



RESEARCH ARTICLE

Evaluating the scientific impact of research infrastructures: The role of current research information systems

Renaud Fabre¹ , Daniel Egret² , Joachim Schöpfel³ , and Otmane Azeroual⁴ 

¹Dionysian Economics Lab (LED), University Paris 8, Saint-Denis, France

²Observatoire de Paris, Université Paris Sciences & Lettres, Paris, France

³GERiCO Labor, University of Lille, Villeneuve-d'Ascq, France

⁴German Center for Higher Education Research and Science Studies (DZHW), Berlin, Germany

an open access  journal



Citation: Fabre, R., Egret, D., Schöpfel, J., & Azeroual, O. (2021). Evaluating the scientific impact of research infrastructures: The role of current research information systems. *Quantitative Science Studies*, 2(1), 42–64. https://doi.org/10.1162/qss_a_00111

DOI:
https://doi.org/10.1162/qss_a_00111

Received: 21 September 2020
Accepted: 22 November 2020

Corresponding Author:
Otmane Azeroual
azeroual@dzhw.eu

Handling Editor:
Ludo Waltman

Copyright: © 2021 Renaud Fabre, Daniel Egret, Joachim Schöpfel, and Otmane Azeroual. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: Common European Research Information Format, current research information systems, research facilities, research information management, research infrastructures (RI)

ABSTRACT

Research infrastructures (RIs) offer researchers a multitude of research opportunities and services and play a key role in the performance, innovative strength, and international competitiveness of science. As an important part of the generation and use of new knowledge and technologies, they are essential for research policies. Because of their strategic importance and their need for significant funding, there is a growing demand for the assessment of their scientific output and impact. Current research information systems (CRIS) have contributed for many years now to the evaluation of universities and research organizations. Based on studies on the application of CRIS to infrastructures and on a recent French report on the scientometric assessment of RI, this paper analyzes the potential of CRIS and their data models and standards (in particular the international CERIF format and the German RDC model) for the monitoring and evaluation of RIs. The interaction between functional specificities of RI and standards for their assessment is outlined, with reference to their own potential to stimulate and share innovation in the networks located inside and outside RIs. This societal challenge, more than an academic issue, is on the way to further harmonization and consolidation of shared and common RI metrics.

1. INTRODUCTION

Research infrastructures (RIs) are facilities, resources, systems, and services needed by scientific communities to carry out large-scale research in cutting-edge fields. The European MERIL project defines RI as a “facility or (virtual) platform that provides the scientific community with resources and services to conduct research in their respective fields. These research infrastructures can be single-sited, distributed or an e-infrastructure, and can be part of a national or international network of facilities, or of interconnected scientific instrument networks” (Beckers, Jägerhorn, & Höllrigl, 2012). Examples of RI are astronomical observatories, particle accelerators, synchrotrons, lasers, and intensive computing resources, as well as data production and management tools. These infrastructures are used by researchers from all disciplines, in astronomy, biology, physics, chemistry, human and social sciences, earth

sciences, etc., who thus have access to high-performance equipment in a high-level scientific environment¹.

The RI road map of the French Ministry of Higher Education, Research and Innovation (MESRI) enumerates 99 infrastructures: large national and international research facilities covering all disciplines, “incredible engines of knowledge, attractors of talent, catalysts for collaboration, bearers of scientific image and prestige [...] not work tools like others,” because of their longevity, their ambitions, and their costs (largely over €1 billion per year) (Ministère de l’enseignement supérieur, de la recherche et de l’innovation [MESRI], 2018). The French road map includes, among many others, the GENCI Company for high-performance computing, the CERN Large Hadron Collider, the CTA Cherenkov Telescope Array, the SOLEIL Synchrotron, the OpenEdition scientific publishing platform for the social sciences and humanities, and the Huma-Num digital humanities platform.

In 2018, the French Ministry of Higher Education, Research and Innovation commissioned a scientometric analysis for a shared assessment of the scientific impact measures of 24 very large RIs and international organizations². The challenge is multiple: a better assessment of the scientific impact of each facility, at the level of disciplines and subdisciplines; a better identification of research collaboration at the national, institutional, and individual levels; the detection of emergent research topics; and a contribution to scientific foresight and advice, as part of the policy-making mechanism.

The results of this analysis were published in November 2019 (Egret & Fabre, 2019). With regard to research information processing, the study reveals a high degree of diversity and specificity. Most RIs make use of some kind of current research information system (CRIS) to provide information about the use of their services, resources, and systems. CRISs are an instrument for the management of research information and are linked to various internal and external systems or databases (finance, SAP, HR systems, project management systems, open access repositories, Web of Science, Scopus, PubMed, national libraries, BASE, CrossRef, EVALuna, Ebsco, equipment management systems such as ULab, and others); they “collect and store metadata on research activities and outputs such as researchers and their affiliations; publications, data sets, and patents; grants and projects; academic service and honors; media reports; and statements of impact” (Bryant, Clements et al., 2017), to support research institutions in the provision of funding information and reporting, in aggregating references for research outputs, and in producing indicators and assessment (De Castro, 2018). With standardized formats and functionalities, they are first and foremost designed for academic institutions, research organizations, and authorities, not for infrastructures. So how can standard CRISs provide solutions to the particular needs and expectations of (especially large) RIs? What is the potential impact of the French approach to RI impact metrics on the further development and implementation of research information systems in this field?

This article contributes to the assessment of the real and potential role of CRISs for the evaluation of RI on three levels.

- First, we review published literature on the topic of research information systems for RI, especially from the euroCRIS seminars, meetings and conferences.

¹ See the interview with Gabriel Chardin, former president of the CNRS RI committee <http://www.cnrs.fr/cnrsinfo/dans-les-tgir-se-construit-la-societe-du-futur-gabriel-chardin-president-du-comite-tres>.

² TGIR (Très Grandes Infrastructures de Recherche) and OI (Organisations Internationales).

- Second, we provide a summary of the French study on RI assessment, in particular of the specific demands and expectations and of the recommendations for further action in the field of RI metrics.
- Third, we discuss how the Common European Research Information Format (CERIF) and the German Research Core Dataset (KDSF/RCD) meet the requirements of present standard information and metrics for RI.

The conclusion will summarize the main findings and make recommendations for further research and development on the evaluation of RI.

2. EVALUATING RESEARCH INFRASTRUCTURES: A REVIEW

This section reviews published literature on the real and potential interest of research information systems for RI, based especially on the papers presented during the euroCRIS seminars, meetings, and conferences. Building RI is one of the priorities of European research policy, to foster international cooperation and integration, to provide tools for the development of open science, and to improve the performance of academic research. For many years now, the European Commission has provided funds for a large variety of e-infrastructure (Buhr, 2014), and similar strategies can be observed at national and regional levels, such as the German Excellence Initiative (Spang-Grau, 2019) or the Finnish Research Information Hub (Puuska & Rydman, 2018). However, funding implies awareness and a good knowledge of existing infrastructures, and it also implies follow-up, reporting, and monitoring; therefore, the need for the evaluation of the output of European or national research policies, in terms of performance indicators of RI, has been clearly identified by authorities and funding agencies.

The need for RI assessment has been highlighted by the OECD (Organisation for Economic Co-operation and Development [OECD], 2019) and by the ESFRI Roadmap³, as well as by the Cour des Comptes (French Court of Audit) in its parliamentary report of May 2019 on the governance and funding of large French RIs. The OECD recommends backing up the socio-economic impact assessment of RIs with a catalog of “core impact indicators” of their scientific performance, such as those metrics developed by the European Spallation Source infrastructure⁴; these metrics include the number of citations, the number of publications in high-impact journals, the number of projects granted, the number of scientific users, the number of patents with commercial use, the number of full-time equivalent (FTE) staff in the RI, and so on.

Florio, Forte, and Sirtori (2016) published a case study to show how a social cost-benefit analysis (CBA) probabilistic model can be applied to evaluate a large-scale RI project, the CERN Large Hadron Collider. Based on empirical methods, they estimated that there is around a 90% probability that benefits exceed costs, with an expected net present value of about €2.9 billion. Their approach combines several categories of data sources, including

- (a) accounting data and expert analysis of capital and operating expenditures, including in-kind contributions; (b) scientometric data to estimate trajectories of publications and their impact in a specific domain; (c) firms’ survey data on technological spillovers expressed in terms of increased sales and cost savings, or increased profits; expert analysis of the technological content of procurement; company accounting data for industries involved in procurement; and expert analysis of the cost savings or other quantifiable effects of open source software or other technological spillovers; (d) survey data and other statistical evidence of

³ European Strategy Forum on Research Infrastructures <https://www.esfri.eu/esfri-roadmap-2021>.

⁴ ESS <https://europeanspallationsource.se/>.

the expected or ex-post effects on salaries of former students and early career scientists; (e) statistics about on-site visitors, web access, use of social media, exposure to traditional media, and data on travel costs, opportunity costs of time, and other information related to cultural effects; (f) contingent valuation data through survey of samples of potential taxpayers about their WTP for potential discoveries related to a specific project” (Florio et al., 2016).

Their multilevel methodology reveals the complexity of the issue, especially for very large RIs. Compliant with the OECD recommendations, they connect socioeconomic assessment with scientometric indicators of the RI scientific performance, in terms of publications and citations (impact). Interestingly, Florio et al. (2016) estimate that the resulting total present value of the publications (€277 million) “per se pays back only a tiny fraction of around 2% of the total cost.” Interestingly, too, their methodology, which needs a high degree of data integration and standardization, does not make use of a CRIS or similar system, neither as a data source nor as a tool for the processing and analysis of collected data.

Perhaps the best-known example of international cooperation in this field is the MERIL initiative (Mapping of the European Research Infrastructure Landscape), started in 2010, funded by the European Commission and coordinated by the European Science Foundation (ESF). Its objective was to provide key information for stakeholders, to achieve an inventory of RI of European relevance and to make the information publicly available through an interactive online portal (Beckers et al., 2012; Dvorak, 2013). The MERIL initiative is interesting for at least four reasons.

1. It defines three basic quality criteria for the inclusion of existing RI in the European portal: They must offer scientific and/or technological performance and support that should be recognized as being of European relevance, they must offer access to scientific users from Europe and beyond through a transparent selection and admission process, and they must have a management structure. If an RI does not fulfil these criteria, it will not be included in the MERIL Portal. The debate over the meaning of “European relevance” has contributed to the definition of common features of European RI.
2. MERIL has defined a couple of elements as relevant for the description of RI, including information useful for evaluation issues, such as the number of users, research services, and equipment.
3. These basic elements have been developed in compliance with the European format for research information, CERIF. MERIL, for this reason, has been described as a “connected e-infrastructure”, interoperable in particular with existing research information management systems (Brasse, 2012).
4. The interoperability allows the creation of tools and services on top of the MERIL directory, such as single national contact points to foster RI cooperation (Houssos & Karaiskos, 2013) or, in the framework of the Research Infrastructures Consortium (RICH project), single entry and access points to information about RI (Tzenou & Bonis, 2016).

As part of the MERIL-2 follow-up project (Baginskaite, 2017)⁵, the European Science Foundation (ESF) introduced in September 2018 a new data visualization tool that allows users to discover and explore data on the European research landscape, such as the RI size and location, user profiles, and research capabilities of over 1,000 research facilities across the continent. Based on MERIL-2, the new catalog of RI services, CatRIS⁶, provides a kind of

⁵ MERIL-2 <https://portal.meril.eu/meril/>.

⁶ CatRIS <https://www.portal.catris.eu/>.

standard framework with minimal data for the description of the service providers and the services themselves, such as access to and use of facilities and instruments, user support and training, and other activities and resources that the RIs deliver to users and customers.

Assessing RI is also a challenge for specific research communities with significant scientific instruments, facilities, and equipment. We give three examples. First, research organizations in the field of solid earth sciences implemented some years ago a new project called European Plate Observing System (EPOS), designed “with the vision of creating a pan-European eResearch Infrastructure for solid Earth science to support a safe and sustainable society [...] the EPOS mission is to integrate the diverse and advanced European Research Infrastructures for solid Earth science relying on new e-science opportunities to monitor and unravel the dynamic and complex Earth System” (Bailo, Ulbricht et al., 2017). The main challenge of this project was the development of a common metadata model to describe and assess in an appropriate way the large variety of persons, services, data, equipment, software, organizations, web services (API), and RI. In particular, the EPOS project identified 12 mandatory and optional elements (attributes) of the IR entity, which are compliant (interoperable) with the CERIF format for research information management (Figure 1).

The second example is from the field of environmental research, where Boldrini, Luzi et al. (2014) demonstrated how to implement the CERIF data model to assess and describe RI especially in global and multidisciplinary contexts.

The third example is the model of the data continuum in photon and neutron facilities developed by the UK PaN-data ODI project (Matthews, 2012), which provides a detailed mapping and description of the research and data life cycle of these facilities. The proposed elements, especially the actors and the stages of the experimental lifecycle, can be considered as basic elements for the evaluation of the performance of RI.

A quite different use case of the evaluation of RI is the European VRE4EIC project (Ivanovic, Theodoridou, & Remy, 2018; Theodoridou, Patkos, & Doerr, 2016). Coordinated by the European Research Consortium for Informatics and Mathematics (ERCIM), this project builds on existing e-RIs providing services, software, data, and resources to develop an enhanced virtual research environment (VRE). The ingestion of information on research, data, and computing infrastructures requires interoperable (standard) metadata on RI. Again, the CERIF is selected as the target format, because of its flexibility, quality maintenance, and political support.

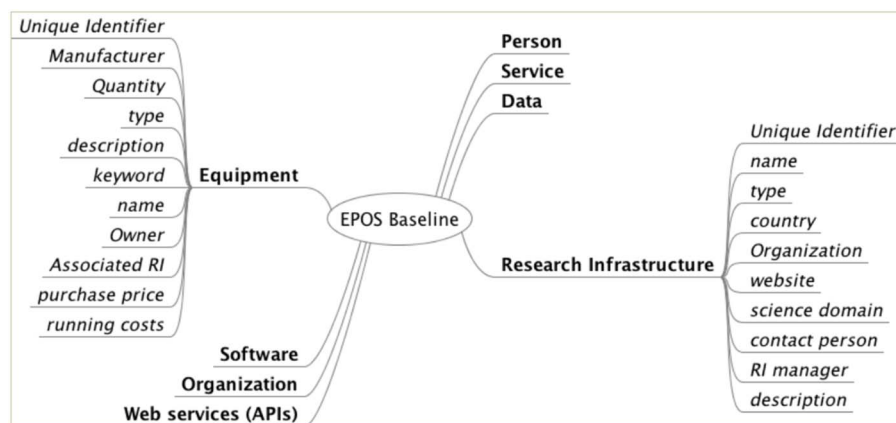


Figure 1. The EPOS metadata baseline (Bailo et al., 2017).

A last and more recent example is the implementation of the European Open Science Cloud (EOSC). One part of their activity is the inventory, description, and assessment of existing data, computational, networking, and thematic infrastructures (Vancauwenbergh, 2019). Yet, so far, (August 2020), EOSC does not provide information on descriptive elements or formats for this assessment, except for general information on national open science and FAIR data policies.

However, in spite of the need for evaluation, most projects dealing with RI and evaluation assess the RI content (i.e., data and documents), and do not consider the RI performance as an object of evaluation on itself. For instance, the initiative for collaborative research information management by the UK Science and Technology Facilities Council (STFC) which “operates large scientific facilities to support experimental research in [...] chemistry, materials science, and biochemistry (including) the ISIS neutron source, the Central Laser Facility and the Diamond synchrotron light source (with) large volumes of data and [...] used each year by many thousands of experimental scientists from around the world” (Crompton, Matthews et al., 2012) focused on the description and linking of data, and not on the assessment of the infrastructures and facilities themselves.

The UK Research Excellence Framework (REF) 2021 assesses RIs as institution-level resources and facilities available to support research, as part of the environment subprofile of universities and other research organizations, and among other supplemental criteria and similar to income, people, strategy, and contribution to economy and society (Research Excellence Framework [REF 2021], 2019). The REF panel criteria consider the investment in RI as a contribution to (or dimension of) sustainability, to ensure the future development of the units and their disciplines. Therefore, evidence is requested for the existence and strategy of RI and about its usage, quality, operation, and benefits (REF5b, Section 3), but without specifying the expected evidence (data sources, indicators etc.).

A recent survey provides some empirical elements on the place of RI evaluation in German public research (Schöpfel, Azeroual, & Saake, 2019). Universities and other research organizations are regularly evaluated and must report on their research activities. To improve the quality of this reporting, many of them have implemented some kind of CRIS, as a central database for the collection, presentation, and evaluation of data related to research. Yet, following a survey with 51 German institutions, only a small percentage (about 10%) make use of their CRIS to evaluate the performance of their own RI in terms of output and input, with appropriate metadata⁷.

This last survey raises two other issues: Insofar as RIs serve different purposes from different communities and institutions, should their performance be assessed differently, according to and for each community and institution, as is done by those German institutions with their institutional CRIS? In this case, assessing the global performance of an RI would require the aggregation of all performance metrics produced by relevant institutions and clearly identified as related to this specific RI—a tedious method whose success would require a high degree of standardization between the institutions involved. In fact, our approach is different, based on the reality of central funding and not on the reality of one community or many; instead of aggregating data and metrics from different institutions, the idea is to produce performance metrics upstream, by the RI itself.

The second issue is about the particularity of RIs. Why do they require a specific assessment, different from universities and other research institutions? CRISs are mainly designed for universities and research institutions; why do they need a specific adjustment for the assessment of RI?

⁷ Unpublished data from Azeroual O. (in preparation). *Untersuchungen zur Datenqualität und Nutzerakzeptanz von Forschungsinformationssystemen*. PhD dissertation.

There are at least four reasons (i.e., four significant differences between RI and other academic and research institutions): RIs provide temporary hosting of scientists and projects (“hotel”); RIs consist of a large equipment with an analysis output; RIs provide methodological support for the research and are not “neutral”; and RIs functioning is based on internal and external networking. We will come back to these characteristics in more detail in the following section, as part of the French study.

In summary, there is a general consensus that RIs are part of research evaluation and that they must be described and assessed. Also, because of the large variety of RIs, a standard and interoperable data model seems appropriate, in particular the only international standard format recommended by the European Commission for research information management system (CERIF). Section 4 will provide more information about the CERIF model and its potential for the evaluation of RI.

3. THE FRENCH STUDY ON RESEARCH INFRASTRUCTURE METRICS

For the reasons mentioned above, the existing procedures and metrics should take into account the particular characteristics and functioning of RI to provide appropriate assessment of RI. Data models and systems made for universities and research organizations are useful but need adjustment for the specific needs of RI. Regarding CRIS in France and compared to other European countries, there is a relatively low degree of standardization among research structures, with few CERIF-compliant systems.

In France, large RIs, known as Very Large Research Infrastructures (TGIRs), are mainly defined by their scientific potential of a national or international nature. The distinction in France between RIs and TGIRs is currently being called into question: It stems, for the most part, from agreements of administrative or financial scope, and it has been recently observed by the Court of Audit that this distinction compromises the readability of national policy (Cour des Comptes, 2019).

The TGIRs are originally French public goods, with a funding which is largely mutualized, around which the Court of Audit observes: “a historical trend towards the pooling of the support of the costs of these infrastructures in the world and, in particular in Europe.” Between 2012 and 2017, according to the Court’s estimate, “the cumulative amount of TGIR resources reached €4.2 billion, half of which came from French budget appropriations.” This method of funding ensures strong international vitality for TGIR networks, but also requires a framework in which the CRIS have their strategic place: Faced with competition in Europe for scientific choices, the Court of Audit observes the need for “mastery of decision-making processes and the conception by France of genuine influence engineering” (Cour des Comptes, 2019).

As mentioned above, the French Ministry of Higher Education, Research and Innovation⁸ commissioned a study on the impact measures of large RIs, the results of which were published in 2019 (Egret & Fabre, 2019). The report provides a review of current scientometric practices and describes the expectations and needs of infrastructure managers; moreover, it makes 15 proposals for the development of shared impact measures, and it discusses some general indicators (“publometrics”) to contribute to a conceptual and methodological framework for further harmonization and standardization of existing metrics, to improve RI evaluation practice, and to develop a common evaluation culture, while respecting the specificities of each research facility and the requirements and standards of the European and international RI landscape, in particular the need for interoperability.

⁸ Direction générale de la recherche et de l’innovation (DGRI).

As noted in the previous section, in France, universities and national research organizations manage a relationship, most often, permanent to the task of scientific discovery, while the TGIRs are generally characterized by the following:

- Temporary accommodation: The 24 TGIR, including the four major international organizations (OIs) in which France participates, host research teams from all institutional sources on scientific projects (universities and organizations with researchers from all countries, sometimes teams from private industrial research, etc.) according to quotas and rules defined at the level of each TGIR with the agreement of the major national scientific authorities (e.g., CEA or CNRS). With the exception of EMBL, which in itself constitutes a special category, TGIRs do not provide any permanent reception beyond a project, which generally spans a short period (less than or equal to 6 months).
- “Self-service” experimentation on a project: The experiment is carried out in “self-service” when the TGIR accepts the scientific project and validates the conditions for its realization, while defining the allocation of reception resources (technicians’ time, adaptation of installations to experience, instrument time, computing time), means of transport (EURO ARGO, oceanographic vessels), beam time (SOLEIL, ESRF, LLB, ...), computing resources (GENCI high-performance computing), etc. All output data are systematically made publicly available after a limited embargo period, and most often result from standard analysis pipelines.
- Technical assistance to experimenters: The survey data, which cannot be developed here, show a very systematic adaptation of the TGIR to the needs for advice, expertise, and scientific support by the teams of permanent researchers of the TGIR to all kinds of scientific projects, from the Humanities (TGIR HUMA-NUM and Progedo) to astrophysics, oceanography, climatology, etc.
- Networking of means and results: Technical assistance is frequently associated with networking. In terms of resources, this is carried out by internationalization of similar resources (e.g., LIGO and EGO VIRGO) or additional resources (neutron lines and X-ray lines coupling the experiments in a mixed program between SOLEIL and the LLB); there are also many opportunities for mutualization of instruments, software, and other resources in astronomy and climate sciences. In addition, pooling is also frequent and currently developing in the sharing of results (standardization of the presentation of acknowledgments, data, affiliations, databases, scientific publications; standardization of the presentation of platforms, European key performance indicators, ERC nomenclature for indexing disciplines, etc.).

In addition, the current trend towards a more user-friendly evaluation system at various scales ranging from local experimentation to major interdisciplinary choices, is the bearer of an original model for optimizing global scientific potential, through a committed approach to sharing of “markers of science.”

3.1. Expectations, Needs and Interests

The 2019 survey reveals the broad interest of large RIs in the development of functions and tools to analyze and share scientific results through new metrics, new software approaches to current metrics, and emerging tools to build numerical functions to support research. Several large research organizations in France have implemented systems to make their research data publicly available⁹. RI managers want to assess the output of their infrastructure in terms of data and

⁹ See, for instance, the MINnD project at the French Geological Survey BRGM (Monitoring of changes in practices and knowledge around the digital model) <https://www.minnd.fr/>.

publication, and its impact in terms of citations, but also in terms of new knowledge, concepts, ideas, etc. The French-Italian Antarctic Station Concordia, for instance, is very interested in all analytical services that can focus on the implementation of metrics to ensure the traceability of scientific production, its thematic semantic analysis, and the genealogy of concepts. The goal is to extend beyond metrics to the analysis of the value of scientific work for all publics.

With the same concern for strategic projection, the European Centre for Medium-Range Weather Forecasts (ECMWF) highlights

There is a need, in particular to trace the genealogy of scientific ideas, and to analyze the ruptures and reorientations of programs, such as those coming from the French community (optimal control, variational assimilation...)

At CERN there is

an interest in a future platform of metrics tools, notably for sorting publications and developing analyses, in relation with the publication committees of major collaborations, and for adopting coherent positions towards funding agencies. The interest is clear on the institutional side (IN2P3¹⁰), but less obvious on the side of the researchers themselves.

The European X-ray Free Electron Laser (XFEL) provides an in-depth analysis of the need:

How to organize an RI access route? This question contains that of the associated services, which can be shared or dedicated according to a “map” that is not yet sketched out ... This publimetric map can be declined according to its various vocations discussed above, and include discovery support services that identify relevant links between work in progress and published in accessible forms in an open science framework.

The French RI in the field of digital humanities, HUMA-NUM, points out

an important aspect of new needs: the implementation of DOI on data is now underway and [...] the RI has launched an OCT (Open Citations Tools) program with all the DOI reservoirs, to build an Observatory, with an “appropriate metric” and this is a hot topic for INSHS¹¹, but also for the French-speaking world, with the development of scientific French.

In a quite different domain, the European Consortium for Ocean Drilling Research/International Ocean Discovery Program (ECORD/IODP) reports the same type of need:

Genealogical analysis of scientific ideas would “bring a lot”. One could also better know and trace the French participation: two French researchers per expedition, this means that a dozen French scientists embark each year on IODP expeditions and then “interact with about a hundred of their colleagues” to process the data from the campaigns. One onboard scientist per year has a knock-on effect on 70 to 100 researchers concerned in one way or another by his approach. The genealogy of the communities in question would undoubtedly be interesting and profitable for the work of the RI.

¹⁰ The French National Institute of Nuclear and Particle Physics.

¹¹ The French National Institute of Social Sciences and Humanities.

Two other examples: ICOS France (Integrated Carbon Observation System) underline

It is certainly necessary to follow the current developments in research on new metrics tools, and, at the same time, make an in-depth analysis of the uses of the RI. In particular, it will be necessary to assess how metrics actually contribute to the scientific options chosen by public policy. In this way, it will be possible to evaluate the precise contribution of science to public policy actions, as in the case of “evidence-based policies.”

The coordinators of the French contribution to the European Southern Observatory (ESO) and (Stocker, Darroch et al., 2020) point out that they have “no current practice of text mining or semantic analysis” and admit that there is, “on the other hand, obvious scientific interest, and it is necessary to follow the advances in the corresponding fields of STI research.” The question of resources is raised by the French oceanographic fleet (FOF): “We are ready to develop sharing with other large RIs, particularly in the field of climate. But on condition that we have the associated resources.”

Based on these and similar findings of the RI survey, the report makes three main, structuring recommendations:

- A. to organize the traceability of the RI results
- B. to build a catalog of shared strategic indicators (publimetrics)
- C. to create a network of these new metrics

These general recommendations are broken down into 15 detailed issues. We present the key points here, as summarized in the report (Egret & Fabre, 2019).

3.2. Recommendation A: Organize the Traceability of Results

A first result of the study is the identification of those areas in which it is important to organize or enhance the monitoring and overall traceability of data and publications resulting from the use of the infrastructure. Therefore, recommendations 1 to 5 of the report have in common the requirement of traceability:

1. Generalize the use of DOI and global traceability
2. Harmonize the main performance indicators (domains, partnerships, equipment)
3. Harmonize the terminology (classification) of research areas
4. Develop new metrics for emerging research fields and monitor the genealogy of ideas
5. Develop open science metrics for publications and data

This first group of recommendations is particularly sensitive for TGIRs, which, unlike university or research institutions, lack visibility in the large bibliometric databases referencing scientific production.

The actors concerned by the recommendations are the persons in charge of the TGIR who must define, in an operational way, the contours of their scientific production: Indeed, this is most often not restricted to that of their teams but must also include that of their users, or even consider more broadly the production of knowledge that has directly benefited from the existence of the infrastructure.

The publishers of large databases are also concerned by these recommendations, who may seek to include specific metadata for instruments and infrastructures, and to develop the nomenclatures of research fields.

3.3. Recommendation B: Catalog Shared Strategic Indicators

The second group of recommendations is based on the current practices and expectations regarding metrics of the RI activities (performance) and their scientific impact.

The main recommendation is the drafting of a Guide with recommendations and best practices for the use of “publimetrics” (i.e., metrics of the scientific impact of the RI output [publications, data]). The framework of this Guide is drawn by the following list of recommendations:

1. List the rules for identifying publications and reaffirm the requirement of an explicit mention of the RI (affiliation)
2. Collect the shared scientific impact indicators based on the prior establishment of a Guide for publimetrics
3. Design an architecture of metrics practices by major purposes and build a typology of current metrics practices
4. Participate in a global modeling of the uses of publimetrics at a European scale
5. Specify the organization and standards of publimetrics services

The production of this Guide aims to encourage the pooling and wide dissemination of good practices, the use and quality of which should be tested at national and European levels. Such a collection of recognized and recommended standards, coproduced with the large RIs, will also facilitate the construction of relevant and flexible digital architectures, appropriate for each infrastructure experiencing the need to complete its digital master plan. To meet the needs of research communities in terms of scientific impact metrics, the diversity of needs and the pluralism of practices must be recognized and supported: These are among the first lessons learned.

This second group of recommendations concerns the deployment of shared indicators. Here again, the main actors concerned are those in charge of the TGIR, but also the supervisory bodies (research organizations, ministry) who will seek to use these indicators in the service of strategic reflection. Finally, these developments must take into account the European and international context in which the TGIRs are deployed and be carried out jointly with the partners of the other countries concerned.

3.4. Recommendation C: Create a Network of Shared Metrics

How to get there? In the short term, a national approach to measure scientific impact of RIs should be set up to contribute to the networking of all the identified and adopted standards and practices. This approach, initiated with the large RIs, should also be developed in line with the whole Higher Education & Research community (i.e., universities and research organizations). The following work directions have been identified:

1. Display the reference charters and support large RIs in their efforts to adhere to international declarations of good practice for the evaluation of scientific results
2. Develop new metrics to support scientific foresight
3. Consolidate the scientific and professional deployment of publimetrics
4. Initiate a national metrics orientation approach
5. Set up a first experimentation process with a few large RIs

This third group of recommendations aims to promote a national network with the know-how and skills for the implementation and monitoring of tracers and indicators of scientific impact. The actors directly concerned are therefore, in addition to those responsible for the

TGIRs (and potentially for the RIs), the national research organizations (such as the CNRS, the CEA, and the IFREMER), the ministerial authorities, as well as the national evaluation and control bodies.

The publimetrics guide, mentioned in recommendation B, would be the means of bringing about the networking of metrics practices common to RIs, universities, research organizations, and other academic institutions.

It can be recalled on this point that, in 2016–2017, on the initiative of the CNRS Department of Scientific and Technical Information (DIST), and in association with the information professionals of Couperin (French academic library consortium), ADBU (Association of academic library directors) and EPRIST (Association of STI directors of research organizations), the higher education (HE) and research institutions had taken the initiative to assess the feasibility of networking the digital objectives and practices of scientific work, particularly in terms of metrics and analysis of scientific publications (Centre national de la recherche scientifique [CNRS], 2017)¹². This former study has shown the potential synergies of resources and projects that can be expected from networking all the approaches, based on the stronger and more detailed recommendations obtained in the survey of large RIs.

4. STANDARD FORMATS AND METRICS

Obviously, there is a growing interest and demand for the assessment of RI by the RI management, and also by funding bodies, research organizations, and authorities (Stocker et al., 2020). The French publimetrics initiative reveals different dimensions of such an approach, including the scientometric evaluation of the RI performance in terms of output and impact as well as the discovery of emerging research trends and the assessment of partnerships, communities, and knowledge production. Large RIs have importance in terms of national and international research strategy, and they need significant, recurrent long-term funding; for both reasons, the French initiative recommends a shared, concerted and mutualized approach to evaluation, based on flexible standard metrics.

In fact, as the French study shows, many infrastructures already do some kind of assessment, often without appropriate tools or models, specific and not standard, and not interoperable. The published projects in the field of research information management show that CRISs, with their standard data models, may be an option for the assessment of RI. Yet, as mentioned above, CRISs are generally designed for the evaluation of research institutions and organizations, not of infrastructures, which are usually considered and assessed by such systems as part of institutional resources, similar to other facilities, services, and equipment. Therefore, the following section analyzes how the main standard CRIS format (i.e., the Common European Research Information Format [CERIF]), and the new German Research Core Dataset (KDSF/RCD) meet the requirements of the present standard information and metrics for RI. Our focus is on the mapping of RCD attributes, RCD entities, and CERIF elements and compare this information with the recommended requirements (metrics) of the French publimetrics initiative. Do RCD and CERIF provide an appropriate solution for the need for evaluation of RI? Are they compliant with the publimetrics recommendations?

4.1. CERIF

Developed with the support of the European Commission and recommended for use by the EU member states, CERIF¹³ is a generic and standard model for organizing and exchanging research

¹² See the COPIST reports at <https://adbu.fr/les-etudes-du-copist-catalogue-doffres-partagees-en-ist/>.

¹³ For more information see <https://www.eurocris.org/cerif/main-features-cerif>.

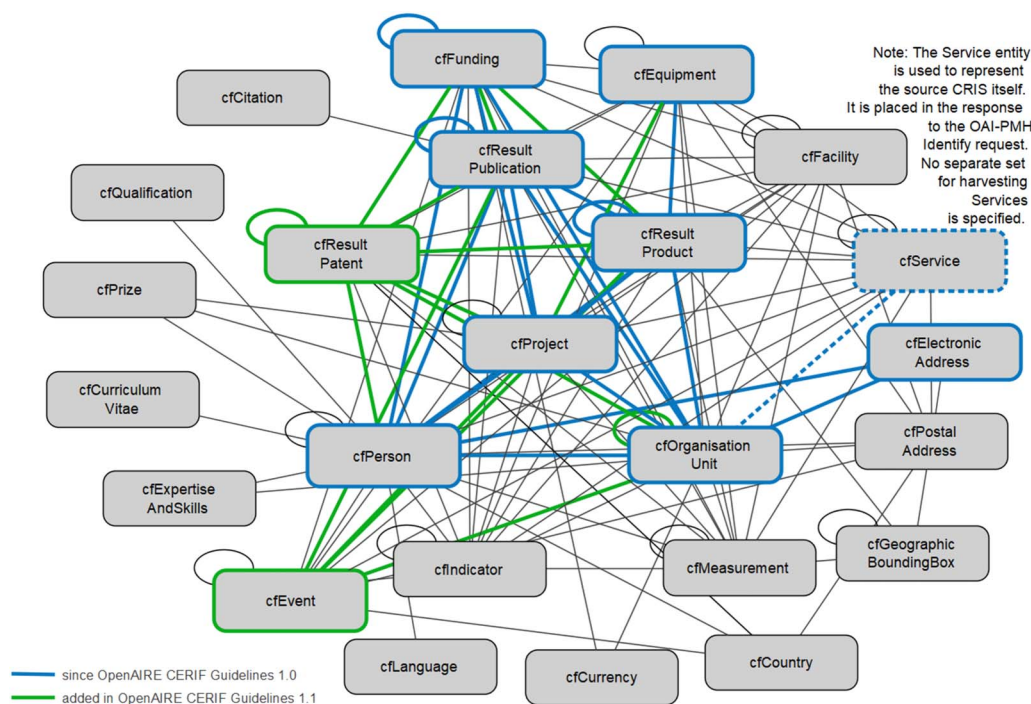


Figure 2. CERIF data model (source: OpenAIRE¹⁴).

information, the research domain and their relationships to each other on conceptual, logical, and physical levels. CERIF is intended to serve as a model for homogeneous access to heterogeneous data systems and as a definition of a data exchange format. The aim of CERIF is to serve as an interoperability level between the digital infrastructure and the research data, and to promote integration and exchange through standardization.

The CERIF data model includes persons, organizations, their projects, funding, and generally everything that arises from or is connected to the research process. At the very heart of the CERIF model are three interconnected core elements: persons, organizations, and projects; all the other elements—outputs, activities, metrics etc. and on another level, identifiers, geographical origin, addresses etc.—are connected with these elements through the semantic layer, in a rich, highly complex but standard network of relations (Figure 2).

We will not, in this context, describe and comment the CERIF data model in detail. Relevant for our study is the fact that the CERIF data model contains three infrastructure entities (i.e., facility, equipment, and service [Figure 3]), with semantic links to all base entities (project, person, organization unit) and result entities (publication, patent, product) and to some second level and link entities, such as funding, event, postal address, measurement, and indicator (Dvorak, 2013).

This data model allows a flexible description (multilingual fields for name, description, and keywords) and assessment of RI and bears the potential for specific extensions, especially for identities (Jörg, Höllrigl, & Sicilia, 2012), classification, and typologies, which may be added and stored in the semantic layer of the CERIF data model. Through the semantic interconnection of the different element levels, CERIF is able to handle RI identifiers, RI classifications and/or

¹⁴ OpenAIRE Guidelines for CRIS Managers https://openaire-guidelines-for-cris-managers.readthedocs.io/en/latest/cris_elements_openaire.html.

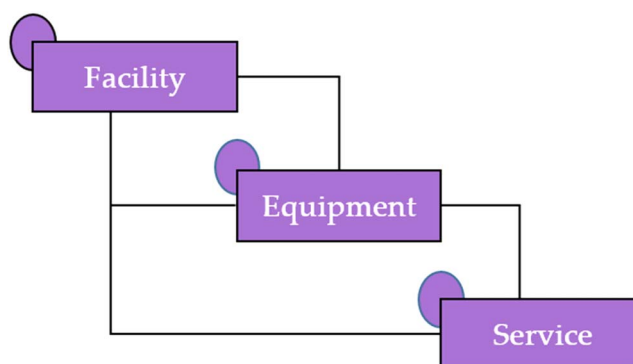


Figure 3. The CERIF data model infrastructure entities (source: CERIF 1.3 Full Data Model¹⁵).

typologies, and an RI directory, and to link specific outcome (result) data such as publications and research data sets.

The OpenAIRE Guidelines for CRIS Managers define “equipment” as an “instrumentality needed for an undertaking or to perform a service,” with one mandatory attribute (internal identifier) and six optional attributes or elements (type of equipment, acronym, name, identifier, description, owner), whereas “service” is defined as a research information management system (CRIS).

4.2. KDSF/RCD

More recently, the German Council of Science and Humanities has funded the development and promotion of the Research Core Dataset (KDSF/RCD)¹⁶, which describes information on research activities in a standardized form (Azeroual, Saake et al., 2019a; Biesenbender, 2019; Biesenbender & Herwig, 2019). This should enable quality-assured research activities for research reports to be compared with little effort and be used multiple times (Azeroual, Schöpfel, & Ivanovic, 2020). The goal is to provide a standard for Germany; the target groups for this are universities and nonuniversity research institutions. As there has been no standardized recording of research activities by institutions in Germany up to now, the RCD standard is intended to contribute to the standardization of research reporting. According to the RCD, research information in the areas of researchers employed by the institutions, young researchers, third-party funded projects, patents and spin-offs, publications, and RIs are to be collected.

These are converted into so-called core data and their characteristics and aggregation measures on the basis of existing definitions and standardization, including CERIF; the mapping between RCD and CERIF shows that RCD is a specialized version of CERIF. The implementation of the RCD standard is supported by the provision of a technical data model based on CERIF in XML format, which describes both basic and aggregate data formats and their respective relationships. The basic data model corresponds to the objects, the description of the objects, and the relationships and properties. The aggregate data model only defines the core data, without characteristics or specializations. Further details about the RCD specification (version 1.0) and the RCD XML schema can be found publicly on the RCD website. Figure 4 shows the semantic linking of the RCD areas as an Entity Relation Model (ERM). This contains the objects on which the specification is based, their attributes, and the relationships between them.

¹⁵ https://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf.

¹⁶ In German: Kerndatensatz Forschung (KDSF). For more information see <https://kerndatensatz-forschung.de/>.

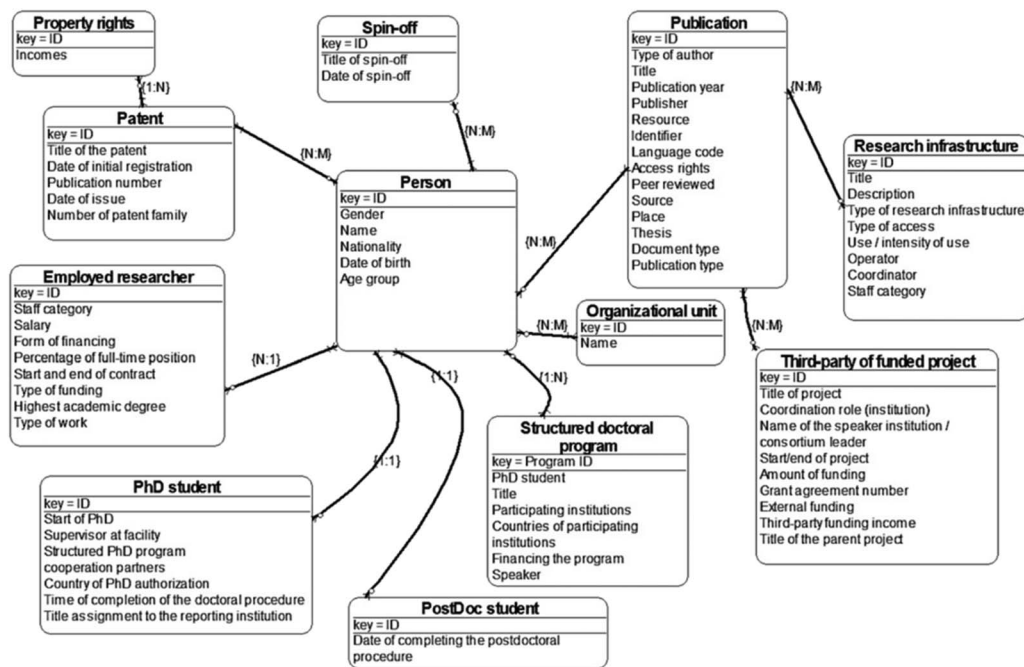


Figure 4. RCD data model (Azeroual & Herbig, 2020).

The German Council of Science and Humanities points out that the RCD is only a recommendation and not an obligation for universities and nonuniversity research institutions, which should improve and not replace the existing recording of research activities in the institutions. Rather, the RCD standard is intended to remove ambiguities in the collection of data and thereby improve the quality of it. Relevant for our study is the fact that the RCD data model includes RI (*Forschungsinfrastruktur*) as core data defined as

large/costly instruments, resources or service facilities for research in all scientific fields, which are characterized by at least supraregional importance for the respective scientific field as well as by a medium to long-term lifetime (more than 5 years) and are available for external use for which access or use regulations have been established¹⁷.

The RCD data model allows free title and description for the core data RI and provides semantics on different levels and for different elements:

- Operator: organizational unit
- Operating personnel: employer/employee
- Coordinator: organizational unit
- Use: use/intensity of use
- Publication: publication
- Type: type of RI
- Access type: type of access

¹⁷ See https://kerndatensatz-forschung.de/version1/technisches_datenmodell/index.html#http://kerndatensatz-forschung.de/owl/Basis#Forschungsinfrastruktur.

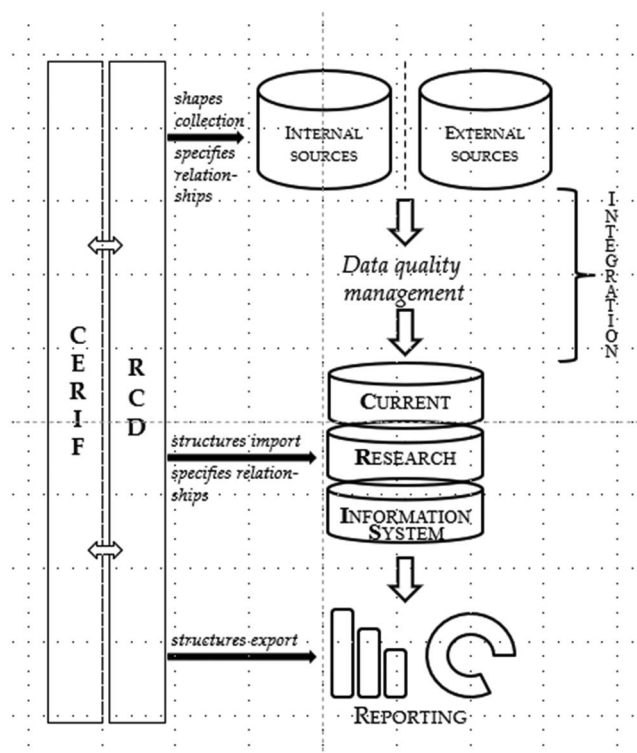


Figure 5. Use of CERIF and RCD in CRIS (Azeroual & Herbig, 2020).

4.3. Mapping CERIF and RCD

The RCD is compliant with the European CERIF format, and the RCD team provides a mapping between CERIF and RCD, to enable the exchange between different CRIS (Azeroual & Herbig, 2020)¹⁸. The RCD core data “Forschungsinfrastruktur” (RI) is mapped against the CERIF infrastructure entity “equipment” but not to the other entities “service” and “facility.” In comparison, CERIF appears more detailed, complete, and flexible for the description and assessment of RI than the German RCD.

RCD and CERIF serve as guidelines for scientific institutions that want to represent RCD and CERIF in their CRIS systems. Implementation can take place at both institutional and CRIS provider level. Both cases can be observed in institutions. The XML schema from CERIF and RCD can be used as a data source before importing into CRIS and/or as an export format to simplify reporting (Azeroual et al., 2020). The use of CERIF and RCD in CRIS can be illustrated using Figure 5.

The figure shows how CERIF and RCD can be used in institutions and offers institutions the opportunity to improve the quality of research information before it is integrated into CRIS.

The data quality is somewhat dependent on the standard application and this will likely improve the data quality. A standardized data model such as CERIF and RCD is an essential prerequisite for data management in terms of monitoring and strengthening data management

¹⁸ See https://kerndatensatz-forschung.de/version1/technisches_datenmodell/Mapping.html.

Table 1. Compliance of CRIS data models with French publimetrics recommendations

#	Recommendation	Compliance with CRIS data models	Comment
A1	ID of results	CERIF, RCD	Standard attribute
A2	Performance indicators	CERIF	Semantic layer (classification, typology...)
A3	Scientific domains	CERIF	Semantic layer
A4	Emergent research fields	CERIF	Semantic layer, attributes
A5	Open science	CERIF	Semantic layer, attributes
B1	RI affiliation	CERIF	Semantic layer
C2	Scientific foresight	(CERIF)	Reporting
C5	Experimental approaches	CERIF, RCD	Reporting

in institutions. This enables the introduction and permanent quality assurance in institutions as an overarching goal for research information (Azeroual & Herbig, 2020).

In the field of infrastructures, this means that if the data on RI are created directly in the RI-CRIS, care must be taken to index the RI in an appropriate way (with identifier, classification, etc.) and to link the RI to each relevant element (publication, data, patent, domain, person, etc.). If the information on RI is ingested from other, internal or external sources, such as repositories, bibliographic databases, or RI systems, care must be taken to control the data quality and to cleanse, enrich, and standardize the integrated data for further RI assessment.

4.4. The Compliance of CERIF and RCD with the Publimetrics Report

The French report makes 15 recommendations for the development of scientometric assessment of RI, summarized above (Egret & Fabre, 2019). How can research information systems cope with these requirements, and to what extent are the standard format CERIF and the German RCD consistent with them? Table 1 provides some elements.

Some comments. First, some recommendations have been excluded because they are not really relevant for research information management systems. In particular, recommendations B2–B5 regarding the usage of metrics (best practices, etc.) will contribute to the development and design of CRIS reporting functionalities, downstream of the ingestion and processing of research information (cf. Figure 4) but should have no impact on the data model itself. Also, recommendations C1, C3, and C4 on networking appear less relevant for data models; even if CRIS will produce useful information and thus support the implementation of a national or international strategy of RI metrics.

4.4.1. A1 Generalized use of unique identifiers for publications

Both CERIF and RCD include identifiers as an attribute of the entity publication. The CERIF attribute ID is for local identifiers, whereas for persistent identifiers such as DOI the link to the CERIF FedID entity should be used.

4.4.2. A2 Harmonization (convergence) of performance indicators (domains, partnerships, equipment)

CERIF handles standard or controlled vocabularies in the semantic layer (classification...). Relationships are established by identifiers of persons, organizations, or projects (attribute column), and fractions are indicated in the classification column, where each value belongs to a scheme.

4.4.3. A3 Harmonization (convergence) of scientific domains

CERIF supports controlled terminology in the semantic layer.

4.4.4. A4 Development of new metrics for emergent research fields

CERIF appears flexible enough to represent and report new indicators, based on semantic relations between entities and attributes and on measurement extensions, elaborated on infrastructure entities and semantics.

4.4.5. A5 Indicators of open science (open access publications, open repository deposits, paywall publications)

CERIF can handle this as semantics and attributes of the result entity publication.

4.4.6. B1 Generalization of RI affiliation

CERIF would represent this as a semantic link between a person (author), an organization (OrgUnit), and an equipment, facility, or service (infrastructure entity).

4.4.7. C2 Development of new metrics for scientific foresight

Depends on the development. Research information systems would at least be able to produce useful information for such new metrics. The CERIF model appears flexible enough for the definition of new metrics, with the entities Indicator, Metrics, and Measurement. Moreover, there is a flexible semantic layer and links between almost all CERIF entities, which can be classified and time framed using the startDate and endDate attributes of link entities.

4.4.8. C5 Experimental approaches

Both data models can handle information about the number of publications, scientific domains, impact metrics (citations), international partnerships, and open science-related metrics (open access publications).

5. CONCLUSION

RIs are facilities, equipment, and services needed by the scientific communities from all disciplines; they provide high-performance equipment in a high-level scientific environment. Because of their strategic importance, and also because of their need for significant, recurrent long-term funding, there is a growing demand for the monitoring and assessment of their performance in terms of research outcomes (publications, data, patents, etc.). A couple of national and international projects from the last decade show that research information management systems, with their standard data models, formats, and procedures may be an option for RI assessment. They also do appear to be consistent with requirements and recommendations of a shared approach to RI outcome and impact metrics, as suggested by the French publimetrics initiative.

Even if the research information management systems are generally designed for institutions, not infrastructures, their data models and, in particular, the main standard format CERIF, would be able to handle RI-related data and to produce relevant indicators for the reporting, monitoring, and assessment of RI performance. Also, they appear flexible enough to cope with the large diversity of RIs and with new metrics. Obviously, the major issue is not system or format but standards, in particular unique RI identifiers, standard classification, and names. On the occasion of the 2020 VIVO conference, de Castro from euroCRIS recently summarized some of the main projects on the identification of research equipment and facilities and recommended a

significant coordination effort (...) at an international level to raise and share emerging best practice case studies, since research is a deeply international endeavour and research facilities used in international projects may be based in any of the partnering countries¹⁹.

The evaluation of RIs is not just an academic issue but represents a societal challenge. As the French Cour des Comptes observed during a recent hearing on its survey on large RIs (Cour des Comptes, 2019; Rapin, 2019), the lack of analysis of the metrics and the impact of RI research is total and deprives society of the means to take full advantage of RIs in a knowledge-based economy:

Beyond the scientific evaluation, whose instruments are partly in place, and the socio-economic evaluation, which is still in progress, a large field of study has not been tackled to date: the evaluation of the positive qualitative externalities linked to the development of knowledge enabled by large RI. However, this impact should be taken into consideration, in order to guide public policies and ultimately to extend the reflection on risks, natural or medical, for example, and the improvement of living conditions²⁰.

Above all, more mutual understanding and coordination between RI management and CRIS development seems required to address this challenge.

As long as RIs have their own specific performance indicators, produced with their own specific systems and for internal use only, it will be difficult to harmonize or consolidate these metrics to assess the overall performance of the different RIs, which is necessary for the development of a reasonable national or international policy. The French publimetrics initiative provides a strategy on how to progress on the way to further harmonization and consolidation of shared and common RI metrics. Research information management systems or CRIS, designed to support research institutions in the provision of funding information and reporting, in aggregating references for research outputs, and in producing indicators and assessment (De Castro, 2018), bear the potential to contribute to this strategy. They have proven their worth in complex research environments, they are based on standards, and they consider the issue of data quality as a critical factor of success (Azeroual, Saake et al., 2019b).

Moreover, these systems would also contribute to a better understanding of scientific discovery and knowledge. At the crossroads of information sciences and bibliometrics, research is advancing towards the construction of “global” traceable document paths (Cabanac, Frommholz, & Mayr, 2020): In this sense, navigation between all the databases accessible

¹⁹ Persistent identifiers for research instruments and facilities? June 25, 2020 <https://www.eurocris.org/blog/persistent-identifiers-research-instruments-and-facilities>.

²⁰ Sénat, Commission des Finances Audition TGIR du 17 juillet 2019, Exposé de Sophie Moati, présidente de la troisième chambre, http://www.senat.fr/rap/r18-675/r18-675_mono.html.

on the Web is recognized as possible (Brickley, Burgess, & Noy, 2019). Furthermore, it is essential for the progress of routes and maps that information search behaviors are modeled and that the semantics of the documentary choices made are stabilized by a solid “topic modeling,” based on a “topic analysis-based approach” built through innovative and exhaustive methods (Tsatsaronis, 2020). The search for these solutions is encouraged by a context of rapid and diverse editorial changes, open to innovation (Conrad, Richardson, & Rinehart, 2020). These evolutions lead directly to the creation of tools for comparing navigation routes; in the words of Atanassova, Bertin, and Mayr (2019), it is necessary to produce “annotated corpora and shared evaluation protocols to enable the comparison between different tools and methods.”

In this environment of query “paths” under construction, a common requirement towards more traceability appears. The paths allow discoverers and users of science to represent their path of hypotheses, discoveries, and ideas, in a more readable and traceable way, through a structured sequence of all published scientific results, based on valid analyses of new maps of documentary choices (Aria & Cuccurullo, 2017). These mappings display their results using new ergonomic and user-friendly tools: This vision is nothing less than the current grail of industries contributing to the exploitation of scientific documentation. A dynamic global offer is thus being developed, with the slogan: “solving the problem of problem solving.”²¹ Some research institutes and infrastructures, such as the European Bioinformatics Institute (EMBL-EBI²²) already do this discovery mapping, which constitutes an initial response to the needs expressed by most large RIs. Improved international standards and cooperation for such marking out of the routes (i.e., a global markup of open routes, such as sea or land air routes) would ensure the scientific integrity of navigation choices and their coherent sharing, and would optimize navigation in digital scientific databases. RI evaluation with research information management systems could be an opportunity for further progress.

Alongside the current community, domain, and institutional platforms, new multiactor, agent, and object infrastructures are now emerging, using a combination of computing and analysis resources to carry out relevant data groupings on a very large scale. The Directory of Research Information Systems²³ shows a broad and structured pool of research information systems, which can enhance research intelligence and contribute to the notion of knowledge infrastructure, as a place for sharing and experimenting with RI publimetrics and for the preparation of what the National Academies of Sciences had called some years ago *The future of scientific knowledge discovery in open networked environments* (Uhlir, 2012). In this dynamic and strategic environment, international synergy and cooperation between the different stakeholders and projects from the communities of RIs (such as the ESFRI working group on the monitoring of RI performance with Key Performance Indicators or the JISC equipment data project), euroCRIS, research information management systems, research organizations, and funding bodies would be extremely useful for the development of relevant standard indicators for the reporting, monitoring, and assessment of the performance of RIs, to meet the academic and societal challenge.

This evolution towards the construction of CRIS, and then of platform networks, operates in the following directions of pooling of results and resources:

- Share scientific results on scientific themes common to several RIs: The French Publimetric Survey recorded this objective in a majority of complementary TGIR around a global

²¹ See <https://www.lens.org/>.

²² See <https://www.ebi.ac.uk/services>.

²³ DRIS <https://dspacecris.eurocris.org/cris/explore/drisc>.

scientific objective (examples: climate change sciences creating an articulation of interest for the structured exchange of data between glaciology, analysis of marine temperatures, traces of carbon, meteorological conditions, etc.).

- Share the technical and scientific resources and practices of the same type of equipment between and in TGIR (astronomy, synchrotron radiation, mainframe computers, large interdisciplinary scientific analysis networks, etc.). As such, TGIRs already practice many ULAB-type procedures to build their different interfaces (uses, experimenters, partners).
- Develop an expression of global interest in research approaches and scientific analysis tools: Even more than others familiar with macroevolutions in concepts and work directions, all TGIRs feel the need to federate approaches on the new semantics of discovery, on the itineraries and maps of knowledge renewed by the present innovations which surround the human sciences and the information sciences. The Bibliometric Survey has collected many testimonies in this direction. In this sense, the analysis graphs of scientific choice routes (Fabre, 2019) are present in our present reflection, as in that of most of the TGIR, which are ready to share experiences on innovative devices on the current orientations of the work of science, as reported in the survey.

The authors are engaged in further work to test the proof of concept of a bipartite Scientific Knowledge Graph (SKG), which was discussed as a research question in Fabre (2019). This SKG compares “routes” of networked users querying scientific information for discovery purposes and uses. Various studies in the literature (Aryani, Fenner et al., 2020; Brack, Hoppe et al., 2020) confirm that SKGs offer powerful means of representation of scholarly knowledge and assessment of research impact. This work will include applications of SKGs to RI uses.

ACKNOWLEDGMENTS

The authors are most grateful for insightful advice and comments from Guillaume Cabanac (University of Toulouse) and Dragan Ivanovic (University of Novi Sad) and for the constructive critics from two anonymous reviewers.

AUTHOR CONTRIBUTIONS

Renaud Fabre: Writing—review & editing. Daniel Egret: Writing—review & editing. Joachim Schöpfel: Supervision, Writing—original draft, Writing—review & editing. Otmane Azeroual: Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

The French Publimetrics study on the scientific impact of large IR was supported by the French Ministry of Higher Education, Research and Innovation (DGRI).

REFERENCES

- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. DOI: <https://doi.org/10.1016/j.joi.2017.08.007>
- Aryani, A., Fenner, M., Manghi, P., Mannocci, A., & Stocker M. (2020). Open science graphs must interoperate! In: L. Bellatreche et al. (Eds.), *ADBIS, TPD and EDA 2020 Common Workshops and Doctoral Consortium. TPD 2020, ADBIS 2020*. Communications in Computer and Information Science, vol. 1260. Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-55814-7_16
- Atanassova, I., Bertin, M., & Mayr, P. (Eds.). (2019). *Mining scientific papers: NLP-enhanced bibliometrics*. Frontiers Media. DOI: <https://doi.org/10.3389/978-2-88945-964-3>

- Azeroual, O., & Herbig, N. (2020). Mapping and semantic interoperability of the German RCD data model with the Europe-wide accepted CERIF. *Information Services & Use*, 40(1–2), 87–113. DOI: <https://doi.org/10.3233/ISU-200076>
- Azeroual, O., Saake, G., Abuosba, M., & Schöpfel, J. (2019a). Solving problems of research information heterogeneity during integration – using the European CERIF and German RCD standards as examples. *Information Services & Use*, 39(1–2), 105–122. DOI: <https://doi.org/10.3233/ISU-180030>
- Azeroual, O., Saake, G., Abuosba, M., & Schöpfel, J. (2019b). Quality of research information in RIS databases: A multidimensional approach. *Lecture Notes in Business Information Processing*, 353, 337–349. DOI: https://doi.org/10.1007/978-3-030-20485-3_26
- Azeroual, O., Schöpfel, J., & Ivanovic, D. (2020). Influence of information quality via implemented German RCD standard in research information systems. *Data*, 5(2), 30. DOI: <https://doi.org/10.3390/data5020030>
- Baginskaite, J. (2017). European Science Foundation: New projects and developments. *EuroCRIS Strategic Membership Meeting*, Bratislava, Slovakia. <https://dspacecris.eurocris.org/handle/11366/626>
- Bailo, D., Ulbricht, D., Nayembil, M. L., Trani, L., Spinuso, A., & Jeffery, K. G. (2017). Mapping solid earth data and research infrastructures to CERIF. *Procedia Computer Science*, 106, 112–121. DOI: <https://doi.org/10.1016/j.procs.2017.03.043>
- Beckers, P., Jägerhorn, M., & Höllrigl, T. (2012). Advances in sharing and managing knowledge about European research infrastructures. In *CRIS2012: 11th International Conference on Current Research Information Systems*, Prague, Czech Republic. <https://dspacecris.eurocris.org/handle/11366/91>
- Biesenbender, S. (2019). The Research Core Dataset: Standardizing the management and reporting of research information in the German science system. *EuroCRIS Strategic Membership Meeting*, Münster, Germany. <https://dspacecris.eurocris.org/handle/11366/1243>
- Biesenbender, S., & Herwig, S. (2019). The German Research Core Data Set (KDSF). *EuroCRIS Membership Meeting*, Espoo-Helsinki, Finland. <https://dspacecris.eurocris.org/handle/11366/978>
- Boldrini, E., Luzi, D., Nativi, S., & Pecoraro, F. (2014). Integrating CERIF entities in a multidisciplinary e-infrastructure for environmental research data. *CRIS2014: 12th International Conference on Current Research Information Systems*, Rome, Italy. <https://dspacecris.eurocris.org/handle/11366/193>
- Brack, A., Hoppe, A., Stocker, M., Auer, S., & Ewerth, R. (2020). Requirements analysis for an open research knowledge graph. In: M. Hall, T. Merčun, T. Risse, & F. Duchateau (Eds.), *Digital libraries for open knowledge. TPDL 2020. Lecture Notes in Computer Science*, vol. 12246. Cham: Springer. DOI: https://doi.org/10.1007/978-3-030-54956-5_1
- Brasse, V. (2012). MERIL: An e-infrastructure to connect research infrastructures. *10th EuroCRIS Strategic Seminar: "Horizon 2020 and Beyond"*, Brussels, Belgium. <https://dspacecris.eurocris.org/handle/11366/297>
- Brickley, D., Burgess, M., & Noy, N. (2019). Google dataset search: Building a search engine for datasets in an open Web ecosystem. *The World Wide Web Conference—WWW '19*, San Francisco, CA. DOI: <https://doi.org/10.1145/3308558.3313685>
- Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., ... Wright, J. (2017). *Research information management: Defining RIM and the library's role*. OCLC. <https://www.oclc.org/research/publications/2017/oclcresearch-defining-rim.html>
- Buhr, C.-C. (2014). Infrastructures for open, digital science. *CRIS2014: 12th International Conference on Current Research Information Systems*, Rome, Italy. <https://dspacecris.eurocris.org/handle/11366/319>
- Cabanac, G., Frommholz, I., & Mayr, P. (2020). Scholarly literature mining with information retrieval and natural language processing: Preface. *Scientometrics*, 125, 2835–2840. DOI: <https://doi.org/10.1007/s11192-020-03763-4>, PMID: 33223580, PMCID: PMC7670972
- Centre national de la recherche scientifique. (2017). *Catalogue d'Offres Partagées d'IST (COPIST). Analyse conjointe des résultats de l'enquête nationale auprès des Universités, des Grandes écoles et des Organismes de recherche sur le partage et la gestion des ressources en IST*. CNRS DIST. <https://www.science-ouverte.cnrs.fr/dist-dossiers/>
- Conrad, L. Y., Richardson, P., & Rinehart, R. (2020). Weathering changes and challenges to society publishing. *Learned Publishing*, 33(1), 2–3. DOI: <https://doi.org/10.1002/leap.1281>
- Cour des Comptes. (2019). *Le pilotage et le financement des Très Grandes Infrastructures de Recherche*. <https://www.ccomptes.fr/fr/publications/les-tres-grandes-infrastructures-de-recherche-tgir>
- Crompton, S., Matthews, B., Yang, E., Neylon, C., & Coles, S. (2012). Collaborative information management in scientific research processes. *IEEE 8th International Conference on E-Science*, Chicago, IL. DOI: <https://doi.org/10.1109/eScience.2012.6404478>
- De Castro, P. (2018). The role of Current Research Information Systems (CRIS) in supporting open science implementation: the case of Strathclyde. *ITLib*, 5, 21–30. DOI: <https://doi.org/10.25610/itlib-2018-0003>
- Dvorak, J. (2013). Research information: The euroCRIS context. *Use Science: Open Infrastructure to Foster Collaboration between Industry and Academia*, Riga, Latvia. <https://dspacecris.eurocris.org/handle/11366/41>
- Egret, D., & Fabre, R. (2019). *Mission "Publimétries". Apprécier l'impact scientifique des TGIR/OI "Comment mieux tracer les dynamiques du travail scientifique."* Ministère de l'enseignement supérieur, de la recherche et de l'innovation. <http://perso.obspm.fr/daniel.egret/Publimetries.pdf>
- Fabre, R. (2019). A "Searchable" Space with Routes for Querying Scientific Information. *BIR@ECIR, Proceedings of the 8th International Workshop on Bibliometric-enhanced Information Retrieval (BIR 2019)* (pp. 112–124). <http://ceur-ws.org/Vol-2345/paper10.pdf>
- Florio, M., Forte, S., & Sirtori, E. (2016). Forecasting the socio-economic impact of the Large Hadron Collider: A cost-benefit analysis to 2025 and beyond. *Technological Forecasting and Social Change*, 112, 38–53. DOI: <https://doi.org/10.1016/j.techfore.2016.03.007>
- Houssos, N., & Karaiskos, D. (2013). The CERIF-based EuroRIs-Net + Observatory. *EuroCRIS Membership Meeting*, Porto, Portugal. <https://dspacecris.eurocris.org/handle/11366/70>
- Ivanovic, D., Theodoridou, M., & Remy, L. (2018). A mapping tool for semantic interoperability of research information systems. *8th International Conference on Information Society and Technology*, Kopaonik, Serbia. <https://www.eventiotic.com/eventiotic/files/Papers/URL/89493fe6-5d45-4ebb-aba7-5d3ba7908bd9.pdf>
- Jörg, B., Höllrigl, T., & Sicilia, M. (2012). Entities and identities in research information systems. *CRIS2012: 11th International Conference on Current Research Information Systems*, Prague, Czech Republic. <http://hdl.handle.net/11366/102>
- Matthews, B. (2012). *Model of the data continuum in Photon and Neutron Facilities (PaN-data ODI. Deliverable D6.1)*. STFC. <http://pan-data.eu/sites/pan-data.eu/files/PaNdataODI-D6.1.pdf>
- Ministère de l'enseignement supérieur, de la recherche et de l'innovation. (2018). *Stratégie nationale des infrastructures de recherche*.

- <https://www.enseignementsup-recherche.gouv.fr/cid70554/la-feuille-route-nationale-des-infrastructures-recherche.html#fr>
- Organisation for Economic Co-operation and Development. (2019). *Reference framework for assessing the scientific and socio-economic impact of research infrastructures*. OECD Science, Technology and Industry Policy Papers, No. 65. **DOI:** <https://doi.org/10.1787/3f4ee43b-en>
- Puuska, H., & Rydman, W. (2018). Linking funding to research output and facilities through the Finnish Research Information Hub. *CRIS2018: 14th International Conference on Current Research Information Systems*, Umeå, Sweden. <https://dspacecris.eurocris.org/handle/11366/637>
- Rapin, J.-F. (2019). *Rapport d'information au nom de la commission des finances sur l'enquête de la Cour des comptes sur le pilotage et le financement des très grandes infrastructures de recherche*. Sénat. <http://www.senat.fr/rap/r18-675/r18-6751.pdf>
- Research Excellence Framework 2021. (2019). *Panel criteria and working methods (REF 2019/02)*. https://www.ref.ac.uk/media/1084/ref-2019_02-panel-criteria-and-working-methods.pdf
- Schöpfel, J., Azeroual, O., & Saake, G. (2019). Implementation and user acceptance of research information systems. *Data Technologies and Applications*, 54(1), 1–15. **DOI:** <https://doi.org/10.1108/DTA-01-2019-0009>
- Spang-Grau, I. (2019). Research Information in the selection process of Germany's universities of excellence. *EuroCRIS Strategic Membership Meeting*, Münster, Germany. <https://dspacecris.eurocris.org/handle/11366/1230>
- Stocker, M., Darroch, L., Krah, R., Habermann, T., Devaraju, A., ... Häggström, I. (2020). Persistent identification of instruments. *Data Science Journal*, 19(1), 18. **DOI:** <https://doi.org/10.5334/dsj-2020-018>
- Theodoridou, M., Patkos, T., & Doerr, M. (2016). Towards contextual awareness and interoperability in an enhanced VRE: mapping individual metadata schemata to CERIF. *EuroCRIS Strategic Membership Meeting*, Athens, Greece. <https://dspacecris.eurocris.org/handle/11366/574>
- Tsatsaronis, G. (2020). Metrics and trends in assessing the scientific impact. *BIR 2020 10th International Workshop on Bibliometric-enhanced Information Retrieval*, Lisbon, Portugal. <http://ceur-ws.org/Vol-2591/paper-02.pdf>
- Tzenou, G., & Bonis, V. (2016). European wide infrastructures and services on research information management: The RICH project. *EuroCRIS Strategic Membership Meeting*, Athens, Greece. <https://dspacecris.eurocris.org/handle/11366/571>
- Uhler, P. F. (Ed.) (2012). *The future of scientific knowledge discovery in open networked environments*. Washington, DC: The National Academies Press. **DOI:** <https://doi.org/10.17226/18258>
- Vancauwenbergh, S. (2019). Analyzing the research data infrastructure in Europe. *EuroCRIS Strategic Membership Meeting*, Münster, Germany. <https://dspacecris.eurocris.org/handle/11366/1225>