Report

Open Access Monitoring – Approaches and Perspectives

2-Day-Workshop, 09–10 April 2018, Vienna

As part of the project *Austrian Transition to Open Access (AT2OA)* a 2-day-workshop on Open Access monitoring was organized. The event, which took place on 09–10 April 2018 in Vienna, intended to put a focus on the different international approaches towards this topic, therefore speakers from across Europe were invited.

The following report includes summaries of all the talks held in the course of the conference on the first day, plus links to their presentations, as well as the documentation of the second day of this workshop.

Day 1
Conference

Photo by Tobias Zarka / CC BY 4.0

Open Access Monitoring in The Netherlands

Summary of the presentation given by
Just de Leeuwe (Publishing advisor, TU Delft Library)[1]

Just de Leeuwe, who works as Publishing Advisor at TU Delft Library[2], started his presentation by introducing the audience to the landscape of Dutch universities.

One big advantage in the Netherlands is that the 13 universities have a common infrastructure called NARCIS[3], which aggregates metadata from every single repository. In some cases, the CRIS systems offer repository functionalities which make it easier. The national repository itself is aggregated by OpenAIRE in order to help researchers to fulfill the Open Access Mandate of the European Commission.

The Netherlands have a national strategy to reach 100 % Open Access of peer-reviewed publications. The Road Map to Open Access[4], launched by the Association of Universities in the Netherlands[5] in 2018, declared the goal of 100 % Open Access publishing by 2020. The presenter stressed that these numbers may never be realized, but more importantly, they show the significance of Open Access and describe the holistic approach to Open Science. This bigger framework of Open Access covers Data Management Plans, approaches for opening up non-disclosure agreements, Open Educational Resources, and promotes using Creative Commons licenses.

Photo by Tobias Zarka / CC BY 4.0

Current goals for the time period 2018–2020 are: negotiations with publishers, international collaboration, archiving and monitoring.

The Open Access movement needs to focus on the researchers. So de Leeuwe pointed out the importance of offering incentives for researchers to publish Open Access. One example is the Journal Browser[6], which allows researchers to check if national agreements exist and shows discounts and fees per journal.

---

1   http://www.at2oa.at/en/Just de Leeuwe (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf
2   https://www.tudelft.nl/library/
3   https://www.narcis.nl/
4   https://www.vsnu.nl/Roadmap-open-access-2018-2020-English/assets/vsnu-roadmap-open-access-2018-2020-en.pdf
5   https://www.vsnu.nl/en_GB/
6   https://www.tudelft.nl/en/library/current-topics/libraryfor-researchers/library-for-researchers/publishing-outreach/journal-browser/ (7.5.2018).

The Association of Universities in the Netherlands (VSNU) will initiate an exploratory study of methods of monitoring Open Access publications. The focus is on keeping monitoring simple and accessible for all parties. For the year 2017, a total amount of 6.286 publications have been published Open Access. While Springer, Taylor & Francis, Wiley, Elsevier and Sage cover 5.622 of these, also smaller publishers (including publishers for Humanities) are listed.

The final version of the Definition framework monitoring Open Access[7] was released by VSNU on April 10, 2017, and defines Open Access as: „Peer-reviewed articles, free of charge and permanently available on the website of the publisher or in a trusted repository". The definition distinguishes between „A = Gold, DOAJ classified", „B = Gold, not DOAJ classified OA/Hybrid" or „C = Green only OA". Articles have to be registered in a CRIS system with a publication date till December 31. Papers uploaded on academia.edu, or those that are delayed or under an embargo, are not counted. The monitoring process consists of three steps: input (incl. cleaning and de-duplication), data matching (A-B-C Open Access categorizing), and reporting. To give an example of the measure, de Leeuwe showed the results of TU Delft in the year 2017: Of 3.130 peer-reviewed articles of TU Delft in the year 2017, 51 % were published Open Access. The number of Open Access journals in which TU Delft scientists are publishing is increasing year by year – with a lot of potential for openness.

Report from Sweden

Summary of the presentation given by
Camilla Lindelöw (National Library of Sweden)[8]

As a starting point, Camilla Lindelöw from the National Library of Sweden[9] gave an overview of the current activities in the context of Open Access, which cover governmental goals (which depend on EC mandates), initiatives from funders and universities, Open APC Sweden[10] and Open Access monitoring activities.

The Swedish Research Bill, which was announced in 2016, is currently in force (for 4 years) and stresses „Open Access to research results" and „Open by default – closed as exception". Even though a national Open Access policy has not been released yet, all publicly funded research publications[11] have to be Open Access when published. In other words, this means that the government aims for Gold Open Access without explaining which version a paper can have. What is lacking are incentives and rewards for researchers as well as infrastructure.

The National Library of Sweden has a coordinating role on the national level. One task among many is monitoring the compliancy of Open Access policies and mandates. This working group consists of researchers and research administrators as well as of the speaker herself. They work on recommendations for unifying Open Access demands, propose a common model for monitoring these demands, recommendations for CC licences as a national standard and, since December 2017, they also formulate criteria to judge if publicly funded publications fulfill the FAIR principles[12]. Lindelöw points out that FAIR doesn't mean data has to be open, which has to be discussed.

Swepub[13] works as the national publication database which aggregates from over 40 repositories and is enriched with Open Access DOI data. As an example for monitoring Open Access, including monographs,

7   https://www.vsnu.nl/files/documenten/Domeinen/Onderzoek/Open access/Definitief Definition framework OA_ VSNU-20160217.pdf (7.5.2018).
8   http://www.at2oa.at/en/Camilla Lindeloew (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf
9   http://www.kb.se/english/
10   https://github.com/Kungbib/openapc-se/wiki
11   It is also valid for research data and artistic work which Camilla Lindelöw didn't cover in her talk.
12   https://www.go-fair.org/
13   http://swepub.kb.se/

the speaker referred to the report *Open Access in Sweden 2010–2016*[14], which shows 14 % of the publications were Gold, 22 % were Green. In comparison, the Swedish Foundation for Humanities and Social Sciences[15], which is an independent funder, has an Open Access rate of 60 %, according to a survey[16] undertaken by the project members between 2010–2014 with a return rate of over 50 %. Their strong demand for Open Access leads to generous funding, but still 40 % of the projects are published in a closed way because of unsatisfying publishing opportunities.

Photo by Tobias Zarka / CC BY 4.0

As a third example, Lindelöw talked about the Open APC Sweden project, which was initiated in 2016 and was based on the INTACT project[17]. Since 2018, the National Library of Sweden has the task to monitor and report the total costs of scholarly publishing for higher education institutes (HEI), which includes licences, APCs and administrative expenses. Some Swedish university libraries monitor APC costs and report to Open APC. The challenge is still finding the answers to what to count, which definitions of Open Access work on a national and a European level, and APC in general.

Open Access Monitoring in Finland

Summary of the presentation given by
Jyrki Ilva (The National Library of Finland)[18]

The Open Science and Research Initiative (2014–2017) – a national project primarily focusing on research data – laid the foundations for the move towards open science in Finland.
Parts of this initiative were several ambitious goals regarding the Open Access rate in Finland (75 % OA in 2018, 100 % OA in 2020).

OA monitoring efforts are integrated in the national data collection conducted for the Ministry of Education and Culture – these data have been collected since 2011, and because of the fact that state funding is linked to the number and quality of publication output, the universities are highly motivated to deliver this information.

14   https://openaccess.blogg.kb.se/2017/12/06/open-access-in-sweden-2010-2016/ (7.5.2018)
15   https://rio.jrc.ec.europa.eu/en/organisations/swedish-foundation-humanities-and-social-sciences
16   https://www.rj.se/globalassets/rapporter/2017/rj_rapport_oa.pdf
17   https://www.intact-project.org/
18   http://www.at2oa.at/en/Jyrki Ilva (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf

Currently, there is a discussion whether the Open Access output of universities should become a factor in the funding of universities.

CSC – IT Center for Science is collecting the data from the CRIS-systems of each organization – the data can be uploaded or harvested in an automated data collection process (since 2016).

The publication data is collected (in XML-format) within the Virta publication information service (developed by CSC). After de-duplication, the publication data can be browsed by using a VUFind-based discovery interface (www.juuli.fi) hosted by the National Library.

In the beginning, there were serious issues with the quality of the OA status data, but this has improved.

Now there are two fields to check the OA status in the national data collection:

- One field shows if the publication is Gold or Hybrid OA
- The second field shows if the publication has been deposited in a repository (Green OA)

URLS should be reported for each version (preferably based on DOIs, URNs or Handles).

The question of what qualifies as Open Access and what fits within the different categories is very difficult to

Photo by Tobias Zarka / CC BY 4.0

answer.

Starting in 2016, OA rules were installed – these rules were made as clear as possible (in order to solve this problem). The minimum requirement is that the publication should be free to read („gratis OA"). Gold/Hybrid and Green OA are all accepted.

After two years of collecting data (2016, 2017) using the new guidelines, there are still many questions to answer:

- Identification of OA-Types
- Technical issues with de-duplication (especially with Gold and Green OA)
- Data on embargoes, licences or APC costs are not collected
- Embargoed publications can be reported as Green OA as soon as the final URL is available.

There is a lot of overlap between Gold/Hybrid and Green OA publications in the system --> some Universities systematically collect all their Gold and Hybrid-publications in their repositories (which would qualify them also as Green OA).

The monitoring results for the years 2016 and 2017 show a slight increase in the share of OA articles at Finnish

universities (from 28.9 % in 2016 to 39.9 % in 2017). This is partly caused by improvement in data quality.

New Project launched by the Ministry of Education and Culture (2017–2020): National Research Information Hub[19].

Development of the Virta service will be continued, but the new system will also cover research data and infrastructures. The collection of OA data will also need further development.

Since the creation of the Finnish OA monitoring guidelines in 2015, there were several developments:

- New tools and sources (like Unpaywall) are now available.
- DOI as a fast and reliable way to check the current OA status of a large number of publications.

It is not yet clear how these tools can be used in the best way (either on a national or on a local level within each organization).

Because of the fact that Open Access categories are subject to change, there are new categories emerging or being introduced (e.g. Bronze OA).

Bronze OA includes both delayed OA and the OA publication channels that don't use CC licences – it is not compatible with the current Finnish OA definition (delayed OA is not accepted at all).

This leads to the question if it might be a good idea to use attributes instead of categories (which are constantly evolving). If categories are needed, they could be generated from these attributes. Downside: probably many OA fields per publication.

Some final suggestions / final thoughts:

- Before adding new fields to the data collection requirements, try to figure out if these fields are really necessary/mandatory: each extra piece of information has a cost (e.g. changes to current information systems have to be made).
- When asking organizations to provide information, there should be a strong motivation to do it because otherwise lots of effort might only lead to unused/unprocessed information.
- Ask yourself the question what you are really trying to measure and why.
- Think about using new data sources and automatic tools, but don't do things in a particular way just because it happens to be easy.

Questions:

Incentives: When distributing money, you said you wanted to give more money (as an incentive) to universities which publish a lot of Open Access – how can you be sure that there won't be universities saying that their publications are Open Access when in reality they are not?

Are there any efforts to expand Virta to make it a Europe-wide system (international attempts at monitoring)?

---

19   https://research.fi/

Jisc OA Dashboard

Summary of the presentation given by
Sarah Fahmy (Scholarly Communications Services Manager, Jisc)[20]

In her talk, Sarah Fahmy gave an overview of one specific project that Jisc evaluated in order to provide an impulse to think about how to monitor Open Access (OA) with a strong focus on discoverability of OA.

The OA Dashboard should look at how the costs of OA are monitored on an institutional level in order to get a wider national picture of OA in the UK.

Jisc has access to a large wealth of data sources and services that aim to support OA in the UK. A recent survey shows an 80 % OA-rate for Green OA, which is very positive.

However, OA information is stored across several disconnected systems, each with different baselines and scope. Many of these services have been developed for one specific usecase (or a specific funder's policy), but the goal has to be to bring all the information together.

An OA Dashboard could summarize information from Jisc services (and other open sources) to better inform institutions, funders and Jisc itself on the current status of OA development. It could help the institutions to reduce costs in terms of effort and resources.

The OA Dashboard project was undertaken by Jisc in partnership with Research Consulting, Pleiade Management and Digerati and was divided into three phases:

1. Defining the use cases

2. Prototype development (important: identify the data sources)

3. Business case and reporting (would institutions be interested to sign up to or buy such a service)

Photo by Tobias Zarka / CC BY 4.0

Sarah Fahmy mentioned that initially there were 5 Dashboard options developed in a workshop, but after a feedback round, the number of pilots which will be investigated further was reduced to two.

Dashboard option A should investigate how the authors' update of OA options develops possible indicators that could identify the share of Green, Pure Gold and Hybrid Open Access articles on an institutional and national level.

Dashboard option B should compare the usage and citation rates of OA and non-OA articles. Fahmy

20   http://www.at2oa.at/en/Sarah Fahmy (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf

mentioned that the question asked was if the citations point towards a favorable impact of OA journals and if these statistics could be used as evidence when talking to researchers who are still sceptic about OA publications.

From the aspect of data there is a link between these two dashboards. In order to understand the usage of OA articles, you first need to know what OA is and what it is not. Because of that, Dashboard A would have to been developed first to implement Dashboard B on top.

The data aggregation starts with the identification of the articles which shall be included in the dashboards (on an institutional level). After that, the OA status has to be determined, and in a final step the data on the OA status is combined with usage information. Only free data sources were planned to be used (Crossref, oaDOI/Unpaywall, SHERPA/RoMEO, CORE, JUSP, IRUS etc.).

After talking about this workflow, Fahmy highlighted three different non-commercial data sources (Crossref, oaDOI/Unpaywall and SHERPA/RoMEO), pointing out their benefits and the challenges faced when working with them (see slide number 6 in her presentation[20]).

When working on the prototypes for the dashboards, it soon became clear that in terms of time, money and effort, realization of Dashboard B would be very difficult. Nevertheless, Jisc developed a Dashboard prototype for further decision.

In summary, it has to be said that Dashboard A (determining OA status) would require significant effort to obtain and normalize data from public data. Due to the problem that it was impossible to obtain complete data in a data quality that is useful for the institution, it was not possible to create a business case for this Dashboard. Therefore, Jisc stopped further development. Nevertheless, Sarah Fahmy strongly believes that there is a need for such solutions which might become realizable if the conditions change.

In the next part of her talk, Fahmy spoke about ways to „filter" OA publications out of the entire body of research literature (which turns out to be quite complicated). Services like Unpaywall might be useful, but the biggest issue is that licence information is NOT being applied consistently. This led Fahmy to state that „the numbers are estimates" and „At best, we can say that approx. 50 % of all articles are OA".

OA Discoverability is a domain in which there is a great deal of very dynamic innovation (Unpaywall, OAButton, Kopernio), but there is also more established public infrastructure like OpenAIRE or CORE as well as library services (Primo, Summon, Ebsco, OCLC) or long-standing commercial services (Web of Science, Scopus, Google Scholar). While pure OA journals are generally easy to find (via DOAJ), Hybrid journals are often not easy to identify and harvest automatically. Possible OA publications on personal websites are (in most cases) also very hard to find.

During the final part of her talk, Fahmy mentioned that we should start thinking about a scheme for OA Discoverability. It is essential to identify the relevant actors within the OA Discovery chain (Data Sources, Indexes, Recommenders etc.) and point out the „hot spots" which make the identification of OA publications difficult in order to start working on them. The bottom line of Sarah Fahmys talk was that in the end it is the quality of the metadata which is crucial for the discoverability of OA content.

After the talk, several questions were asked and additional input was given:

- The first question asked was if Jisc had ever thought about letting institutions provide their own data so that Jisc was only doing the calculations?
  According to Sarah Fahmy, this option was discussed but it wasn't very popular. Probably because of the fact that UK institutions have numerous reporting duties and they would see this as just an additional site to which they have to upload their publication data.

- Another input emphasized the problems with OA definitions:
  Instead of talking about definitions what is considered to be OA and trying to find each article that could possibly be OA, it might be an option to define a particular place, and only what is discovered at

that one place should be counted as OA. In this case, institutions would have to make sure that affiliated publications are found there, otherwise they won't get funded.

- Another question raised was if the software behind the Jisc OA Dashboard was going to be released as open source (since Jisc can't build a business case out of it)? Sarah Fahmy replied that this would definitely make sense but hasn't been discussed yet.
- One of the questions was about the CRIS infrastructure in the UK which seems to be quite diverse (a lot of different systems are used).
- Finally, speaker and audience were discussing OA publication culture: especially the questions if researchers who are publishing OA also use more OA material and how happy they are using preprints/final drafts?
  The conclusion of Sarah Fahmy was that there is strong evidence that publishing OA is encouraging citation and usage of OA material. In terms of using final drafts, it seems to depend on the research fields. In many cases, preprints (e.g. arXiv) are sufficient.

Report from Germany & experiences with the OpenAPC workflow

Summary of the presentation given by
Dirk Pieper (Bielefeld University Library)[21]

In his talk, Dirk Pieper from Bielefeld University Library presented an overview of the Open Access workflows and administrative tasks related to Open Access (OA) Article Processing Charges in Germany. The INTACT ("transparent infrastructure for Open Access publication fees") initiative – funded by the German Research Foundation (DFG) from 2015 to 2018, executed by the partner organizations of Bielefeld UL, Max Planck Digital Library and the Institute for Interdisciplinary Studies of Science (I²SoS, Bielefeld U) and supported by the DINI working group for Electronic publishing – has played a pivotal role on the route towards a working framework of APC cost monitoring in Germany.

The INTACT initiative was preceded by the preparatory and complementary framework of ESAC (Efficiency and Standards for Article Charges) that has the development and dissemination of guiding principles and best practices around Gold OA charges as its main incentive (cf. Customer Recommendations for Article Workflows and Services for Offsetting[22] or the organization of workshops, e.g. in 2017, Vienna[23]).

Then he named the initiative O@ analytics[24] located at I2SoS at Bielefeld University that provides bibliometric indicators of Open Access publishing at German universities and major research organizations as a key element on the route towards comprehensibility of OA publishing costs in Germany. Within the project, the dataset enabling the automated matching of Gold OA Journals (ISSN-GOLD-OA[25]) that is used, e.g., in OpenAPC for the distinction of Gold OA and Hybrid journals, has been built.

In the next part of his talk, Dirk Pieper, using data gathered in the O@ analytics initiative, demonstrated that the share of OA publications is relatively low globally and on the European scale. Secondly, he showed that it is countries in South America, South-Eastern Europe, Asia and Africa which proportionally chose to publish OA most. Hence, while global big players like the US, China, the UK or Germany may account for the largest amount of OA publications worldwide, it is in smaller nations – considered to have less developed scientific infrastructure – where the potentialities of OA publishing are realized most clearly and which have

---

21   http://www.at2oa.at/en/Dirk Pieper (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf
22   http://esac-initiative.org/wp-content/uploads/2017/04/ESAC_workflow_recommendations_1st_draft20march2017.pdf
23   http://esac-initiative.org/2nd-offsetting-workshop-9-10-march-2017-vienna/
24   https://www.intact-project.org/oa_analytics/
25   https://pub.uni-bielefeld.de/data/2906347

the biggest shares of OA publications nationally (cf. https://pub.uni-bielefeld.de/publication/2912807[26]).

He finished this first part of his presentation with 3 hypotheses that still need to be discussed and tested in academia:

- *"Growth rate of Gold OA percentages per year between 2008–2016 was too small to meet (national/European/global) political goals of Open Access or Open Science guidelines*
- *Number of Gold OA Journals is too small to reach higher growth rates, which affirms OA2020 approach to transform existing journals*
- *Current Springer Compact agreements increase the number of Open Access articles in hybrid journals (2016–2017), but not enough to turn journals entirely"* (Presentation Pieper)

Photo by Tobias Zarka / CC BY 4.0

In the second part of his talk, Dirk Pieper presented the OpenAPC project that was initiated at Bielefeld University Library in June 2014 with the goals of:

- releasing a dataset on OA article processing costs paid by publicly funded research institutions and on research funded through major research funders;
- finding a way towards reporting on the costs of OA publishing that is more transparent and reproducible throughout institutions;
- providing reliable data on the costs of OA publishing.

The tool OpenAPC was implemented as a centralized infrastructure that largely builds on the Crossref API and is designed to require as little information as possible provided by contributing institutions (see OpenAPC Data Submission Handout[27]). Data is then enriched with identifiers guaranteeing the interconnectivity of stored data with sources for metadata, such as DOAJ, Europe PMC, ISSN or Web of Science. In the OpenAPC project, no separation of data, documentation and reporting code is allowed and data is stored in a common format (CSV, JSON), including an OLAP Server API. The tool is made accessible through GitHub including full documentation and historization of versions. Visualization of data – treemaps[28] – is finally realized through Cubes OLAP Server and accessible through the OpenAPC web-appearance under an Open Database

---

26   https://pub.uni-bielefeld.de/publication/2912807
27   https://github.com/OpenAPC/openapc-de/wiki/Data-Submission-Handout
28   https://treemaps.intact-project.org/

License (ODbL) v1.0[29].

In the last part of his presentation, Dirk Pieper reflected on the roles that INTACT and OpenAPC have played in the transition towards OA publishing in Germany. He described how the initiatives have contributed towards guaranteeing the transparency of the process and fostering the idea of Open Access to all publicly funded research. Even though he had to admit that data provided through OpenAPC was far from being complete, it nonetheless provides the most comprehensive dataset on APC payments by far. Finally, he concluded his talk with his takeaway messages from the successful implantation of an OA cost monitoring tool:

- Nothing is complete, there is a need for trustworthy databases (DOAJ, Crossref, ISSN-L, Scopus, Web of Science, …)
- Bibliometricians like terminated databases for calculations (basic totality of all academic articles cannot be exactly determined)
- Keep workflows as simple as possible for data providing institutions, enrich data with trustworthy identifiers
- Check data quality not manually but automatically
- Measurement of Open Access shares should be contextualized (e.g. Open Access transformation, CRIS, funder policies, new deals and consortial structures)
- APC expenditures can be estimated also for those institutions which don´t have central institutional APC services (WP for the next OpenAPC funding phase)
- Use and demand for publisher reports to counter-check data sources (Crossref, libraries, Web of Science, …)
- Tagging the Open Access status (license information) on the level of articles is unfortunately still a challenge for everyone involved in publication workflows
- Future potential of ORCID for integrating (Open Access) metadata into repositories?
- Data may be incomplete, but we have enough to manage Open Access transformation and to develop a truly Open Access publication system

The possible role of CERIF and CRISs for Open Access Monitoring

Summary of the presentation given by
Pablo de Castro & Jan Dvořák (euroCRIS)[30]

In their talk, Pablo de Castro and Jan Dvorak put forward the idea that local Research Information Systems (RIS), preferably based on the Common European Research Information Format (CERIF), are valid providers of data for and facilitators of Open Access monitoring in first place. In addition, they highlighted one important takeaway of the workshop in their joint presentation: workflows for Open Access monitoring should be made as simple as possible and as accessible as possible for collaborating institutions and individual researchers. Further reliance on metadata exchange standards like CERIF-XML will facilitate the development of successful monitoring tools. Apart from the ready availability of high-quality standardized publication metadata (incl. OA information) in local RIS, the interoperability between these and any monitoring tools as well as publication repositories will be a crucial challenge for RIS's in the nearer future.

Pablo de Castro, from University of Strathclyde, presented the conceptual framework of OA monitoring on a European scale, whereas Jan Dvorak, from Technical University and Charles University of Prague, highlighted

---

29   https://opendatacommons.org/licenses/odbl/1.0/
30   http://www.at2oa.at/en/Pablo De Castro & Jan Dvorak (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf

the strengths of CERIF and CERIF-XML for the interoperability of Research Information Infrastructures. Both agreed that a large number of infrastructures that store information on research publications exist in the academic landscape, but that we lack robust common standards for reporting publication outputs and OA publishing on a trans-national level. From their point of view, a commonly used metadata standard – like CERIF – and a standard for the inter-institutional exchange of data – like CERIF-XML – can be considered as optimal frameworks for the consolidation of national as well as international monitoring of research outputs.

Pablo de Castro provided us with some examples of best practice, where data from heterogeneous local CRIS – e.g. the P-O-P-F project in the Netherlands – or from a national CRIS – like CRISTIN in Norway – can provide valid and valuable data for OA monitoring.

Photo by Tobias Zarka / CC BY 4.0

Jan Dvorak then gave a short overview of the huge potentialities of CERIF as a metadata standard within the ecosystem of academic publishing, in which he put the individual researcher into the center of interest. One, if not the major, strong point of CERIF is that it allows metadata on research publications to be flexibly augmented and re-used in different contexts.

Besides the reflections voiced during the workshop, Pablo de Castro shared his thoughts on Open Access monitoring in two recent blog posts from the euroCRIS.org-Blog: part 1[31] and part 2[32].

---

31   https://www.eurocris.org/blog/few-thoughts-oa-monitoring-and-criss-i
32   https://www.eurocris.org/blog/few-thoughts-oa-monitoring-and-criss-ii

Open Access Monitoring – State of the discussion in Austria

Summary of the presentation given by
Patrick Danowski (AT2OA)[33]

Patrick introduced a concept to model different Open Access types beyond the Open Access colors. After the workshop, he published the idea in a paper on Zenodo[34].

Photo by Tobias Zarka / CC BY 4.0

Outlook/Teaser/Preview – Day 2

On the day after the AT2OA-Workshop „Open Access Monitoring – Approaches and Perspectives", the organizing committee invited all the speakers from day 1 as well as members of AT2OA Working Group to collect ideas for a prospective Open Access monitoring in Austria. For this purpose, this second day was organized in a World Café format where first two main themes, „Metadata/Sources" & „Processes", were discussed in smaller groups to jointly draft a best practice-model for Austria in the end.

---

33   http://www.at2oa.at/en/Patrick Danowski (AT2OA-OA-Monitoring-Workshop, 2018 04 09).pdf
34   http://dx.doi.org/10.5281/zenodo.1244154

Day 2
Workshop

Photo by Tobias Zarka / CC BY 4.0

Introduction

On April 9th and 10th, a workshop on "Approaches and Perspectives" in Open Access monitoring was held in the context of the Austrian HRSM project AT2OA in Vienna. While the first day of the workshop was open to the public, the second part was reserved for discussions between European specialists and Austrian stakeholders in Open Access monitoring. This second part of the workshop, permitting in-depth reflection of best-practice models of Open Access monitoring, revolved around two guiding questions that were discussed in two separate parallel groups. The groups of discussants consisted of the invited presenters of the first part of the workshop and members of the AT2OA project. Even though the two discussions were held in parallel, all participants had the opportunity to join and participate actively in both groups. While the one group focused its discussion on infrastructures and processes that need to be set at national as well as institutional level in order to facilitate a successful national OA monitoring, the other group discussed viable sources for publication metadata within OA monitoring and reflected on the level of granularity necessary to ensure successful monitoring.

Topic: Processes

In the session dedicated to the identification and definition of processes and workflows related to Open Access monitoring, the discussants identified 6 main aspects that are promising and could be beneficial for a successful launch of a national Open Access monitoring initiative:

1. The first question was concerned with how data can be gathered in order to develop and operate a comprehensive and successful national OA monitoring tool. Basically, two different approaches were identified. Firstly, a bottom-up approach, where individual institutions guarantee the provision of complete publication metadata. A second – rather top-down – approach involves the utilization of publication metadata registered in multidisciplinary citation databases like Web of Science (WoS) or Scopus. Such data would then need to be aggregated on an institutional level. While in the first approach individual institutions inevitably need to take part actively in the monitoring exercise, the second approach would allow for an exclusive central service for the curation of publication metadata.

   The decision whether the overarching strategy of metadata collection requires institutions or individual researchers to contribute actively to the generation of data inevitably entails implications regarding the quality of metadata within a national monitoring project. A solution based on metadata provision through internal reporting systems (i.e., Research Information and Management Systems = RIMS) would facilitate a more complete depiction of the entire publication output of an institution. Solutions based on standardized bibliographic data sources would provide a more consistent and intercomparable set of data. Hence, a combination of both approaches was regarded as optimum by the discussants, as web-based bibliographic resources , e.g. WoS, Scopus or Crossref, could then be used, firstly to normalize metadata provided by individual institutions and secondly to act as reference for the completeness of the compiled metadata.

2. The collection of metadata in order to back up a monitoring tool for the transition towards open approaches of academic publishing comes along with a set of complexities that need to be taken into account from the beginning. First of all, data in a national monitoring system needs to be standardized and intercomparable regardless of its source or provenience. Thus, mechanisms enabling the deduplication and matching of metadata in participating institutions need to be set in place. This service could be achieved by a central deduplication and matching facility, like already established in Sweden. But still, it needs to be pointed out that such services rely heavily on existing and well-accepted unique digital identifiers, like DOI

or (E-)ISSN, as identification and deduplication of non-standardized publication metadata stemming from multiple sources is not only extremely time-consuming but also tedious and at times ambivalent. Hence, optimum functionality of any monitoring approach of academic publication output can only be guaranteed by the provision of unique digital identifiers at the level of individual records and institutions/contributors.

Thus, before starting the implementation of an Open Access monitoring tool, a best practice model – able to provide publication metadata via a standardized interface that is both complete and consistent – should be determined. Nevertheless, a task group engaged in the development of a monitoring strategy should consider that not all participants of the initiative might be able to deliver metadata at the same level of completeness and standardization. Hence, a situation of maximum disparity in data quality among individual suppliers of data should be defined as a worst-case scenario.

3. One major issue when designing a comprehensive and reliable Open Access monitoring tool is completeness of reported data. As researchers are often forced to report their academic output via multiple systems, grasping the whole picture of academic publishing at national level or for a distinct institution can be problematic. Further, researchers often lack the motivation to register their complete publication output in a RIMS/CRIS, as such administrative tasks are often perceived as cumbersome and overly time consuming. Hence, in order to improve the level of integrity, quality and reliability of publication metadata, it would be sensible to develop incentives for the complete documentation of research output.

   In any project relying on external data sources, the quality of the service is related to the accuracy, completeness and interoperability of the transferred source data. Hence, it will be crucial for the success of an Austrian national Open Access monitoring initiative to collect information on existing institutional RIMS/CRIS and to assess critically how these important sources of publication metadata on institutional level can be integrated into a national monitoring tool.

4. Another important aspect that needs to be taken into account when designing a national Open Access monitoring tool is fractionalization of record counts. Generally, the choice is between two viable options: fractionalization by institution or fractionalization by disciplinary subject categories. The experience of the German OA monitoring initiative has shown that fractionalization by institutions is more promising than fractionalization by subject categories.

5. Another issue related to the design and implementation of a national Open Access monitoring initiative in Austria are external sources of data that could and can ease the normalization of publication metadata. First and foremost, the availability of DOIs was pointed out by all discussants. Further resources like Ringgold[35] (Institutional denomination) and ORCID[36] (personal profiles) were discussed as potentially rewarding sources of metadata related to publication output. The discussants identified these resources as valuable facilitators, not only for the enrichment of metadata, but also for the correct attribution of authors and collaborators to institutions and hence for the fractionalization of research output.

6. A last, yet still viable, option that was discussed in the group concerned with the strategic outline of an Austrian national Open Access monitoring strategy focused on the adoption of existing discovery and monitoring services like *unpaywall.org*. A similar discussion regarding the most promising existing resources for the discovery and traceability of Open Access scholarly literature has recently been started in the UK. So far, unpaywall.org was identified as the most advanced tool, but stakeholders and experts on an international level still need to evaluate whether such services can be adapted to serve as adequate monitoring tools in the future.

---

35   https://www.ringgold.com/
36   https://orcid.org/

Topic: Sources

The session on Metadata for OA monitoring and its most promising sources was concluded with the maybe most important take-away message of the entire workshop: "*Keep it as simple as possible*". This straightforward assertion should remind us, as researchers, evaluators of and within research, research managers, librarians, CRIS managers or repository managers that the environment of academia – that we work in and that we try to describe analytically at the same time – is diverse and multiple. Hence, when conceiving a project aiming at showcasing the successful transformation of academic publishing towards practices of Open Access publishing, we first of all need to scrutinize the actually important bits of information necessary to accomplish this arduous endeavor thoroughly. Secondly, at the beginning of a project aiming at the demonstration of a transition to OA, all stakeholders should be aware that presentation of results should be publicly available in a timely manner, as readily available information on the state of the project has proven crucial for the success of processes of transformation. Enriched publication metadata stored in a central registry and presented through a highly accessible platform can here be regarded as one of the most powerful facilitators and drivers of narratives of successful transformation. Consequently, those responsible for the monitoring of the transition process must be aware of the quality of registered metadata from the very start of the project. The two most prominent issues are: how much centralized editing of metadata is necessary in order to guarantee the due completeness and accuracy of data; and, what is the depth and breadth of metadata that individual/ institutional sources of data like local RIMS or institutional repositories are able to provide at high quality. The guiding rationale of the project – keeping metadata standards as simple and unpretentious as possible, but also as accurate and complete as necessary at the same time – should lead to a monitoring tool that is timely available and easily accessible by stakeholders in academic publishing.

The second most important reminder of the session consisted in ascertaining that successful monitoring tools need to be sustainably and independently (free of any commercial interests) funded in order to be able to provide a mid- and long-term resource for the temporal analysis of the process of transition of academic publishing towards Open Access practices.

Apart from these fundamental considerations, the participants of the workshop discussed the importance of discerning the scope of data collection when defining minimal requirements of metadata depth. First of all, it should be clear from the very start of a monitoring project how results are to be reported / presented. Secondly, a national monitoring project needs to take into account the diverse institutional and technological situation of data suppliers. Thus, different metadata standards as well as the technological development status of systems – like institutional RIMS and institutional repositories – need to be taken into account when defining a common minimal set of metadata for Open Access monitoring. This, first and foremost, because these decentral institutional infrastructures are the most important sources of exhaustive information on academic publications produced in scientific institutions at a national level. As not all institutions are running systems based on a common technological infrastructure in Austria and the status of technological development and refinement is diverse across institutions, it is crucial to account for this situation and define metadata standards with the lowest technological barriers for data providers / sources, in a manner that interaction between institutional and national monitoring infrastructure(-s) is possible to a maximum while maintaining the highest possible quality standards. In a nutshell, when it comes to defining common metadata standards for Open Access monitoring, joint agreement on standards and extent of data between partners is more important than the sophistication of the metadata standard, both in terms of depth and format of metadata exchange and storage. Hence, metadata exchange protocols should provide the possibility to exchange data in very basic (flat) data formats, e.g., csv. For example, the Finnish system of Open Access monitoring provides the opportunity to ingest data via .csv files – then transformed by a data curation team into the specified XML format – if institutions do not run a technical environment able to connect via an API.

Another important consideration when defining common metadata standards are chaperoning infrastructures facilitating the parsing of data issuing from multiple sources, e.g., RIOXX[37] in the UK.

In addition to the discussion on minimal requirements of metadata standards, the participants of the workshop discussed the most promising sources of information regarding Open Access status and costs (e.g. unpaywall or Open Access button[38]), affiliation data (ORCID), national bibliographies, bibliographic databases, journal metadata and journal ratings (e.g. ERIHplus[39]). While all of these specialized sources of information around academic publishing provide valuable information for a national Open Access monitoring infrastructure, their funding models and sustainability need to be taken into account and assessed critically. And even more importantly, if we, as stakeholders within the ecosystem of academic publishing, want data providers to be reliable and sustainable, public academic institutions as well as commercial publishers and data providers need to take care of and guarantee the sustainable funding of infrastructures collecting and distributing data on Open Access scholarly literature.

Recommendations

Keep it as simple as possible!

In the early stage of the deployment of a national Open Access monitoring strategy, it is crucial not to be overambitious and to keep the scope of the project as simple and unpretentious as possible. Hence, a strategy only covering publication data that

1. can be identified unambiguously via a DOI;
2. has been published in a journal;
3. is peer-reviewed;
4. is not a contribution to a conference

has been consented on as a viable option in the early stage of development.

Further, for the success of a monitoring tool, it is crucial to design a system that is purpose-built and based on existing data sources and adjacent services. A national Open Access monitoring system should not try to satisfy other needs or tasks than its focal functionality or to position itself as a supplementary system in regard to existing infrastructure. Still, such a system could be expanded with supplementary modules (e.g. scientometric overlay) at a later stage of development. Further, it was highlighted throughout the discussion that it is not only an appropriate strategy to separate the monitoring of pervasion of Open publication strategies from Open Access cost monitoring, but that a distinct separation of these two aspects can be instrumental for the overall success of monitoring exercises.

In an early phase of operationalization of an Open Access monitoring tool, it is in any case favorable to start the implementation within a pilot project where only a restricted number of institutions are involved in order to keep the level of complexity as low as possible. The scope of the project should only be eytended to cover the entire national landscape at a later stage of development.

Joint agreement on standards and extent of data between partners is more important than the sophistication of the metadata standard, both in terms of depth and format of metadata exchange and storage. Thus, it will prove productive for any monitoring strategy not to engage in definatory work on Open Access categorizations (e.g. so-called Open Access Colors) and to content itself with describing access, licensing or re-use conditions with highest accuracy instead, in order not to lose momentum due to a war of opinions between different

---

37   http://rioxx.net/

38   https://openaccessbutton.org/

39   https://dbh.nsd.uib.no/publiseringskanaler/erihplus/

approaches towards Open Access strategies.

During any phase of the project, the team in charge of the development of an Austrian national Open Access monitoring tool should consider and reflect on ongoing monitoring and reporting initiatives alongside the major Austrian funding agencies (like the Austrian Science Fund).

For the purpose of data aggregation via external sources of metadata, the OAI-PMH protocol has been identified as the best choice. Technical interfaces should be designed in accordance with the OAI-PMH protocol, as this facilitates the interoperability with well-accepted publication harvesters and discovery tools like OpenAIRE. Offeríng an ingest interface that is compliant with the OAI-PMH protocol would also allow a more flawless integration of an Open Access monitoring tool in the landscape of existing RIMS.

From an Austrian perspective, it would probably also be a viable option to refrain from developing a custom reporting tool and to build a monitoring strategy on existing systems instead, similar to the national OA monitoring system in Finland, or to re-use codes developed and made available in other projects.

As in any academic domain, readily available funding for the development and sustainable operation of Open Access monitoring tools and sources of metadata is scarce. Thus, all stakeholders in the ecosystem – scientific institutions, libraries, fund raisers, publishers – should, from the very start of a national Open Access monitoring program, be aware of and make sure that not only the project, but also all services integrated in the landscape of the monitoring project are resourced with funds that enable a sustainable operation.

Strategic Outlook

After two days of interesting input and insightful presentations, the guiding rationale for organizing a workshop on Open Access monitoring in Austria remains unsolved: how should an Open Access monitoring Tool for Austria be conceptualized and operationalized in order to serve as a comprehensive instrument for decision making and reporting success of the transition towards Open academic publication strategies?

Since the uprise of Open Access publishing models, we have known multiple options and categorizations of Open Access. While these have been intensively and extensively debated in scholarly literature and in political debates, until today the definitions of Open Access have not been stabilized and still are object to ongoing negotiation. In the aftermath of the workshop, we thus developed a first reference model for a classification of Open Access literature beyond the often contested color based categorization. The theoretical model for the Classification of Open Access Tuples (COAT) has been published as a proof of concept as *R Code* on GitHub[40]. Nevertheless it needs to be pointed out that this proof of concept – through extensive testing – needs to be refined, especially in order to showcase its performance for larger and big data sets. Thus, the next steps consist in finding pilot institutions which are willing to provide their data and participate in the testing of the proposed infrastructure. The final goal of this process of refining and testing existing bits of code and infrastructure is to demonstrate the possibility of integrating data issuing from multiple institutions and environments into one common operable monitoring tool.

Further, we decided to take the expert's advice – keep it as simple as possible – seriously and decided to focus on publication metadata that includes a unique Digital Identifier (i.e. DOI) in an initial step. So far it has not been possible to take a decision whether it would be preferable to focus on specific document types (e.g. only considering research articles). This decision needs to be postponed to a later stage of the project in which the functional and technical operability of the proposed tool has already been demonstrated.

Finally, we aim at keeping up the communication regarding most recent developments with other countries which are working on Open Access monitoring approaches – e.g. Germany and Switzerland – for the moment and strive to develop a vivid platform for international exchange. As OpenAIRE has been identified as an important and influential player regarding the situation in Europe from all participants of the workshop in Vienna, from our perspective, OpenAIRE could be the hub for further discussion and probably take a key role in establishing a European network for the exchange of ideas and novel approaches in Open Access monitoring.

---

40   https://github.com/patrickda/COAT