Text Literature Database for the Life Sciences and Platform for Innovation.' *Nucleic Acids Research* 43(D1): D1042–8. DOI: <https://doi.org/10.1093/nar/gku1061>
- Greenhouse Associates. 2004. 'Business Impact of the Digital Object Identifier (DOI).' http://www.greenhousegrows.com/publications/#id_of_div6.

- Groom, Colin R., Ian J. Bruno, Matthew P. Lightfoot, and Suzanna C. Ward. 2016. 'The Cambridge Structural Database.' *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* 72(2): 171–79. DOI: <https://doi.org/10.1107/S2052520616003954>
- Haak, Laurel L., Martin Fenner, Laura Paglione, Ed Pentz, and Howard Ratner. 2012. 'ORCID: A System to Uniquely Identify Researchers.' *Learned Publishing* 25(4): 259–64. DOI: <https://doi.org/10.1087/20120404>
- Hartley, John, and Jason Potts. 2014. *Cultural Science*. London: Bloomsbury 3PL.
- Helfrich, Silke. 2015. 'Patterns of Commoning: How We Can Bring about a Language of Commoning.' In: *Patterns of Commoning*, edited by David Bollier and Silke Helfrich, 1–12. Amherst, MA: Levellers Press.
- Huala, Eva, Tanya Berardini, Donghui Li, Robert Muller, Leonore Reiser, and Emily Strait. 2016a. 'TAIR's Successful Transition to a Sustainable Business Model.' *SciDataCon 2016*. <http://www.scidatacon.org/2016/sessions/45/paper/51/>.
- Huala, Eva, Tanya Berardini, Donghui Li, Robert Muller, Leonore Reiser, and Emily Strait. 2016b. 'Witness Statement – TAIR.' *SciDataCon 2016 Session: Witness Statements on Repository Business Models*. http://wiki.codata.org/w/images/wiki.codata.org/2/2f/Witness_statement-TAIR.pdf.
- Kamens, Joanne. 2015. 'The Addgene Repository: An International Nonprofit Plasmid and Data Resource.' *Nucleic Acids Research* 43(D1): D1152–57. DOI: <https://doi.org/10.1093/nar/gku893>
- Larivière, Vincent, Stefanie Haustein, and Philippe Mongeon. 2015. 'The Oligopoly of Academic Publishers in the Digital Era.' *PLOS ONE* 10(6): e0127502. DOI: <https://doi.org/10.1371/journal.pone.0127502>
- Lawson, Stuart. 2017. 'Access, Ethics and Piracy.' *Insights* 30(1). DOI: <https://doi.org/10.1629/uksg.333>
- Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Reprinted edition of 1971. Princeton, NJ: Harvard University Press.
- Ostrom, Elinor. 1991. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom, Elinor. 2005. *Understanding Institutional Diversity*. Princeton: Princeton University Press.
- Paglione, Laura. 2016. '2016: The Year in Review.' *ORCID Blog*. December 29, 2016. <https://orcid.org/blog/2016/12/29/2016-year-review>. Archived at: <https://perma.cc/8PS6-EWCN>.
- Reiser, Leonore, Tanya Z. Berardini, Donghui Li, Robert Muller, Emily M. Strait, Qian Li, Yarik Mezheritsky, Andrey Vetsushko, and Eva Huala. 2016. 'Sustainable Funding for Biocuration: The Arabidopsis Information Resource (TAIR) as a Case Study of a Subscription-Based Funding Model.' *Database* 2016 (January). DOI: <https://doi.org/10.1093/database/baw018>
- Smith, Adam. 1776. *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W Strahan and T Cadell. DOI: <https://doi.org/10.1093/oseo/instance.00043218>
- Star, Susan Leigh, and Karen Ruhleder. 1996. 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces.' *Information Systems Research* 7(1): 111–34. DOI: <https://doi.org/10.1287/isre.7.1.111>
- Wilson, David Sloan, Steven C. Hayes, Anthony Biglan, and Dennis D. Embry. 2014. 'Evolving the Future: Toward a Science of Intentional Change.' *The Behavioral and Brain Sciences* 37(4): 395–416. DOI: <https://doi.org/10.1017/S0140525X13001593>

How to cite this article: Neylon, Cameron. 2017. Sustaining Scholarly Infrastructures through Collective Action: The Lessons that Olson can Teach us. *KULA: knowledge creation, dissemination, and preservation studies* 1(1): 3. DOI: <https://doi.org/10.5334/kula.7>

Submitted: 17 March 2017

Accepted: 14 July 2017

Published: 27 December 2017

Copyright: © 2017 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.