From collecting to connecting – the role of libraries in Open Access

A short history of libraries and Open Access

In the beginnings of Open Access, libraries were not directly involved. The Budapest Open Access Initiative had no representative of a major library among them (cf. Chan et al., 2002). The Berlin Declaration had the German library association as an initial signatory but still a systematic involvement of libraries worldwide was missing (Berliner Erklärung, 2003). However, first systematic accounts of the role of libraries were discussed (Suber, 2003). Predominantly, these started from the hope that Open Access would be a solution to the ‘serial crisis’, i.e. the negative spiral of price increases for academic journals introduced by publishers forcing a reduction of subscriptions by libraries, yielding, in turn, to increased pricing for journals since publishers wanted to keep revenues. Whether or not the serial crisis was (or is) the appropriate motivation for Open Access services in libraries remained controversial, even among the Open Access proponents (Bosc & Harnad, 2005). On the one side, rationales dominated that claimed to keep the publishing system largely untouched by putting copies of subscription journal articles in institutional and subject-based repositories, often called the Green route to Open Access or Open Access Archiving. On the other side, rationales that introduced direct forms publishing in journals or other formats dominated, often called Golden route to Open Access. The Golden route was either financed by pre-payment for an individual article (the so called “Author-Pays model”, later termed “Article Processing Charges”, APC) or through in-kind contributions of mostly public organisations financing free-of-charge Open Access journals (Björk et al., 2010).

Open Access Services in Libraries

Ever since the beginnings of Open Access at the turn of the millennium, the world of Open Access kept differentiating, resulting in the current situation of a ‘mixed economy’, ranging from subscriptions and archived versions in repositories over full open access journals, either free-of-charge or payable and hybrid journals (see below) to interim models that contain only components of one of those models. And as differentiated as this mixed economy is, the roles of libraries varies. The following description of the role of libraries in Open Access has components
of a historic, developmental perspective and shall lead to a service categorization that can be used by newcomers to the field or students of library and information science. The examples given will be biased towards experiences and observations of the author and therefore contain over-proportionally many examples from Germany, UK and Europe.

Advocacy and reference services

*Service Definition:* Authors receive information from a librarian or a library website on what Open Access is, why it is important, how it can be supported and where to get further information.

Libraries offer advocacy and reference services since the beginnings of Open Access. Initially, existing lists in the library community such as ‘Liblicense’ (Okerson, 1999) were used to discuss librarians’ roles in Open Access. Libraries started to develop information resources. Conferences such as the Berlin conference gained increasing interest among librarians. Today, there is probably almost no library that has not at least one contact point for Open Access among staff and almost no website of an academic library does not mention Open Access. Even national information portals have been developed, for example in Germany (Hätscher, 2007), a continental coordination such as the European National Open Access Desks (Rettberg & Schmidt, 2012) or international bodies for advocacy and exchange such as the Confederation of Open Access Repositories (Peters & Lossau, 2011).

Institutional (Open Access) Repositories

*Service definition:* An internet service that stores and publicly offers Open Access versions of publications authored by researchers at a given academic institution.

At the third edition of the abovementioned Berlin conferences, it was controversially discussed how to prioritize and support the Green and the Golden route to Open Access, resulting in a simple formula, also known as the keystroke strategy: “(1) Implement a policy to **require** their researchers to deposit a copy of all their published articles in an Open Access repository. And (2) **encourage** their researchers to publish their research articles in Open Access journals where a suitable journal exists and provide the support to enable that to happen.” (Harnad, 2005). At this time Open Access Repositories were already being established
It is noteworthy, though, that early accounts did not necessarily have Open Access in the centre of the institutional repository. A frequently used definition does not even mention Open Access: „... a university-based institutional repository is a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members. It is most essentially an organizational commitment to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution“ (Lynch, 2003). However, the notion of an Institutional Open Access Repository became paradigmatic. With the registries ROAR ¹ and OpenDOAR ² (see also below „Registries and Aggregators“), worldwide lists of repositories were established: ROAR shows exponential growth with individual precursors in the 1990s and over 4000 repositories in 2016. OpenDOAR shows registrations of some dozens of repositories in 2005 up to over 3000 in 2016. In sum, it can be assumed that the majority of academic libraries today operate an institutional Open Access repository or participate in a shared service.

Subject based (Open Access) Repositories

Service definition: An internet service that stores and publicly offers Open Access versions of publications authored by researchers from a given subject community.

Libraries are mainly focussing on institutionally created content in their institutional Open Access repositories (see above). But the story of Open Access and library services for Open Access cannot be told without mentioning subject based Open Access repositories. These are services that provide articles for a given subject community. And these services even precede the story of Open Access. In the 1990s, arXiv (Ginsparg, 1994) was developed to share preprints, i.e. versions of documents that have been submitted to a journal for review but have not yet been published by the journal. It started from high energy physics but expanded into other mathematically formalized disciplines, e.g. computer science (Luce, 2001). The main commonality between all papers in arXiv is actually that they are produced with LaTeX (Lamport, 1994), a type-setting language that allows to produce high-quality layouts of articles already at the stage of submission. Another subject based service is RePEc (Krichel, 2000) that was developed to collect working paper series in Economics before they were published and even before submitted to a journal. These papers were shared by Economy departments around the world on local (institutional) web-servers. It is noteworthy that both arXiv and RePEc deal with pre-published versions of articles but they differ substantially in that arXiv papers serve a function of an early surrogate to the final journal article while RePEc papers also fulfill a discussion function. Additionally arXiv is a genuine repository that stores and shares papers uploaded by the authors in a single, integrated service on the web.

¹ http://roar.eprints.org
² http://www.opendoar.org
while RePEc is primarily a bibliographic aggregation service that collects information from a globally distributed system of local web-servers. A third major subject based Open Access repository is PubMedCentral, PMC\(^3\). Based on the bibliographic services Medline and PubMed, PMC displays that section of articles that have an Open Access full-text attached. It is functionally similar to arXiv but partly gets the full-text through pipelines to publishing houses rather than relying on direct upload from authors. Recently, EuropePMC\(^4\) has been developed to extend the bibliographic data and supplement the Open Access corpus for medical and life sciences with added value services such as automatic classification, annotation and linking to further databases. There are many other subject based repositories that will not be further mentioned for the sake of simplicity.

Libraries play an important role for subject based Open Access repositories. But it is more subtle than the one for institutional repositories. arXiv was originally developed in a lab but then brought to Cornell University and is still operated within Cornell University Libraries. RePEc is based on web-servers from academic institutions – often Universities – and those are partly operated by libraries. And PMC is operated by the National Library of Medicine. These examples indicate that libraries play a role as operator but not directly as initiator and driver. Libraries thus provide a sustainable operational basis for subject based initiatives. This is also becoming clear through a counterexample: the Social Science Research Network, SSRN\(^5\). SSRN was not operated by a library and has now been acquired by the publisher Elsevier.

Electronic Thesis and Dissertations

*Service definition: A service that implements local requirements of departments at a given University to store or publish electronic thesis and dissertations.*

Institutional Open Access Repositories may be mentioned alongside services for electronic thesis and dissertations, ETDs. ETD services implement University policies for managing official works in a graduation process, i.e. for registering, retaining and possibly publishing them (Fox et al., 1997). Obviously, ETD services refer to a different document type than the journal articles in institutional repositories. However, ETD services in libraries preceded many institutional repositories, were among the first systematic document servers in libraries and, thus, set standards in the technological implementation of institutional repositories. And they could provide an existing staff contact in the library for supporting open access and repositories. Indeed, many libraries operated both as the same service, which sometimes caused confusion among researchers, who did not want to see their highly valued journal articles among a collection of students’ ETDs. Either a separated display of journal articles and dissertations or a fully comprehensive bibliographic database of

---

\(^3\) https://www.ncbi.nlm.nih.gov/pmc/
\(^4\) https://europepmc.org
\(^5\) https://www.ssrn.com
the whole university, in which ETDs are not the majority, seem to be an established solution for libraries to overcome this problem, today.

Software Development and Deployment

_Service definition: Library staff participating in repository software development and operations of repositories._

Technological support of repositories is distinct from other Open Access activities in libraries (see also chapter 6b). It is often located in departments such as “library technology” or “digital library”. Most of the software used for repositories is open source, notably DSpace, ePrints or Fedora (Crow, 2004) and relies on contributions from the library community. DSpace, for example, has originated from MIT Libraries. In many cases, staff in libraries contributes to the code base in a collaborative fashion so that local use cases can be supported and sustainability of the software is jointly achieved. Also expertise on the local deployment of the software can be increased by active contributions to the code base and supports stability and usability. In other cases, libraries do not contribute to the software development but only focus on local configuration and deployment or even pay companies to deploy their repository. A technologically specialized international community has been built through the conferences of the Open Archives Initiative in Geneva (Harnad, 2001) or the Open Repositories conferences.

The technical _de facto_ standard Open Archives Initiative – Protocol for Metadata Harvesting OAI-PMH (Sompel, Nelson, Lagoze, & Warner, 2004) has been a crystallization point for the technological community even though alternatives that are web-native and support more diverse use-cases have been put forward (Lagoze et al., 2008).

**Open Access Policy Support**

_Service definition: Activities that contribute to the formulation and implementation of policies for Open Access._

Institutional Open Access repositories and institutional Open Access policies have been closely related since the beginning and subject to intensive debate (see also chapter 1c). It has been proclaimed that repositories are the tools for implementing an institutional policy and that a mandate should enforce the incorporation of all institutional publication output in the repository (Harnad, 2011). The policies have been documented in ROAR (see above). Libraries have been involved in formulating Open Access policies for institutions and, in most cases, are responsible for implementation and monitoring. A typical Open Access policy of an institution recommends or mandates to both archive a journal article in a subject based repository or the institutional repository or publish an article in an Open Access journal. Open

---

6 [http://sites.tdl.org/openrepositories/]
Access policy support in libraries may extend to regional, national or international scale. The library community active in DRIVER (Feijen, Horstmann, Manghi, Robinson, & Russell, 2007) and OpenAIRE (Manghi, Manola, Horstmann, & Peters, 2010), for example, can be attributed to having played a vital role in the Open Access policy of the European Commission. A different type of policy support in libraries is checking Open Access policies of journals to identify whether and how which version of a journal article can be registered and shared on a given repository. These journal policies can be intellectually looked up or automatically linked to a given repository through SHERPA/RoMEO\(^7\), which itself has also originated in a library.

Registries and Aggregators

*Service definition: National and international services collecting characteristics and content of a globally distributed repository network.*

Institutional and subject based Open Access repositories as well as Open Access journals are distributed by design and show a distinct fraction of the global scholarly output. They are not meant as a resource for scholars to find publications, for example when they write a paper. The aggregation of all repositories and Open Access journals worldwide, though, allows indeed to search bibliographic information and be referenced to an Open Access version of an article in a given repository. A longstanding effort to achieve this is the Bielefeld Academic Search Engine, BASE (Pieper & Summann, 2006). Between 2004 and 2016 BASE has increased its aggregation to almost 5000 sources worldwide and thereby builds a resource of over 100 million document references. In a similar fashion, GoogleScholar\(^8\) references to repositories and thereby promotes Open Access versions when scholars search for literature, for example, when writing a paper. Aggregators and registries also play an infrastructural role of organizing and standardizing the repository landscape. This becomes particularly obvious in the European network OpenAIRE that issues guidelines for repository operators (Schirrwagen et al., 2013). An earlier and highly influential guideline has been issued by the German Initiative for Networked Information (DINI) that additionally implements a personal validation process for testing guideline compliance\(^9\). Other registries such as ROAR, OpenDOAR and SHERPA/RoMEO were already mentioned. The World Ranking of repositories even evaluates and displays the web impact of repositories (Aguillo, Ortega, Fernández, & Utrilla, 2010). The Directory of Open Access Journals, DOAJ, registers ‘golden’ journals that allow direct publishing of articles in an Open Access mode\(^10\). Libraries play a vital role in the field of aggregators and registries: BASE is operated by a library, OpenDOAR and SHERPA/RoMEO have originated from a library and OpenAIRE is operated by a consortium with many library partners.

\(^7\) http://www.sherpa.ac.uk/romeo/
\(^8\) https://scholar.google.com
\(^9\) https://dini.de/dini-zertifikat/english/
\(^10\) https://doaj.org
Journal Publishing Services

Service definition: Platforms for editors and editorial teams to organize the calls, submissions, review and publishing of complete journals in an Open Access mode.

Due to the massive simplification of the publishing process by the introduction of digital methods, university libraries have started over the last years to establish support for predominantly local editors for publishing complete Open Access journals (see also chapter 6a). The processes involved are different from operating repositories because calls for papers, submissions, review processes, typesetting and publication, including persistent identifiers such as DOIs or URNs (Hilse & Kothe, 2006) have to be implemented and sustainably supported. In many cases, libraries are operating the basic infrastructure, while the journal is driven by a lead editor and an often internationally distributed editorial team. The central role of the editorial manager who is keeping track of all communications, processes and plans is sometimes directly provided by the library but also often located with the local editor because a subject specific education is required to communicate with authors, the editorial team and the subject community. The software Open Journal Systems has become a major player in this field and is used in many initiatives as a platform (Edgar & Willinsky, 2010).

Book Publishing Services

Service definition: Support for the selection, review, typesetting, printing and electronic publication of Open Access books.

Books have not been as central to the Open Access landscape as journal articles. However, libraries are also engaged in Open Access book publishing, particularly with university presses. Not every university press is an Open Access book publisher and not every university press is attached to or operated by a library. But similar to the rise of locally operated Open Access journals (see above) a pattern of university presses (in libraries) that act as Open Access book publishers emerged with the simplifications introduced through electronic publishing. It is now possible to provide economically lean services to distribute an Open Access version of a book digitally and sell the print version on the market. Prints now can be produced on demand with commercial service providers. As for individual journals, it is customary for university presses to have an editorial board for quality assurance and for managing the publishing profile. Presses can be focussed on certain works from select authors at the university or on specialized series including ‘extramural’ authors or provide a means for excellent doctoral dissertations to be formally published as a monograph. A registry of open access books has been established[^1].

[^1]: http://www.doabooks.org
Publication Funds

*Service definition: Management of an institutional budget for (‘golden’) Open Access publications in journals often involving contract management and invoice processing.*

Paying for individual articles before publishing existed before Open Access: journals required so-called page charges by authors for putting graphics, tables or coloured photos in articles. But the idea to charge authors upfront and then offering an article and the whole journal free for download for everybody on the internet is a genuine Open Access phenomenon. Pioneers such as the Nucleic Acids Research\(^{12}\), New Journal of Physics\(^{13}\) or BioMedCentral\(^{14}\), BMC, or the Public Library of Science, PLoS\(^{15}\), introduced this model at a larger scale. Initially based on a direct process between publisher and author, BMC and PloS introduced early institutional memberships that gave discounts to institutional authors and brought libraries, often holders of the membership, into the business. Through the membership, institutional publication funds were introduced, e.g. in Germany by Bielefeld University Library (see also chapter 3b). Systematic funding through funding organisations like the German Research Foundation became available\(^{16}\). Other experiments introduced a ‘buy-out’ of individual articles of an otherwise subscription-based journal, the so-called hybrid Open Access. This triggered a discussion about ‘double dipping’, i.e. the question whether publishers get paid twice: one time through the Open Access fees and a second time by the subscription fees (Pinfield, Salter, & Bath, 2015). Transparent models that linked between both revenue sources were tried to be developed, e.g. by Oxford University Press, but were dropped again. It was the Finch Report and the following UK Open Access policy in 2012 that made hybrid Open Access viable at a larger scale (Horstmann, 2013). Since then, institutional publications funds became a standard in many universities and are mostly managed by libraries. Differences between the policies of individual Open Access funds remain, i.e. whether or not hybrid Open Access is an eligible cost.

Bibliographic and Bibliometric Services

*Service definition: Production and maintenance of a complete institutional bibliography, sometimes involving bibliometric reporting and research analytics.*

Producing an annual institutional bibliography was a traditional task of some libraries. But soon after the introduction of institutional repositories and systematic publication funds, the question about the percentage of Open Access at a given institution became prevalent. Computing the Open Access percentage requires to have an overview of the complete institutional output, equal to an institutional bibliography.

---


\(^{13}\) [http://iopscience.iop.org/journal/1367-2630](http://iopscience.iop.org/journal/1367-2630)

\(^{14}\) [https://www.biomedcentral.com](https://www.biomedcentral.com)

\(^{15}\) [https://www.plos.org](https://www.plos.org)

\(^{16}\) [http://www.dfg.de/foerderung/info_wissenschaft/2010/info_wissenschaft_10_01/index.html](http://www.dfg.de/foerderung/info_wissenschaft/2010/info_wissenschaft_10_01/index.html)
Based on an institutional bibliography, reports about research performance in terms of bibliometrics can be produced and libraries have started to gain importance in these tasks (see also chapter 1e). Within the university, this also moved libraries closer to research administration, which formerly was often the sole manager of institutional research indicators. Today, libraries sometimes even operate the research information system, taking on leadership in a completely new role: as a provider of research analytics (Dempsey, 2012). Genuine library tasks such as subject classification, disambiguation and deduplication now appear in utterly different contexts. One example is the ORCID initiative that provides unique identifiers and thereby means for author disambiguation (Haak, Fenner, Paglione, Pentz, & Ratner, 2012).

Negotiation Services

Service definition: Negotiating contracts for a library or a consortium that contain Open Access components.

With the introduction of electronic journal packages from publishers, the acquisition departments in libraries became contract negotiators. Early Open Access models such as the membership model of BMC (see above) already brought Open Access into the acquisition departments in libraries. Over the years, contracts expanded up to regional or national scale and sometimes contain Open Access components. In Germany for example, the national licensing activities address Open Access\(^\text{17}\). Recently, the Netherlands negotiated contracts with major publishers at a national scale that also include certain Open Access conditions\(^\text{18}\). Programmatic accounts proclaim that these negotiations will result in the total transformation of the market without adding more money to the academic publishing system (Schimmer, Geschuhn, & Vogler, 2015). Initiators and negotiators are usually libraries.

The Open Access Situation in Libraries So Far

Once a provocative idea of a few activists, Open Access has become mainstream over the last 15 years. Accounts of the Open Access percentage vary. But it can be stated that at least a third of the global production of scholarly journal articles is available in Open Access. And there are certainly scholars who write their articles by using solely GoogleScholar and the articles available in Open Access – without using library subscriptions. Open Access has changed the libraries’ role in the academic communication system: from collecting information to connecting articles in repositories and authors to funds. The question whether libraries undermine their own justification of existence is allowed. Simplifying access for scholars to academic publications otherwise not or only laboriously available is a disappearing task for libraries in the end-game of Open Access. A main pillar of the libraries’ service


\(^{18}\) http://www.openaccess.nl/en/news-and-events
profiles over the last decades regarding the electronic era, and over the last centuries regarding the print era is dwindling.

But two obvious points should have become clear in this chapter. First, Open Access may make subscriptions obsolete but it also moves the library profile from focussing on the distribution of scholarly materials to the production of scholarly materials. And this brings academic libraries back to the scholar where academic libraries originated from. It lifts the veil between the library and the scholar that has been silently fallen with the industrialisation of academic publishing and the commodification of electronic access. Second, Open Access can also be seen as a phenomenon of an academic publishing system as a mixed economy. Scholarly communication has become much more complex, requiring libraries to develop an adequately differentiated service portfolio that is customized to smaller scholarly cultures. There are no indications that this process of differentiation and the need for libraries to continuously develop new roles will terminate.

For libraries, one of the subtle but deeply transformative effects of Open Access refers to skills development. Whether directly or indirectly, Open Access introduced a variety of new services and workflows into libraries. The skills’ profiles are variable and differ substantially from traditional acquisitions and cataloguing services – not even mentioning the circulation of printed media. Even in this short chapter, a dozen substantial services can be identified. Some of them represent a further development of rudimentary or traditional library services but many of them are entirely new. They range from computer science to law and economics as well as from data cleaning to transformative strategy. And they may be the basis of conversations between a scholar and a librarian in the local bakery as well as conversations between programmer in Japan and a financial officer in Kenia.

This variety of skills poses the question of curricula for open access in Library and Information Science. Many experts for Open Access in libraries do not have a library education or degree. Will this change in the next decade? The world of libraries keeps changing: disruptive innovations such as SciHub keep shedding new light on Open Access and with Research Data management and Data Analytics new challenges for libraries have long begun.

Literatur


Schimmer, R., Geschuhn, K. K., & Vogler, A. (2015). Disrupting the subscription journals’ business model for the necessary large-scale transformation to open access.

