

Panel Discussion / CRIS2024 Conference Vienna (TU Wien, May 16th, 2024)

CRIS Data: Potential and Challenges for Open Access Monitoring and Negotiations

Chair: Ursula Ulrych (AT2OA², University of Vienna & University of Veterinary Medicine, Vienna)

Scientific Committee / Panellists: Patrick Danowski (ISTA, AT2OA², Austria), Natalia Manola (OpenAIRE), Arjan Schalken (UKB, Dutch University Libraries Network), Joachim Schöpfel (University of Lille/France)

Patrick Danowski: Patrick Danowski is manager of the library at the institute of Science and Technology Austria (ISTA). He studied Computer Science and did a traineeship for academic librarians at ZLB combined with a Master Study at Humboldt University (MA LIS). He is active in IFLA and the national library association of Austria (VÖB). In the Austria Transition to Open Access 2 project (AT2OA2) he is co-lead of the part project for the development of the Austrian Datahub for Open Access Negotiations and Monitoring. <https://orcid.org/0000-0002-6026-4409>

Natalia Manola: Natalia Manola is the CEO of [OpenAIRE](#), a pan European e-Infrastructure supporting scholarly communication and open science Europe since 2009. She is also a research associate in “Athena” Research and Innovation Center and in the University of Athens, Department of Informatics & Telecommunications, and a co-founder of an Athena RC spinoff on AI driven policy and business consulting. She holds a Physics degree from the University of Athens, and an MS in Electrical and Computing Engineering from the University of Wisconsin at Madison. She has several years of employment as a Software Engineer and Architect employed in the Bioinformatics commercial sector. She has expertise in Open Science policies and implementation, having served in the EOSC Executive Board (2018-20, chairing the Working Group on Skills and Training) and in the Open Science Policy Platform (2016-18), an EC High Level Advisory Group to provide advice about the development and implementation of open science policy in Europe. Her research interests include the topics of e-Infrastructures development and management, scientific data management, data curation and validation, text and data mining and research analytics. Natalia has also served in the EC Future Emerging Technology (FET) Advisory Group (2013-2017).

Arjan Schalken: Arjan Schalken is Program Manager UKBsis at UKB, the network of Dutch University Libraries and the Royal Library. His work focusses on the realization and implementation of sustainable Open Access strategies for the Dutch consortium. Arjan advises and supports UKB in negotiating and managing Publish & Read agreements and he is also national project manager for implementing consortium agreements with full Open Access publishers. As project manager 'You Share, We take Care', Arjan helped universities to implement article 25fa of the Dutch Copyright Law, supporting researchers to share their research publications in the institutional repositories more easily.

An important result of the program UKBsis is a data warehouse to monitor, manage and analyze publication related data as input for Open Access strategy, negotiations, contract fulfillment and tracking the open access life cycle of publications.

Arjan works as UKB Program Manager since 2020. From 2017 – 2020 he was deputy director of the Free University Amsterdam Library.

Joachim Schöpfel: Joachim Schöpfel is associate professor in information and communication sciences at the University of Lille (France) and member of the GERiCO research laboratory. His research interests are in the field of scientific and technical information, academic publishing, research assessment, and library development. His latest research projects are on the usage of the HAL repositories by French research laboratories and on the impact of open science on research assessment, including ethical issues. He is

consultant in scientific communication with two consulting offices (Ourouk, scidecode).
<https://orcid.org/0000-0002-4000-807X>

Ursula Ulrych: Ursula Ulrych is project coordinator of Austrian Transition to Open Access Two (AT2OA²) at the University of Vienna and library deputy head at the University of Veterinary Medicine, Vienna. She studied Microbiology and successfully finished the postgraduate programme Library and Information Studies, both at the University of Vienna. After gaining expertise in international collaboration and project management at a biotech company, she was part of the Bibliometrics and Publication Strategies Team at the University of Vienna from 2015 to 2022, where she additionally held the position of deputy head from 2020 to 2022. Her current interests focus on the multifaceted changes and the transformation in scientific communication.

<https://orcid.org/0000-0002-0549-2716>

Abstract:

Open Access (OA) poses some challenges, especially in terms of monitoring its development and progress, and negotiating with publishers and other stakeholders to ensure fair and sustainable OA policies and practices. In this context, Current Research Information Systems (CRIS) can play a crucial role, as these systems collect and manage information and detailed metadata about the research activities and outputs of institutions.

Our Discussion aims to explore the potentials and challenges of using CRIS data for OA monitoring and negotiations, and to present some existing approaches and initiatives in this field. It also proposes some questions and topics for further discussion among the participants.

OA monitoring is the process of collecting, analyzing and reporting data on the state and trends of OA development, at institutional and national level. Among other reasons, OA monitoring is needed for evaluating the impact and effectiveness of OA policies and initiatives, and to identify gaps and opportunities for improvement and decision making. Also it can support increasing the transparency and accountability of research, and to ensure the quality and reliability of OA data and information.

Using CRIS data for OA monitoring entails some challenges, such as:

- The quality and completeness of CRIS data may vary across institutions and countries, depending on the standards, practices and systems used for data collection and management.
- The interoperability and integration of CRIS data with other data sources may require technical and semantic solutions, such as common identifiers, formats, protocols and vocabularies.
- The ownership and openness of CRIS data may depend on the legal and contractual agreements between institutions and data providers, and may raise issues of privacy, security and ethics.
- The trust and acceptance of CRIS data among the research community and other stakeholders may depend on the credibility and reputation of the data providers, and may require mechanisms of verification and validation.

There are different approaches to gather data for OA monitoring, depending on the level of centralization and coordination involved. Two main approaches can be distinguished:

Top-down approach: This approach involves the reuse of existing systems and services that provide OA data and information at the national or international level, such as Web of Science, Dimensions, Scopus, Open Alex etc. These systems and services aggregate and analyze data from various sources, and offer indicators and reports on OA development and performance.

The main advantages of this approach are that it requires only a few players to coordinate with, and that it offers a ready-made and standardized solution. The main disadvantages are that the database is only as complete and accurate as the source system (especially humanities and publications without DOIs are missing or under-represented in most of the commercial systems), and the ownership and control of the data lies with an external company, which may limit the access and reuse of the data.

Bottom-up approach: This approach involves the collection and integration of data from individual e.g. CRIS systems at the institutional level, and the creation of a common database and service that provides OA data and information at the national or international level. This approach requires the collaboration and coordination of multiple institutions and stakeholders, and the development of technical and organizational solutions to ensure the interoperability and quality of the data.

The main advantage of this approach is that the CRIS database is more complete and comprehensive because of the reporting function of these systems for other stakeholders. The main disadvantage is that it can require more time and resources, and that it may face challenges of harmonization and standardization of the data.

Examples will be the following implementations:

[Austria/AT2OA²](#): National Datahub - This initiative is funded by the Austrian Federal Ministry of Education, Science and Research (BMBWF) and is part of the project AT2OA2. It aims to create a national datahub that collects and integrates data on OA publications from various CRIS systems from project partners. The CRIS landscape in Austria is quite diverse from PURE to inhouse-development. The data is enriched using different sources via APIs (e.g. unpaywall, crossref, OpenAlex).

[UKB/ The Netherlands](#): Data service - UKB, the network of Dutch University libraries, built a data warehouse which delivers data services for their consortium and libraries. The datahub contains a rich, structured and controlled metadata set for peer reviewed articles written by Dutch authors inside and outside publishing OA agreements. Besides data from publishers, university CRIS systems and commercial databases, the data warehouse also harvests open databases like Crossref, Unpaywall and OpenAPC.

[OpenAIRE](#): European approach - OpenAire collects and aggregates data from CRIS systems and repositories, which are OpenAire compliant since 2009. The collected data can be enriched in the OpenAireGraph and presented in a monitor that has to be configured for institutional and national views as a paid service. In addition to this, they host a working group for the exchange of European Open Science monitoring initiatives. The [Irish OA Monitor](#) will be presented as a current practical example.

[HAL / France](#): Archive and platform - HAL is a multidisciplinary open archive for the sharing of published and unpublished research. It is at the service of researchers in academic institutions, both public and private. In France, HAL is the national archive chosen by the scientific and academic community for the open dissemination of its research results. Launched in 2001 as a multidisciplinary open archive, HAL has become a platform with more than 140 institutional repositories of universities and research organizations, with thousands of collections of research laboratories, institutes, and projects, and with an increasing number of functionalities and features similar to research information management systems, for individual and institutional assessment.