



SPECIALISED PORTALS, ONLINE INFORMATION SERVICES, SCHOLARLY ONLINE NETWORKS: THE IMPACT OF E-INFRASTRUCTURES ON SCIENCE COMMUNICATION AND SCHOLARLY COMMUNITY BUILDING

Open infrastructure and community: the case of astronomy

Niels Taubert

Abstract

This comment focuses on an early case of an open infrastructure that emerged in the 1990s in international astronomy. It targets the reasons for this infrastructure's tremendous success and starts with a few comments on the term 'digital infrastructure'. Subsequently, it provides a brief description of the most important components of the infrastructure in astronomy. In a third step, the use of one component — the arXiv, an open access repository for manuscripts — is analyzed. It concludes with some considerations about the success and acceptance of this infrastructure in astronomy.

Keywords

Scholarly communication; Science communication: theory and models

DOI

<https://doi.org/10.22323/2.17020302>

Introduction

The creation and establishment of a discipline-specific digital infrastructure is challenging. In many cases it is highly questionable before or during its development whether the requirements of the targeted user group [Van Zundert, 2012] will be met, and when completed whether the extent of its utilization may stay behind the initial expectations [Kaltenbrunner, 2017, p. 303] often due to nebulous reasons. The aim of this comment is to approach at least some of the difficulties of the development of a digital infrastructure in science by drawing on a successful case: the digital infrastructure in international astronomy. In a first step, a heuristic model is given. It illustrates that the development of an infrastructure is not merely a technical task, but a creation of a complex arrangement that includes a two-sided social embedding of the technology. The second step focuses on the technical side and describes the digital infrastructure in astronomy with its most important components and major characteristics. The third step narrows the perspective to one of its components — the *arXiv* repository — and reconstructs how it is being used. The comment concludes with some considerations about the conditions for success drawn from the example.

A few remarks on digital infrastructures in science

From a sociological point of view, a digital infrastructure should not be regarded as a merely technical ‘thing’ but as a complex arrangement where three layers can be distinguished analytically (see Figure 1).¹

The first layer is a *system of action* for which a digital infrastructure provides resources. It is important to note that ‘resources’ are context sensitive: the output of the technical layer acts as a resource only in the context of a specific system of action. In the course of the action, the output of the technical layer is activated by specific rules applied by the actor [Schulz-Schaeffer, 1999; Schulz-Schaeffer, 2000]. Regarding science, the formal communication system of science where new findings are registered, certified, and circulated within the scientific community is an example for such a system of action. Another one is the research system in which validity claims are tested, data are gathered and experiments are conducted.

The second layer comprises all *technological components and rules* that are necessary for the proliferation of an output. Given that the output of the technical layer constitutes a resource only in a specific context of use, the question whether or not a component is part of the infrastructure can only be determined with reference to the system of action that is supported by the infrastructure.² For example, the publication infrastructure supports the formal communication system of science and comprises publication media like journals, conference proceedings, journal platforms, open access repositories, online editorial management systems, classifications embodied in online catalogues, (inter-)disciplinary portals, citation databases and many other components.

Service organizations that follow the mission of the creation, development and maintenance of components of the infrastructure during its use constitute a third layer of the heuristic model. In the case of the publication infrastructure these organizations include publishing houses, information providers, libraries, and research organizations.

The three layers can only be distinguished analytically as they form a tripartite structure and are connected through relationships of enabling. As noted above, the service organizations ensure the development and maintenance of the publication infrastructure, while the latter is a prerequisite of the formal communication system of science. In spite of their differences, all three layers are social phenomena, which can be subjected to sociological analysis.

Digital infrastructure of astronomy

An interesting case is the digital infrastructure in international astronomy as it comes with some striking features and can be regarded as a dream come true for open science protagonists. It consists of large components that are closely connected and offer immediate or delayed access to a large share of publications and research data. It has been in place since the mid-1990s and is widely being used by astronomers all over the world.

In astronomy, there is a core of *large peer reviewed journals* that publish the bulk of articles. According to the Web of Science Citation Report, the three largest journals

¹For a more theory-based introduction to the heuristic, see Taubert [2016] and Taubert [2018].

²In this respect, the heuristic model picks up a relational understanding developed in Star and Ruhleder [1996], Star [1999], Ribes and Lee [2010].

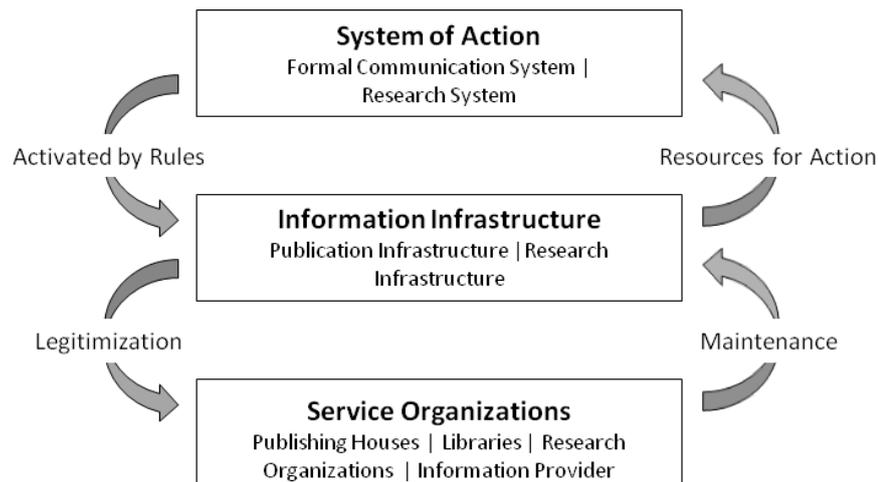


Figure 1. Digital infrastructure — heuristic model.

are the *Monthly Notice of the Royal Astronomical Society* (MNRAS) with 2,790 citable items,³ the *Astrophysical Journal* (ApJ) with 2,785 citable items, and *Astronomy & Astrophysics* (A&A) with 1,735 citable items in 2014. The next four core journals are also important but smaller. These are the *Astrophysical Journal Letters* (ApJL) with 669 citable items, the *Astronomical Journal* (AJ) with 296 citable items, the *Astrophysical Journal Supplement Series* (ApJS) with 159 citable items, and the *Publications of the Astronomical Society of the Pacific* (PASP) with 90 citable items. In addition to the concentration, two other characteristics are remarkable. All of the journals are owned by learned societies or research organizations and are, as a result, freely accessible to a large extent. The four journals of the *American Astronomical Society* (ApJ, AJ, ApJL, and the ApJS), A&A, and PASP apply a moving wall open access model that allows free access after a period from 12 to 24 months.⁴

The astronomical journals deliver abstracts and metadata of their publications to a second component of the infrastructure, the *Astrophysics Data System* (ADS) located at Harvard University (U.S.A.), which can be described as a central information hub in astronomy. It started as a subject database and abstract service but developed and now supplies the community with all older relevant astronomical literature [Eichhorn, 2004]. Regarding currently published research, it provides links to the publications in journals and conference proceedings. In cases where publications are also self-archived, ADS also links to the document on the repository. If a publication reports findings about an astronomical object and if there are data on the object, ADS provides a link to the data sets. Since ADS is run by a publicly funded research organization, it is freely accessible online.

Prominent databases for observational data are run by the *Centre de Données Astronomiques de Strasbourg* (CDS, France) and include *SIMBAD*, *VizieR*, and *ALADIN* [Genova et al., 1998]. *SIMBAD* is a database with descriptions of astronomical objects, *VizieR* is a collection or star catalogue, and *ALADIN* is an

³According to the definition of the Web of Science citable items include the document types article, review and proceedings paper.

⁴For a more detailed description and the history of the astronomical core journals, see Osterbrock [1995] (ApJ), Dalterio et al. [1995] (ApJS and ApJL), Bracher [1999] (AJ), Murdin [2005] (MNRAS), and Pottasch [2011] (A&A).

interactive software-based star atlas. Like ADS, the databases of CDS can also be used without restrictions.

The last component of the astronomical digital infrastructure that should be mentioned is the subject specific repository *arXiv* based in Cornell University (U.S.A.). It was introduced in astronomy in the year 1992 and allows depositing manuscripts. Like other repositories, the *arXiv* does not offer any peer review and should therefore not be understood as a substitute for the journal but as a second channel for the dissemination of research. It allows depositing papers at any time before, during, and after the peer review and makes them freely accessible online.

The use of arXiv

The next step will focus on the *arXiv* and will give a brief reconstruction about how it is used in astronomy. Self-archiving publications and making them freely accessible online (often called 'green open access') is widespread in astronomy. In the course of a research project,⁵ I found out that 61.57% of the publication output of 102 randomly selected astronomers was self-archived mainly on the *arXiv*. In addition to bibliometric analyses, I conducted in-depth interviews to gain more information about the background and decision of self-archiving.⁶ One focus was on the question at what point of the publication process astronomers self-archive their publications. The interviews give evidence that a considerable number of publications appear as pre-prints, not only before publication in a journal but even before completion of the peer review process. The interviewees gave two reasons why to choose such an early point in time: first, there is a high level of competition in some fields in astronomy. The research frontier is moving fast and there is a need to publish first. In this context of use the repository acts as a registry that protects priority.⁷ Second, to improve the chance of getting their work published, authors in astronomy are interested in receiving feedback from colleagues before submitting their papers to journals. Within this context of use the repository acts as a two-way medium that addresses a specific community and that allows them to react to a paper.⁸

When shifting from the authors' to the readers' perspective, it becomes apparent that there is one important difference between pre-prints on the *arXiv* and publications in a journal. Early self-archiving before completion of peer review de facto bypasses the evaluation procedure which is a precondition for trust in published research. Therefore, the pre-print self-archiving routines of some of the authors evoke the question whether readers deal with pre-prints in a specific way, taking their potential non-peer-reviewed nature into account. In the interviews four types of routines could be identified. First, readers are highly aware of context information of pre-prints. They are especially interested whether or not a pre-print

⁵The research project and the methods are described in Taubert [2014] and Taubert [2018].

⁶The ten interviewees are located in Germany and South Africa coming from a variety of different institutes and observatories. The sample included astronomers from the European Southern Observatory, the Max-Planck-Institute for Astronomy, the Hamburg Observatory, the South African Astronomical Observatory, the Cape Town University and the North-West-University (Potchefstroom).

⁷An example from the interviews: "I think everyone wants to get their work out into the public domain as soon as possible. That's the driving reason [for self-archiving, NT]" (I 15, 00:45:28).

⁸One astronomer explains this motive: "They would be regarding [...] the *arXiv* as a reviewing body, seeing the entire community saying, "Oh they could give me feedback. And they can strengthen my paper by saying it's crap or it's good"" (I 12, 00:54:03).

has already been accepted for publication in a journal.⁹ Second, information about the author or the author group is interpreted in terms of trustworthiness. If the colleagues or the working group are known for good quality in the past, it is likely that the reader also trusts in findings reported in a pre-print.¹⁰ Third, astronomers limit the citation of pre-prints. For example, when writing a paper they avoid basing their own argument on a non-published manuscript.¹¹ Fourth, they sharply distinguish between trustworthy and untrustworthy components of a paper. In observational astronomy data are often generated by large observatories and do not have to be subject to peer review. From the astronomers' perspective, they can be used right away. In contrast, interpretations of data should be peer reviewed before being cited.¹²

Self-archiving and use of pre-prints in astronomy shows that there is a co-evolution of the technical infrastructure, on the one hand, and routines of scientists in which technical means are employed, on the other. Routines of authors and readers are complementary and allow speedier dissemination of new results and findings within the astronomers' community. Speed is not only a result of early self archiving but also of the readers' routines that help to make pre-prints useable.

Conclusions

Put into a broader perspective, the example of the digital infrastructure in astronomy invites to ask for possible causes for its success. Regarding self-archiving, two epistemic characteristics of astronomy may be important for the development of routines of action: first, competition for priority gives strong incentives for self-archiving and also for reading pre-prints as early as possible. Second, the benefit of feedback from colleagues that helps to pass peer review is high, if a community evaluates research according to the same criteria and has a shared understanding about what is good research. In astronomy, this is the case¹³ and makes it likely that such routines develop.

A technical infrastructure that provides an output that acts as resources for action will hardly evolve by accident. The case of the infrastructure in international astronomy is also instructive here as it points to three factors that make a co-development of the output of the technical layer and the routines of action likely. First, the impulse for the creation of the components of the infrastructure came from the scientific community. This holds for the large astronomy journals that were created by scientific communities as well as for ADS and CDS. The *arXiv*

⁹Such interpretation is explained by interviewee I 4: "If manuscripts are on a preprint-server but not published in a refereed journal after half a year or so, one would not use them. Or I wouldn't use them" (I 4, 00:11:29).

¹⁰Interview I 12 explains that the name of the author acts as a proxy for trust: "[...] depends who the author is [...] you sort of know the work of certain people. [...] It is a small community, there are a few hundred people [...] so you know most of the people who are working on the same kinds of things" (I 12, 00:40:59).

¹¹As an example for the restriction of the use of pre-prints: "So it's something insignificant in a sense in that it's the latest news [...] Then it's okay [to cite a pre-print, NT]. So I wouldn't really place big important things on pre-review papers [...] There is a small role for that I would say, but yeah, keep it to a minimum" (I 3, 00:15:25).

¹²"Especially in this area where I'm quite interested in the observation on astronomy so the simple just reporting of observations doesn't necessarily need to be peer reviewed. It's the interpretation of the results, of the data that needs peer-reviewing really [...]" (I 15, 00:19:21).

¹³An expression of shared quality standards are low rejection rates in astronomy journals between 10–18% [Abt, 2009].

is also an example here, since the innovative impulse came from a neighboring discipline (physics, Ginsparg [1994]). Second, all components of the infrastructure are controlled by the scientific community, and the service organizations that maintain the components are (a) actively chosen by astronomers (like the publishing houses that edit the journals) or (b) strongly embedded in the community. The latter is the case for ADS und CDS both run by highly respected astronomical research organizations. Third, the community of astronomers is well organized and is able to articulate their requirements and needs. These characteristics of the discipline should not be understood as a necessary pre-condition for the creation of a digital infrastructure but make it more likely to succeed.

Acknowledgments This comment reports findings from a research project supported by Deutsche Forschungsgemeinschaft (grant number TA 720 1/1).

References

- Abt, H. A. (2009). 'Reviewing and Revision Times for The Astrophysical Journal'. *Publications of the Astronomical Society of the Pacific* 121 (885), pp. 1291–1293. <https://doi.org/10.1086/648536>.
- Bracher, K. (1999). 'The Astronomical Journal: A Mirror of Astronomy'. *The Astronomical Journal* 117 (1), pp. 12–16. <https://doi.org/10.1086/300681>.
- Dalterio, H. J., Boyce, P. B., Biemesderfer, C., Warnock, A., Owens, E. and Fullton, J. (1995). 'The Electronic Astrophysical Journal Letters project'. *Vistas in Astronomy* 39 (1), pp. 7–12. [https://doi.org/10.1016/0083-6656\(95\)91995-s](https://doi.org/10.1016/0083-6656(95)91995-s).
- Eichhorn, G. (2004). 'Ten years of the Astrophysics Data System'. *Astronomy and Geophysics* 45 (3), pp. 3.07–3.09. <https://doi.org/10.1046/j.1468-4004.2003.45307.x>.
- Genova, F., Bartlett, J. G., Bonnarel, F., Dubois, P., Egert, D., Fernique, P., Jasniewicz, G., Lesteven, S., Ochsenbein, F., Wenger and M. (1998). 'The CDS Information Hub'. In: *Astronomical Data Analysis Software and Systems VIII. ASP Conference Series 145*. Ed. by R. Albrecht, R. N. Hook and H. A. Bushouse. San Francisco, U.S.A.: Astronomical Society of the Pacific, pp. 470–473.
- Ginsparg, P. (1994). 'First Steps Towards Electronic Research Communication'. *Computers in Physics* 8 (4), p. 390. <https://doi.org/10.1063/1.4823313>.
- Kaltenbrunner, W. (2017). 'Digital Infrastructure for the Humanities in Europe and the US: Governing Scholarship through Coordinated Tool Development'. *Computer Supported Cooperative Work (CSCW)* 26 (3), pp. 275–308. <https://doi.org/10.1007/s10606-017-9272-2>.
- Murdin, P. (2005). 'Monthly Notices of the Royal Astronomical Society'. In: *Communicating Astronomy*. Ed. by T. J. Mahoney. Santa Cruz de Tenerife, Spain: Instituto de Astrofísica de Canarias (IAC).
- Osterbrock, D. E. (1995). 'Founded in 1895 by George E. Hale and James E. Keeler: The Astrophysical Journal Centennial'. *The Astrophysical Journal* 438, p. 1. <https://doi.org/10.1086/175049>.
- Pottasch, S. R. (2011). 'The history of the creation of Astronomy & Astrophysics'. *EAS Publications Series* 49, pp. 23–31. <https://doi.org/10.1051/eas/1149002>.
- Ribes, D. and Lee, C. P. (2010). 'Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories'. *Computer Supported Cooperative Work (CSCW)* 19 (3–4), pp. 231–244. <https://doi.org/10.1007/s10606-010-9120-0>.

- Schulz-Schaeffer, I. (1999). 'Technik und die Dualität von Ressourcen und Routinen. Zur sozialen Bedeutung gegenständlicher Technik'. *Zeitschrift für Soziologie* 28 (6), pp. 409–428. <https://doi.org/10.1515/zfsoz-1999-0601>.
- (2000). *Sozialtheorie der Technik*. Frankfurt/New York: Campus.
- Star, S. L. (1999). 'The Ethnography of Infrastructure'. *American Behavioral Scientist* 43 (3), pp. 377–391. <https://doi.org/10.1177/00027649921955326>.
- Star, S. L. and Ruhleder, K. (1996). 'Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces'. *Information Systems Research* 7 (1), pp. 111–134. <https://doi.org/10.1287/isre.7.1.111>.
- Taubert, N. (2014). 'Green Open Access in Mathematik und Astronomie'. In: *Forschung und Publikation in der Wissenschaft. Jahrbuch Wissenschaftsforschung 2013*. Ed. by H. Parthey and W. Umstätter. Berlin, Germany: Wissenschaftlicher Verlag Berlin (wVB), pp. 42–75.
- (2016). 'Digitale Publikations- und Forschungsinfrastrukturen'. In: *Handbuch Wissenschaftspolitik*. Ed. by S. H. D. Simon A. Knie and K. Zimmermann. Wiesbaden, Germany: Springer, pp. 1–18. https://doi.org/10.1007/978-3-658-05677-3_32-1.
- (2018). *Fremde Galaxien und abstrakte Welten. Open Access in Astronomie und Mathematik: eine soziologische Analyse*. Bielefeld, Germany: Transcript.
- Van Zundert, J. (2012). 'If You Build It, Will We Come? Large Scale Digital Infrastructures as a Dead End for Digital Humanities'. *Historical Social Research* 37 (3), pp. 165–186.

Author

Niels Taubert, Ph.D. studied sociology at the Universities of Hamburg and Bielefeld and obtained his PhD from Bielefeld University for a thesis on open source software development. He currently completed a habilitation thesis on open access in astronomy submitted to the University of Kassel. He is the head of the bibliometric working group at the Institute for Interdisciplinary Studies of Science (I²SoS), Bielefeld University. E-mail: niels.taubert@uni-bielefeld.de.

How to cite

Taubert, N. (2018). 'Open infrastructure and community: the case of astronomy'. *JCOM* 17 (02), C02. <https://doi.org/10.22323/2.17020302>.



© The Author(s). This article is licensed under the terms of the Creative Commons Attribution — NonCommercial — NoDerivatives 4.0 License. ISSN 1824-2049. Published by SISSA Medialab. jcom.sissa.it