

# Roadmap to FAIR Research Information in Open Infrastructures

## Christian Hauschke

German National Library of Science and Technology, Hannover, Germany

[christian.hauschke@tib.eu](mailto:christian.hauschke@tib.eu), [orcid.org/0000-0003-2499-7741](https://orcid.org/0000-0003-2499-7741)

## Serhii Nazarovets

State Scientific and Technical Library of Ukraine, Kyiv, Ukraine

[serhii.nazarovets@gmail.com](mailto:serhii.nazarovets@gmail.com), [orcid.org/0000-0002-5067-4498](https://orcid.org/0000-0002-5067-4498)

## Franziska Altemeier

German National Library of Science and Technology, Hannover, Germany

[franziska.altemeier@gmail.com](mailto:franziska.altemeier@gmail.com), [orcid.org/0000-0001-7086-6211](https://orcid.org/0000-0001-7086-6211)

## Nataliia Kaliuzhna

State Scientific and Technical Library of Ukraine, Kyiv, Ukraine

[natali.kalyuzhna91@gmail.com](mailto:natali.kalyuzhna91@gmail.com), [orcid.org/0000-0003-3154-8194](https://orcid.org/0000-0003-3154-8194)

## Abstract

The FAIR Principles were designed to improve the findability, accessibility, interoperability and reusability of data holdings by humans and machines. The principles can be applied to research information too. We present the results of the discussions that took place during the series of online workshops with experts on Research Information and FAIR Guiding Principles. We provide high-level criteria on how to foster findable, accessible, interoperable and reusable, and we hope that our roadmap for FAIR research information in open infrastructures bring many benefits to a diverse group of stakeholders of the scientific ecosystem.

**Keywords:** FAIR Principles; Research infrastructures; Open Infrastructures; interoperability; reproducible research

## 1. Introduction

There is an increasing need and growing importance of research information for different use cases in the scholarly world. Funding organizations are aiming at getting more information about the outcomes of the research they financed, research institutions want to have well-founded information for evidence-based decisions on their institutional governance, researchers use the information to find works in the area of interest, to discover potential collaborators or experts in a particular scientific domain. Moreover, career decisions are being made based on scientometric information. The use cases of research information in the scholarly world include, but are not limited to, controlling, reporting, research analytics, and research intelligence.

Furthermore, the FAIR movement (Wilkinson et al., 2016) gains momentum. FAIR refers to findability, accessibility, interoperability, and reusability of research data. However, these principles can be applied to research information, too. These two developments converge in the area of research evaluation and its underlying information base. Initiatives that call for open and transparent data in research evaluation and bibliometrics are getting traction. The most prominent examples are the Leiden Manifesto (Hicks et al., 2015), the Metric Tide (Wilsdon et al., 2015), and the San Francisco Declaration on Open Research Analytics (Alberts, 2013). This leads to a need for an ecosystem of FAIR data in an academy-owned and open infrastructure (Open Data, Open Source, Open Standards etc.).

There are some differences between research data and research information. The production of research data is often very expensive, and can often times not be repeated. It seems obvious to publish it in a reusable way, but this is often associated with considerable technical, legal or organizational obstacles. The FAIR principles aim at overcoming these hurdles to make research data useful to a wider audience. They were established in 2014 in a workshop in Leiden, and the concept of research data ideally being Findable, Accessible, Interoperable and Reusable (FAIR) received notable uptake by the scientific community.

Research information is a distinct term and can be as metadata about research activities (Science Europe, 2016) or any information that describes the research output as well as the context in which research is being conducted (Jeffery et al., 2014). Another term, that is often times used interchangeably, is “scholarly metadata” (Gregg et al., 2019). Naturally, research information includes various elements such as actors involved in the research process (i.e. researchers, research administrators, programmers and technical team personnel which often support projects etc.), multiple types of research outputs (i.e. publications, patents, dissertations, data sets, software etc.), organizations, organizations’ departments, information on the workflows, equipment, facilities, research awards, honours and events related to the research etc. Traditionally, in order to maintain a wide array of research information consistent, well-structured and up to date Current Research Information Systems (CRIS) have been developed. Over the past years, some countries like Norway (CRISin), the Netherlands

(NARCIS), the Czech Republic (IS VaVaI), Slovenia (SICRIS) and India (IRINS) have successfully implemented national CRIS systems (Blümel et al., 2014; Castro, 2020). In other countries like Finland, Peru, Croatia and Ukraine national CRIS systems have been under construction. In the special case of science studies research information can also be the subject of research and then takes on the character of research data. However, the terms should be clearly distinguished.

It is important to mention that original FAIR principles emphasizes the significance of metadata in turning data FAIR where metadata is seen as any description of a resource that can serve the purpose of enabling findability and/or reusability and/or interpretation and/or assessment of that resource (Jacobsen et al., 2020). Metadata is an umbrella term that refers to all data about data (Boeckhout et al., 2018) and in fact is a constituent of research information. In our study research information treated as a sufficient digital object which must be FAIR in its own right.

## **2. Related works**

Most notable when discussing FAIR research information is the 14th International Conference on Current Research Information Systems (CRIS2018) with the conference theme “FAIRness of Research Information”. Several contributions dealt with this topic, for example Ivanovic et al. (2019) who addressed the issue of making research information accessible across language boundaries. Engelman et al. (2019) describe their approach of building a FAIR electronic archive using the CERIF data model. Guillaumet et al. suggest including the FAIR guiding principles in the analytics processes of research information.

Other contributions to the field include Peroni & Shotton (2020), who introduce their take on FAIR scholarly metadata in regard to citation metadata. Hagemann-Wilholt et al. (2019) explain, how the FAIR guiding principles align to conference metadata.

## **3. Method**

To gather information about the application of the FAIR principles on research information, we decided to conduct a series of workshops, which – due to the Corona pandemic – had to be held virtually. The participants were selected to represent different areas of expertise about FAIR principles and about research information. We were able to include the perspective of information scientists, librarians, PID providers, researchers, research funders, research information experts, and scientometricians. In total 24 experts from 23 institutions participated in the workshops. We held one workshop for each principle of FAIR. The structure of the workshops was as follows:

1. Introduction of the participants and their expertise in the topic

2. Presentation by an input speaker (an expert on the respective criteria)
3. Round of statements by the participants
4. Open discussion
5. Wrap-up and closing

The workshops were held via the virtual meeting software Cisco WebEx. The participants agreed beforehand in the recording of the workshop. The recordings were transcribed. A summary of the transcript of each workshop was sent to the participants to create another opportunity for feedback, and for quality assurance purposes. An additional means of input for the experts was one Google Document for each workshop where all participants could collaboratively write down their thoughts and questions on the topic.

In preparation for the workshops an email was sent to all participants with an outline of what will be the scope of the meeting. This included a list of questions that we strived to answer. The following list shows the questions from the interoperability as an example. It was sent out for each of the FAIR principles.

1. How can “interoperability of research information” be defined?
2. What are the criteria/requirements to realize interoperability?
3. Who are the stakeholders?
4. What might be barriers?
5. What are the elements that need to be agreed, standardized and implemented by research communities to support RI interoperability?
6. What services, tools that currently exist in the research ecosystem are capable of providing FAIR research information?
7. What are best-practices studies for supporting FAIR research information?

In accordance with the purpose of this work, the principles of FAIR were repeated, rethought, interpreted for use in research infrastructures (RI) of the two countries. The workshops took place between September 2020 and January 2021.

#### **4. Results**

In the following we will outline the results of the workshops in four sections, one for each of the four FAIR principles. The order corresponds to the FAIR principles, not to that of the workshops. Literature cited in this section is part of the results, as it was mentioned or posted during the workshops, in the slides, in the meeting notes or in the chat function of the video-conferencing software.

#### ***4.1. Findability***

It was stated that in order to find something you need to be able to index and discover it. This means that metadata is the key to findability. The discussion on this premise led into different alleys.

Persistent identifiers were a highly debated topic. To cover the different aspects of research information, there is a need for a variety of PIDs, which each provide different entry points. This is very well reflected in the idea of research graphs, where a diversity of entities can be interconnected via various kinds of relationships. It was an open question, if persistent identifiers are necessary to make RI findable. One point of view was that PIDs facilitate identifying a certain entity. And this in turn increases the findability of entities and their relationship to each other.

The former refers to the function of a PID as an identifier. Persistence was also addressed, as findability should not be linked to the life cycle of a research activity. Even after the completion of a research project, metadata describing this project and its output should be sustainably available. From this perspective, the PID-issuing institution fulfils a social contract that guarantees the sustainable availability of the referenced metadata. The importance of this function in a fluctuating research landscape was rated as rather important by the participants of the workshop.

A third feature for the findability of RIs is that they are usually accompanied by metadata. All common PIDs for academically relevant entities are accompanied by metadata schemas that allow at least basic descriptions. Popular examples here are DOIs (with the metadata schemes of Crossref and Datacite) or ORCID. This metadata is often times machine-readable, which allows consumption and ingest by other information systems to allow more opportunities for findability. PIDs may not be necessary, but they are an established and helpful means of maintaining the sustainable discoverability of research information.

Several stakeholders for the findability of research information were mentioned. Researchers, as well as research organizations, are interested in discovering research. A special emphasis was put on findability for the public. The reasons for this include better advertisement to taxpayers, administrations and politicians, who usually fund research, and engagement of the society to generate the impact of societally relevant research.

Solutions like activity stream 2.0 (<https://www.w3.org/TR/activitystreams-core>) were mentioned, to enable automatic discovery of new content by other systems. This fits well with the observation that even if something is easy for humans to decipher, it can still be opaque to machines. The general consideration of established web standards (as from the W3C) could provide a remedy here.

The general infrastructure of research information systems was seen as problematic. On the one hand there are a lot of institutional CRIS systems that keep their data behind closed walls, which was seen as not only a waste of public resources, but also as very problematic for the general commons of research information. It was stated, that institutional CRIS “could be a searchable resource as well” therefore CRIS managers and content providers should more actively register their systems in the Directory of Research Information System (DRIS), similarly like Re3data for data repositories, or DOAR for open access repositories.

#### ***4.2. Accessibility***

The discussion in the workshop on accessibility quickly led to the finding that all scholarly metadata without any exception and with explicit inclusion of abstracts should be licensed under a CC0 licence. This is a binary question. RI is either completely open, or it is closed. There are valid reasons for institutions to keep RI internally, but that automatically leads to the fact that these RIs might not be FAIR by definition. Furthermore, it is the legitimate interest of researchers to retain control over the RI associated with their digital identity. One example is the possibility to keep RIs in research profiles but not to make them visible. A big issue for the FAIR RI commons is an alarming number of research and data that is not accessible, especially from industry research. A prominent case is the research in the area of artificial intelligence. The exact amount of information that corporations like Google, Twitter, Facebook etc. collect, share and provide, is unknown.

Accessibility for RI means making research institutions responsible for collecting information, managing, preserving and disseminating it. The institutions should do it through open standards, open protocols, open formats, open technologies. There is a duty for institutions to use open standards which means to distribute this information and to make it available. It is also the responsibility of the institutions to curate the data in order to guarantee high data quality.

On a technical level some observations were made. Accessibility means that there is a human-readable view of RI and APIs for retrieval by machines. The use of open formats, standards and protocols such as CERIF was recommended to lower the barriers to access. Other means include compliance with the OpenAIRE guidelines for CRIS managers.

One concern was the availability of RI beyond current research activities. Once a project is finished, the interest in maintaining the project-relevant infrastructure drops abruptly. Applied to the attention span of research information, this means that after an evaluation cycle, there is no longer a high priority on documenting research activities in metadata and then making this RI accessible. One suggested solution were backups - regularly produced data dumps of

CRIS - or aggregations of RI in dedicated open infrastructures. The use of persistent identifiers, with the ensuing distribution of information chunks, can help.

Another major concern was the restriction of access by RI monopolies. The publication system with a few very strong players was used as a negative example. Currently, there are often dependencies on a few players who pull RI together from different sources and make it proprietary - sometimes by adding things like accession numbers or keywords - and thus inaccessible to all other stakeholders.

The roles of the different stakeholders were defined as follows: Researchers should be responsible to provide the information to research institutions, governments, funders and policymakers. Research institutions should be responsible for recording, collecting, curating and sharing RI. Key stakeholders for ensuring RI accessible are research funders. Public funders are usually able to see the value in making “their” research accessible, but private ones will often make the information available, too. The stance of policymakers varies, of course, and the level of disclosure of information varies from one to another.

Accessibility can also mean overcoming language barriers. The rich metadata of the object that is available in several languages (for example, local and English) facilitates its use among users. Accessibility also offers great opportunities for the use of automatic translation programs. This greatly simplifies the understanding of information and is very important for health care (Samkange-Zeeb et al., 2020).

A good example of the implementation of this FAIR principle is the register of unique identifiers of ORCID for scientists, which offers users public and private access to the information. Access to public information is open to all Internet users while access to private information requires appropriate permissions, while scientists independently adjust the parameters of information availability in their profiles.

This applies not only to, for example, research data sets, but also to laboratories or equipment, as it is important for potential users to know exactly where and on what equipment this data was created. It is clear that the availability of RI metadata depends primarily on the institutions that manage it, so it is critical that the creation, collection and dissemination of this metadata and all related processes be done through open standards, protocols, formats and technologies.

### ***4.3. Interoperability***

Interoperability was seen as the crucial principle for FAIR RI, being the base of the others. It incorporates standards, procedures, ontologies, workflows and mappings that allow the information exchange across systems designed to host such research information entities. In-

formation entities and characteristics are clearly defined from a conceptual, technical and organizational (data gathering processes) perspective.

A core means to ensure interoperability are open protocols for metadata exchange. W3C standards typically ensure just the syntactic interoperability. Another level of interoperability that has to be addressed is semantic interoperability. This can be addressed on a higher level of abstraction by standards like CERIF or the Research Core Dataset, which provide definitions of research information entities. The distinction between technical and semantic interoperability was deemed to be very important. Focussing very narrowly on machine interoperability based interoperability is not sufficient. It was stated that semantic interoperability is difficult to implement. Research libraries can play an increasing role in supporting the implementation of interoperable research information management, as they are in a privileged position to advocate these values towards the administration of research institutions.

Different requirements need to be fulfilled to realize interoperability. First of all there are standards. Standards for each type of research information need to be developed and, most importantly, disseminated to all relevant stakeholders. These standards should be open and under defined licences.

Other requirements include:

- Each record of research information (research activity / result) has to be clearly identifiable. Persistent identifiers can be a means to fulfil this requirement.
- Data quality and curation processes have to be guaranteed.

There are different stakeholders in the field of RI interoperability. On the institutional level there are universities and other research institutions, where research is carried out, and where researchers, librarians, programmers, managers of institutes and universities are located. These institutions usually have to ensure the interoperability of RI, because they are the place where the information originates. Policymakers should have a high interest in RI interoperability to allow aggregation of RI for purposes like policy monitoring or measuring research performance. Funders usually have first-hand information on the research they fund. Due to their importance to all other stakeholders they have strong levers to obtain RI of the very highest quality, and thus can enforce interoperability of RI. Other stakeholders include scientometricians, who want to process RI for science studies, service providers and research infrastructure providers.

Several relevant barriers to interoperability of RI were mentioned. Critical vendor lock-in was seen as one of the biggest risks of RI interoperability. Other issues were the lack of standards, which by some was interpreted as being a result of the existence of too many standards. "My standard is the best!" was a sentence that most participants have heard before. In

the same vein is the question raised as to what we do if we have 30 different PIDs for one kind of entity.

The lack of awareness among researchers and research organizations was another issue. If there are no direct constraints, often no reason is seen for aspirations to ensure interoperability. If the information is produced e.g. by a university only or with a strong focus for internal purposes, then there is no incentive and need to open up the RI and spend resources on interoperability. This is amplified if there is a lack of aggregation systems on local, regional or national level, for example if there is no national CRIS.

The mechanisms for producing research information were also addressed. Often times this happens highly distributed, without curation at the source level, without binding standards. This frequently results in information loss and multiple levels of editing and curating. Tools, standards, platforms, processes are missing to allow RI to be entered only once, to be curated only once and to be shared by many.

Different elements need to be standardized and implemented by research communities to support RI interoperability. The participants emphasized that this is not a complete overview.

#### **4.4. Reusability**

It has been noted that there is no need to look for distinguishing mark between “use” and “re-use” as it is too blurry. When we talk about reusability of research information in fact it implies usability for external stakeholders and even for ourselves in the future. In the input talk on reusability Stefanie van de Sandt presented her understanding of this term: *“I define (re)use as the use of any research resource regardless of time, purpose, transformation and user. Special characteristics of (re)use beyond ‘using’ a research object are not considered”* (van de Sandt et al., 2019). There is no limitation in the FAIR principles to what reuse means.

An important aspect associated with reusability for computers is consistent use of persistent identifiers for each entity like organizations, persons, conferences, funders and ideally for grants because it helps minimize the risk of getting the wrong metadata. Some funders, for example, Wellcome Trust have already started using Crossref grand Identifier. Similarly, ORCID has an intention to add a new relationship type “funded by ” to enable connection between a work and its funding.

It was noted that there is a need for proxy mechanisms to ensure data quality. The data providing institutions should deliver reliable information about the completeness and the coverage of the RI. Harmonization and matching of data from different sources was seen as a challenge, too. PIDs were suggested to serve a function here, if the PID-issuing institutions can vouch for the quality. These institutions fulfil a further role by enabling usability beyond the

life cycle of the actual research activity, too, with the persistent function of the PIDs. One challenge seen here is that often different PIDs are used for one type of entity. And different PID types require different metadata fields that are maintained by different agencies that come with different standards. This is a challenge to be solved in the field of interoperability.

Research information systems were seen as both producers and consumers of research information. Discussion on NORA system (Hytteballe Ibanez et al., 2020) which harvests metadata from Dimensions and local Danish CRIS systems revealed issues of metadata quality and its openness which need to be addressed.

As a possible solution to enable reuse the idea was issued to establish national research information systems with high quality scholarly metadata for the most common research information entities. This would consist of “a basic layer of data or metadata in this case that are truly open and [on top of that we can] build several service layers and not all of these need to be open, if the underlying metadata layer is open”, allowing for a combination of proprietary and open services. A special emphasis was made on the importance of proper data provenance, especially when it comes to aggregators like OpenAire or future Open Knowledge Base in the Netherlands that collect different types of data from different sources and there is a kind of risk that at the too early stage data is already processed in a certain way.

Among key barriers of reusable RI are poorly described metadata, mismatched vocabularies and absence of agreed minimal (optimized) set of metadata elements.

The legal act that constitutes reuse was also addressed. As a rule, there are numerous legal obstacles, including discrepancies in the legislation regulating the activities of the various participants in research. Second, there is no unified or prioritized policy at the state level in collecting data on research conducted and funded in Ukraine and abroad. The differences between national legislation in an international ecosystem such as research make it difficult to guarantee the reusability of RI. A related issue is the (non)use of free licences for scholarly metadata, and incompatibility between different licences.

The stakeholder groups for the reusability of RI are all those who could use research information. The stakeholder groups for the post-usability of RI are logically all those who could use research information. A variety of different examples were collected in the workshop. This includes “basically everyone who produces, aggregates, (re)uses research information”.

Organizations:

- Government and non-government agencies
- research agencies
- foundations
- universities
- businesses

- publishing houses
- funding bodies
- Research Groups & Departments
- University management
- Research Support (Libraries, Funding Support)
- Companies
- special interest groups (lawyers, patients, etc)
- information service providers
- distributors of data

Individuals:

- librarians
- archivists
- lawyers
- journalists
- citizens
- Individual Researchers

Another finding was that the reuse of RI often helps to improve data quality, as errors can be discovered and subsequently corrected through reuse. Likewise, reuse has been seen as the way to overcome the reproducibility crisis of research results.

#### ***4.5. Initiatives to realize FAIRness of RI***

There are several initiatives which contribute powerfully to FAIR implementation of Research information. For instance, the recently launched Initiative for Open Abstracts (I4OA) (Waltman et al., 2020) that promotes the open availability of scholarly abstracts and its sister organization Initiative of Open Citations (I4OC) (Levchenko, 2018). Both initiatives demonstrate profound impact in increasing availability of open RI.

OpenAIRE infrastructure represents another significant development toward open metadata of publications, datasets, projects, organizations, data source and persons. The infrastructure provides services to continuously harvest from data sources (e.g. literature repositories, data repositories, CRIS systems) metadata descriptions related to research outcome and aggregate, harmonize and enrich by inference such content (Houssos et al., 2014). An important role here also plays CERIF (de facto known as a standard for RI) and implementation of the OpenAIRE Guidelines for CRIS Managers based on CERIF-XML, which allow Current Research Information Systems (CRIS) to be harvested by the OpenAIRE content aggregator. In the long run OpenAIRE might serve as a CLOCKSS (CLOCKSS, 2021) for RI.

Another example is a project on constructing the Open Knowledge Base (OKB-NL) in the Netherlands which aims to collect and store metadata from Dutch research institutions and universities in accordance with existing standards, identifiers, ontologies and thesauri. The future infrastructure will be open for others to access and reuse, whether via an API, SPARQL endpoint, a data dump, or a simple interface, typically via a CC0 licence (Open Working, 2020). Big drivers of FAIR RI are Crossref and ORCID.

## **5. Discussion**

Various research activities are being accompanied by high production of research information which makes it an essential component of the research ecosystem. Ideally, research information should be consistent, up to date, openly available (Blümel et al., 2014) and accessible but in fact, very often, it is unstructured, does not provide provenance information and resides in multiple heterogeneous and often non-accessible systems.

Professionals, librarians, administrators and other scientific stakeholders need to be brought together to implement best practices and standards. Research libraries around the world are playing an increasing role in supporting the implementation of data in research management systems. Just an example: Despite the fact that the Ukrainian library network is one of the largest in the world and most academic libraries use automated library information systems, the issues of creating unified search library services and improving the interoperability of library systems remain completely unresolved. Ukraine mainly uses two formats for storing bibliographic descriptions - MARC21 and UNIMARC. It is necessary to integrate metadata from different Integrated Library System (ILS), convert them into a Resource Description Framework (RDF) format consistent with the subject ontology, as well as analyse the existing mechanisms for publishing data in Linked Open Data (LOD). Implementing these steps will create up-to-date services for scientists and improve the visibility of this data.

A similar ambiguous situation is connected with the state planning of the development of national academic resources of open access. According to the Registry of Open Access Repositories, there are more than 100 open repositories in Ukraine. As of January 2021, 370 journals are represented in the Directory of Open Access Journals. Despite such quantitative achievements in terms of openness, today Ukrainian scientific publications are officially aggregated on the portal "Scientific Periodicals of Ukraine", which does not offer API. Instead of improving the policy of the existing portal, in 2016 Ukraine started another project the "National Repository of Academic Texts". It is known that the new portal should use the capabilities of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and SWORD (Simple Web-service Offering Repository Deposit), but this seems to limit all requirements for interoperability of the future service, which is still in the process of development.

Many repositories (Figshare, Zenodo) and academic social networks (ResearchGate) automatically generate DOIs for deposited documents. For example, the author presented his pub-

lication in the institutional repository (especially if the institution has adopted a policy of mandatory self-archiving). At the same time, the author presented the same publication also on a popular thematic repository or social network. Also, each co-author of a work may submit a copy of one work to the institutional repository of his or her institution. Thus, one object can receive several identifiers at once.

Persistent identifiers are used primarily for the correct identification of objects, including the correct citation of publications and references to relevant components of the research infrastructure. It is obvious that the existence of the risk of assigning several identifiers to one object at once will not allow to correctly assess its impact in the system of scientific communications, because the scattering of links between different identifiers will distort and underestimate such an assessment. Similarly, it prevents the correct assessment of scientific impact using altmetrics tools.

Moreover, serious competition from commercial players (Kim & Oh, 2020) and possible mistakes made in planning the development of institutional archives have made it difficult for institutions to justify the financial feasibility of creating separate repositories for research data. Separate repositories are usually created to present research data. This type of repository is becoming increasingly popular, as modern science is experiencing a crisis of reproducibility of results, and the practice of presenting the original data of research in the public domain can help the academic community to solve this problem. In Ukraine, publications and data are uploaded to a single institutional repository, and the need to create rich metadata for datasets is generally ignored by repository administrators. It is clear that the situation with other RI objects is not the best either.

In the relevant FAIRIO workshop, the experts agreed that research information would comply with the principle of availability if all elements of metadata were distributed by default under an open licence, which would allow all stakeholders to freely improve, integrate and reuse this metadata. For example, today's abstracts of scientific papers are mostly closed elements of metadata. Copyright applies to abstracts, as well as to the full text of the article, so they are not open to third-party algorithms for text mining.

In September 2020, the Initiative of Open Abstracts was created, which calls on all scientific publishers to open abstracts of their publications and make them freely available (Waltman et al., 2020). Open access to abstracts will help scientific publishers expand the readership of their publications, and scientists will be able to find important publications faster with the help of text mining, natural language processing and artificial intelligence technologies. Active support for this initiative by publishers will accumulate a lot of open abstracts, which will positively affect the visibility of RI components on the Internet, as well as help improve the functionality of specialized services for scientists, including search engine and citation database Open Ukrainian Citation Index (OUCI), which uses open data from Crossref (Cheberkus & Nazarovets, 2019).

Also, it should be noted the rapid development of the Initiative for Open Citations (I4OC), because in 2021 it was joined by two major publishers of scientific literature - Elsevier and American Chemical Society.

Ukrainian officials responsible for the development of education and science are often convinced that active international scientific cooperation is extremely important for the development of science in Ukraine. However, Ukraine has not yet created a national register of scientific infrastructures, undeveloped recommendations and requirements for metadata, in the implementation of joint research projects are used standard forms of agreements that do not take into account all issues of legal compatibility. The use of FAIR principles in the state policy in the field of scientific and technical activity would contribute to the creation of better opportunities for the development of international scientific cooperation in Ukraine.

As we have indicated above, some original FAIR principals are absolutely clear and directly applicable for research information whereas others require careful rethinking and practical definitizing. Our analysis highlighted that production of research information is highly distributed, often without curation of the source and non-standardized publication of research information. This results in information loss and the need for unnecessary and costly additional work for reuse. It must be the imperative for RI-providing institutions - those responsible for collecting information, managing, preserving and disseminating it to do this by using open standards, open protocols, open formats and open technologies.

Another important aspect which was stressed during the discussions is that all actors should organize their workflow in a way which minimizes or excludes the possibility of monopolizing research information by for profit organizations like it happened, for instance, with scientific publication. Key stakeholders for ensuring RI is findable and accessible are research funders.

Future research should focus on moving beyond local data structures and semantics to a global knowledge graph where all research information can be conceived as features and subsets of a global knowledge graph of research information. Local systems are needed to support local operations, maintain non-shareable information, and then should provide data to the global scholarly knowledge graph.

### **Funding acknowledgement**

The contribution of the authors associated with the TIB was partially funded by the Federal Ministry of Education and Research of Germany in the framework of FAIRIO (project number 01DH20080). The contribution of the authors associated with the SSTL was funded by the Ministry of Education and Science of Ukraine.

## References

- Alberts, B. (2013). Impact Factor Distortions. *Science*, 340(6134), 787–787. <https://doi.org/10.1126/science.1240319>.
- Blümel, I., Dietze, S., Heller, L., Jäschke, R., & Mehlberg, M. (2014). The Quest for Research Information. *Procedia Computer Science*, 33, 253–260. <https://doi.org/10.1016/j.procs.2014.06.040>.
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: fair enough? *European Journal of Human Genetics*, 26(7), 931–936. <https://doi.org/10.1038/s41431-018-0160-0>.
- Castro, P. de, Kimidi, S. S., & Palavesm, K. (2020). The Rise of Current Research Information Systems (CRIS): The Case of the Indian Research Information Network System (IRINS). <https://doi.org/10.35543/osf.io/uv5pg>.
- Cheberkus, D., & Nazarovets, S. (2019). Ukrainian open index maps local citations. *Nature*, 575(7784), 596–596. <https://doi.org/10.1038/d41586-019-03662-6>.
- CLOCKSS. (2021). *How CLOCKSS Works*. CLOCKSS. <https://clockss.org/about/how-clockss-works/>.
- Engelman, A., Enkvist, C., & Pettersson, K. (2019). A FAIR archive based on the CERIF model. *Procedia Computer Science*, 146, 190–200. <https://doi.org/10.1016/j.procs.2019.01.076>.
- Gregg, W., Erdmann, C., Paglione, L., Schneider, J., & Dean, C. (2019). A literature review of scholarly communications metadata. *Research Ideas and Outcomes*, 5. <https://doi.org/10.3897/rio.5.e38698>.
- Hagemann-Wilholt, S., Hauschke, C., & Plank, M. (2020). Confident – an open platform for FAIR conference metadata. *Grey Journal*, 16(2), 95-100.
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429–431. <https://doi.org/10.1038/520429a>.
- Houssos, N., Jörg, B., Dvořák, J., Príncipe, P., Rodrigues, E., Manghi, P., & Elbæk, M. K. (2014). OpenAIRE Guidelines for CRIS Managers: Supporting Interoperability of Open Research Information through Established Standards. *Procedia Computer Science*, 33, 33–38. <https://doi.org/10.1016/j.procs.2014.06.006>.

- Hytteballe Ibanez, K., Sandfær, M., Steensboe, C., Lauridsen, N. D., Parraguez Ruiz, P., Lowe, B., Falcoz, F., & Kirkegaard Lunn, B. (2020). *NORA, National Open Research Analytics, OPERA Conference 2020* (Version 1). figshare. <https://doi.org/10.6084/m9.figshare.13272932.v1>.
- Ivanović, D., Ivanović, L., & Layfield, C. (2019). FAIRness at University of Novi Sad - Discoverability of PhD research results for Non-Serbian scientific community-. *Procedia Computer Science*, 146, 3–10. <https://doi.org/10.1016/j.procs.2019.01.071>.
- Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., ... Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. *Data Intelligence*, 2(1–2), 10–29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024).
- Jeffery, K., Houssos, N., Jörg, B., & Asserson, A. (2014). Research information management: the CERIF approach. *International Journal of Metadata, Semantics and Ontologies*, 9(1), 5. <https://doi.org/10.1504/IJMSEO.2014.059142>.
- Kim, Y., & Oh, J. S. (2020). Researchers' article sharing through institutional repositories and ResearchGate: A comparison study. *Journal of Librarianship and Information Science*, 096100062096284. <https://doi.org/10.1177/0961000620962840>.
- Levchenko, M. (2018, February 16). *Why every researcher should care about open citations*. [web log]. Europe PMC Blog. <http://blog.europepmc.org/2018/02/why-every-researcher-should-care-about.html>.
- Open Working. (2020, May 29). *What is an Open Knowledge Base anyway?* [web log]. Open Working. <https://openworking.wordpress.com/2020/05/29/what-is-an-open-knowledge-base-anyway/>.
- Peroni, S., & Shotton, D. (2020). OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies*, 1(1), 428–444. [https://doi.org/10.1162/qss\\_a\\_00023](https://doi.org/10.1162/qss_a_00023).
- Samkange-Zeeb, F., Samerski, S., Doos, L., Humphris, R., Padilla, B., & Bradby, H. (2020). “It’s the First Barrier” – Lack of Common Language a Major Obstacle When Accessing/Providing Healthcare Services Across Europe. *Frontiers in Sociology*, 5. <https://doi.org/10.3389/fsoc.2020.557563>.

- Science Europe. (2016). *Science Europe Position Statement on Research Information Systems*. Brussels, Belgium: Science Europe. [https://scienceeurope.org/media/qbzi-uyj2/se\\_positionstatement\\_ris\\_web.pdf](https://scienceeurope.org/media/qbzi-uyj2/se_positionstatement_ris_web.pdf).
- van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A., & Petras, V. (2019). The Definition of Reuse. *Data Science Journal*, 18(1), 22. DOI: <http://doi.org/10.5334/dsj-2019-022>.
- Waltman, L., Kramer, B., Hendricks G., Vickery, B. (2020, September 25). *Open Abstracts: Where are we?* [web log]. Crossref Blog. <https://www.crossref.org/blog/open-abstracts-where-are-we/>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., ... Hill, J. (2015). *The Metric Tide: Report of the Independent Review of the Role of Metrics in Research Assessment and Management*. <http://doi.org/10.13140/RG.2.1.4929.1363>.