

# Advances in Sharing and Managing Knowledge about European Research Infrastructures

Paul Beckers<sup>a</sup>, Martin Jägerhorn<sup>b</sup>, Thorsten Höllrigl<sup>b</sup>

<sup>a</sup> European Science Foundation, <sup>b</sup> AVEDAS AG, Germany

## Summary

Research excellence requires excellent research infrastructures. Given the important role of research infrastructures for scientific discoveries and innovation, the European scientific community and policy makers have expressed their needs for a comprehensive database of research infrastructures in Europe. The European Commission funded MERIL initiative (Mapping of the European Research Infrastructure Landscape), which is coordinated by the European Science Foundation (ESF), aims to achieve an inventory of research infrastructures of European relevance and to make the information publicly available through an interactive online portal, which will be launched during the Autumn 2012. In this paper, we describe the initiative and discuss the driving forces, challenges and solutions for establishing a system which sustainably provides its stakeholders with the desired key information on Research Infrastructure of European relevance.

## 1 Introduction

Research Infrastructures (RI) are a vital pillar under the European Research Area (ERA). In order to provide policy-makers and scientists in Europe with data for analysis of and access to the current RI landscape, it is essential to produce a comprehensive database of existing research infrastructures with European relevance.

In the report ‘EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions’ (2009), the European Science Foundation (ESF) and the European Heads of Research Councils (EUROHORCs)<sup>1</sup> highlight the need to develop shared funding and exploitation of research infrastructures. A two-year mandate was given to the Deutsche Forschungsgemeinschaft (DFG), a member of both EUROHORCs and the ESF, to accomplish this action. In January 2010, the ESF launched a Member Organisation Forum, chaired by the DFG, to address this issue among its members and other key players in Europe. One of the activities entangles a continued updating of the inventory of national research infrastructures with European significance. The importance of RI as a corner stone of the ERA has also been endorsed in the ‘Europe 2020 Flagship Initiative Innovation Union’ communication of the European Commission (European Commission 2010).

As a result of the above drivers, the MERIL initiative (Mapping of the European Research Infrastructure Landscape) was started in October 2010 with the aim to achieve the desired comprehensive inventory of RI of European relevance and make the information publicly available through an interactive online portal.

---

<sup>1</sup> EUROHORCs (European Heads of Research Councils) has been dissolved in 2011 and a legal entity called ScienceEurope has been installed instead

## 2 About MERIL

MERIL faced many fundamental challenges at the beginning, and a set of key questions had to be answered in order to lay the foundation for a sustainable online portal of RI:

- What is meant by research infrastructures?
- Who are the stakeholders?
- What are the assets of an inventory of research infrastructures?

### 2.1 What is meant by research infrastructures?

Considering existing definitions from the European Commission and the European Strategy Forum on Research Infrastructures (ESFRI), MERIL established the following definition of Research Infrastructures: “A Research Infrastructure is 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.”

Furthermore, it is important in this context that a European Research Infrastructure fulfils certain quality criteria, i.e. it should:

- offer scientific and/or technological performance and support that should be recognised as being of European relevance
- offer access to scientific users from Europe and beyond through a transparent selection and admission process
- have a management structure

If the RI does not fulfil these criteria, it will not be included in the MERIL Portal. It is generally acknowledged that the aspects that determine the European relevance of an infrastructure may differ from one scientific domain to the other or from one particular type of RI to the other. Therefore a common understanding of what European relevance means, as well as criteria that can be objectively applied, needs to be developed. MERIL, the European Strategy Forum on Research Infrastructures (ESFRI) and the ESF Member Organisation Forum on Research Infrastructures are hence jointly working on a common understanding of ‘European relevance’.

A working definition of ‘European relevance’ provides the basis for further discussion. The working definition of European relevance contains the following elements:

- The RI should be planned and operated with the scope of being capable to attract, offer access to and provide support to a significant fraction of non-resident users from Europe and beyond on the basis of clear access and/or admission rules, and
- The RI should have a certain level of uniqueness and relevance in its scientific and/or technological performance as a single-sited research infrastructure or as part of a distributed research infrastructure, or network, and this uniqueness and relevance should be by the satisfaction of users after having been selected/admitted and having used the RI successfully.

## 2.2 Who are the stakeholders?

The stakeholders of the MERIL Portal are the scientific community, EUROHORCs, ESF, European Commission, ESFRI, European Association of National Research Facilities (ERF), EIROforum, and Ministries. Most of these are represented in MERIL's governance through its Steering Committee.

At the operational level, the scientific committees of the ESF have nominated experts to represent their scientific domains in the MERIL Science Advisory Board. Support from the scientific units at the ESF is provided in order to ensure coherence and liaison with the initiative.

## 2.3 What are the assets of an inventory of Research Infrastructures?

MERIL is a unique resource for the scientific community, policy makers and other communities such as research performing and research funding organisations. As the Member States of the European Union continue to develop a more co-ordinated and strategic approach to scientific research, it is important to have a clear picture of the research infrastructure landscape. A single source of information about all the principal research infrastructures throughout Europe can:

- Help the scientific community to identify and gain access to a wide range of resources, services and facilities.
- Allow policy makers to assess the state of research infrastructures throughout Europe to pinpoint gaps and possible duplications and make decisions about where best to direct funding.
- Encourage cross-border and cross-sector collaboration and more efficient use of resources within the European research community.
- Promote individual research infrastructures by raising their profile and fostering a greater sense of partnership across Europe.
- Contribute with this state-of-the-art analysis to the planning for future needs in cooperation with the European Strategy Forum on Research Infrastructures (ESFRI).

Collecting, storing, exchanging and updating quality assured information related to research infrastructures is of key importance to achieving these ambitions. The latest developments in information systems and technology make it possible to build sustainable and reliable knowledge management tools that support this. The MERIL portal will become such a tool for key data about research infrastructures of European relevance. To this end the MERIL Steering Committee has chosen the CONVERIS Research Information System developed by AVEDAS as its solution. Through customisation of CONVERIS the MERIL Portal meets the specific needs for collecting, quality assuring and presenting information on the research infrastructures.

One of the assets of the CONVERIS system is the fact that it is compliant with the Common European Research Information Format (CERIF). This lays the foundation for facilitating future exchange with other data sources on research infrastructures that would adopt the CERIF standard. In addition, MERIL as a leading initiative for sharing and managing information about research infrastructure in Europe is thus able to demonstrate the practicability and meaningfulness of CERIF, but also contribute effectively to improve CERIF for the handling of information on research infrastructure.

### **3 Challenges & possible solutions for establishing an information infrastructure for Research Infrastructure**

When building a tool to collect, share and manage quality assured information about research infrastructures one faces challenges on different levels. These challenges on the societal, organisational and technical levels, and the way MERIL copes with them are briefly described below.

#### **3.1 Societal challenges**

For the coming years, 2014-2020, the European Commission has identified the following six main areas of challenges:

- Health, demographic change and well-being;
- Food security, sustainable agriculture, marine research and the bio-economy;
- Secure, clean and efficient energy;
- Smart, green and integrated transport;
- Climate action, resource efficiency and raw materials;
- Inclusive, innovative and secure societies...

These challenges are global and the scale is such that they require research infrastructures, which are often beyond the resources available at national or regional levels. As such the global challenges are drivers for the identification, connection, operation and construction of both single sited and distributed research infrastructures, linked by high-speed networks for optimal sharing of data and resources.

To be instrumental in the process of identification, MERIL collects information about which Research Infrastructures consider themselves relevant in the light of tackling these challenges. Potentially this information is a first step towards creating new clusters or endorsing new initiatives of existing clusters that stimulate innovative approaches. Research Infrastructures, who are in the heart of the knowledge triangle (Expert Group on Research Infrastructures 2010) can foster the necessary innovation by attracting excellent scientists who develop new knowledge, by training and educating new generations of scientists and by spinning off enterprises based on new knowledge. Integration of knowledge within these clusters requires harmonisation and standardisation.

#### **3.2 Organisational challenges**

In order to ensure that the MERIL Portal becomes sustainable and a continuously updated, reliable source of information, it is a prerequisite that consensus is achieved about a set of common parameters. MERIL has achieved consensus among its stakeholders about a number of these parameters, including a list of scientific domains and RI categories to group the RIs; a definition of 'European relevant RIs'; the minimal set of data to be collected and cut-off criteria for qualifying RIs to be included in the portal. These achievements enable the collection of RI meta-data across the European countries in a harmonised manner.

In close consultation with the stakeholders of the MERIL initiative a set of eight Scientific Domains has been established to be used as backbone for the database and portal:

1. Information Science and Technology
2. Biological and Medical Sciences
3. Earth and Environmental Sciences
4. Physics, Astronomy, Astrophysics and Mathematics
5. Chemistry and Material Sciences
6. Engineering and Energy
7. Social Sciences
8. Humanities and Arts

In a similar way a list of 71 RI categories was established. Each Scientific Domain contains a subset of RI categories to classify each individual research infrastructure. Depending on the nature of the RI it may be relevant to more than one Scientific Domain and may fit more than one category. Each category has been described in a way that it is clear which facilities should be included. The portal will facilitate searches based on several entities that will be collected in the database. Figure 1 shows an example of a potential set of filters. The presented data are preliminary since the portal is still under construction and the full data set needs to be collected.

The screenshot shows the MERIL portal interface. At the top left is the MERIL logo with the tagline 'MAPPING OF THE EUROPEAN RESEARCH INFRASTRUCTURE LANDSCAPE'. Navigation links for 'FAQ | ABOUT | GLOSSARY | LOGIN' are at the top right. A search bar is located below the logo. The main content area displays '36 Research infrastructures' with a 'Sort by: Name | Acronym' option and 'Elements per page: 10 | 25 | 50 | 100' settings. A filter bar contains letters A through U and 'All'. The left sidebar lists filters for 'Scientific Domains' (Biological and Medical Sciences: 36), 'RI Categories' (Genomic, Transcriptomic Facilities: 16; Structural Biology Facilities: 6; Bio-informatics Resources: 3; BioSecurity Level-4 Laboratories: 2; Biobanks including Seed banks: 2), 'Countries' (France: 19; Germany: 4; Italy: 2; Netherlands: 2; Sweden: 2), and 'Status' (Being upgraded: 1). The main list shows entries like 'Animal Health Biosafety Research Center' (Acronym: CISA), 'BRENDA Enzyme information system' (Acronym: BRENDA), 'Center for Biomolecular Magnetic Resonance (BMRZ), Frankfurt/Main' (Acronym: BMRZ), 'Center for Systems Neuroimaging', 'Centre Risonanze Magnetiche, CERM', 'DNA Array Centre Nantes', 'European Bioinformatics Institute (EBI). N.B. The EBI is part of the European Molecular Biology Laboratory (EMBL)', 'European Centre for Training and Research on Highly Contagious Diseases (EUTRICOD)', 'European Clinical Research Infrastructures Network' (Acronym: ECRIN), and 'European Light Microscopy Initiative - Centre for Cellular Imaging' (Acronym: CCI). At the bottom, there are links for 'Create list as PDF | RSS 2' and '1/4' page indicator. The footer includes 'DISCLAIMER | CONTACT' and logos for 'European Research Infrastructure Consortium' and 'European Union'.

Figure 1. Example of filter page for the MERIL portal (under development)

Based on the analysis and prioritisation of the needs of users agreement has been reached about the following set of metadata to be collected for each RI (Table 1).

Table 1. Set of metadata to be collected in the MERIL database

Information	Definition
Name	The name of the RI for which data is provided
Acronym	Abbreviation used to reference the RI
Website	URL to the RI's web site.
Location	Position or point in physical space that the RI occupies on the Earth's surface or area
Scientific domain(s)	Widely-recognised field of study dealing with a specific discipline of science <i>A Research Infrastructure can be associated to one or more Scientific Domains, which all should be listed.</i>
Grand Challenge	Relevant to the societal grand challenges as listed by the European Commission; to be indicated by the RI operators
Type of RI	A classification given to an RI to further describe its characteristics
RI category	Further classification of RIs within scientific domains <i>A Research Infrastructure can belong to one or more RI categories, which should all be listed.</i>
Research Services at the RI	Service provided to users (access to a technology or a laboratory, user support, training...)
Equipment at the RI	Tool generally used by scientists to either perform an experiment or to take measurements and gather data; larger or more sophisticated equipment is generally called a scientific instrument.
Size	Total number of (national, European, International) users
Year of start of operation (i.e. Status and Status Date)	Status: life-cycle stage of the RI Status date: Year related to Status
Keywords	Words or concepts of great significance when describing the RI
Description of the RI	General description of the RI, can include RSS/social network addresses
Hosting Organisation	The name of the organisation or institution which is hosting the RI
Contact email	The email address used to contact the RI

This data set has been translated into a standard format for transferring data into the database. This will likely trigger the next step towards standardisation of data storage and connecting RI data. For this purpose the euroCRIS Task Group for CERIF has worked explicitly to enhance CERIF for the handling of information on Research Infrastructures. This work is further described in section 3.3 ‘Technical challenges’.

European relevance will be the cut off criterion for inclusion of RIs in the MERIL portal. A procedure for objectively checking this criterion is under construction, based on the working definition that is mentioned in section 2.1. Representatives of all organisations in the European countries, involved in the funding of RIs are consulted for this procedure to create a broad commitment. The process of the European relevance check will be formalised in order to validate new entries to the portal that will appear after the contractual end of the MERIL project.

As part of its deliverables the MERIL initiative developed the so-called ‘Common features of European Research Infrastructures’ to create consensus among the national funding organisations what the ultimate requirements for high quality research infrastructures should be in Europe. These standards are recommendations to be considered for future development of funding programmes.

In view of building a sustainable portal for quality assured RI metadata, which goes beyond the current EC funded contract, an adequate governance model need to be developed. Currently the different possible scenarios for hosting, maintenance and upgrading of the portal are being considered in consultation with the stakeholders represented in the MERIL Steering Committee. The governance model should describe the decision making process in case of necessary software updates or the implementation of additional features.

### 3.3 Technical challenges

The main technical challenge for the MERIL portal is to ensure a continuous collection and quality assurance of the desired information on the RIs. This is in particular a challenge since the RIs describe and structure their information in different ways and make use of different systems with different data models. To overcome the data heterogeneities, generally three different solutions could be applied (Hoellrigl et al. 2010):

- The RIs could adapt their local data models to a common one.
- Each RI could submit the data to MERIL in their own local data schema.
- The data model of the local RI could be mapped to a common intermediary data model.

Due to political as well as efficiency reasons option 1 and 2 are in the case of MERIL practically not feasible. In federated scenarios it is quite common to make use of an intermediary scheme, therefore, MERIL also decided to make use of option 3. In general, this option can be implemented automatically, for instance, based on a middleware service such as a connector provided by CONVERIS or manually by mapping the local data schemes to the intermediary data schema. In case of MERIL, CERIF is used as an intermediary schema.

One of the reasons for the selection of CONVERIS was its CERIF compliance, and secondly its flexibility. ESF not only wants to benefit from the mature and powerful CERIF data model, but in addition has specific requirements that CONVERIS fulfils. Thus, CONVERIS is compatible with CERIF on two levels: CERIF is used natively as far as possible in the local data model and CONVERIS both imports and exports valid CERIF XML files to facilitate CRIS interoperability.

In the course of the MERIL initiative it has even been possible to actively contribute to the euroCRIS community, as the requirements concerning the data model have been directly taken up by the CERIF Task Group.

Another major technical challenge is the way the data is accessed (browsed, filtered, searched) and rendered (visualisation, numerical aggregation, statistical analysis...). In this area, the development of the MERIL portal provides a context for enhancing CONVERIS features and CERIF data access in general:

- one example is in relation to some of the recently enhanced CERIF attributes such as the ones related to geo-location and the ability to display the information on geographical maps;
- another example under investigation is the ability to link internally stored Research Information with externally indexed data, as available in the individual Research Infrastructure websites.

Finally, a last example under investigation is the use of user-friendly visualisation tools to visually navigate through the classifications (Scientific Domains, RI Categories...).

Hence, for the continuous data collection MERIL will recommend each RI to deliver its information in CERIF-XML, and where needed and possible collaborate with the RI to enable this.

## **4 The Future: Sustainability and leveraging the results?**

The main challenges society is facing today require the collection, storage, integration and analysis of huge amounts of data. Together with creating the necessary physical e-infrastructure networks that connect different parts of the world and cross disciplinary borders, global standards for data storage and sharing are essential elements to optimise the use of data that science produces.

The MERIL portal will be an e-infrastructure that provides meta-data about the Research Infrastructures of European relevance. This should enable the identification of potential partner facilities that in the future could be further leveraged in that they share their data in a linked open data world (Jörg et al. 2012). Supporting mechanisms are currently under development. The newly proposed Strategy programme (2014-2020) of the European Commission, Horizon 2020, will support the implementation and operation of ESFRI and other world-class research infrastructures (Geoghegan-Quinn, 2012). It will integrate national research infrastructures and facilitate transnational access to these infrastructures, as a continuation of the very successful FP7 Integrating Activities. Horizon 2020 will also support the development, deployment and operation of e-infrastructures; foster the innovation potential and the human resources of research infrastructures; and reinforce European policy and international cooperation.

Horizon 2020 offers an integrated approach. It will allow linking the activities on research infrastructures with other funding to increase excellence, as well as tackling the societal challenges and boosting industrial leadership. While research infrastructures are critical for basic science, they can also act as an anchor for technology and innovation development and provide the focal point for developing innovation clusters. Such clusters stimulate knowledge transfer from academia to business – from research to commercialisation - bringing together universities, research centres, enterprises and regional authorities.

Since it is obvious that the European Commission administers only a minor part of the available funding for Research & Technological Development, the European Commission, national governments and national research organisations should create appropriate platforms for the development, maintenance and use of both large and medium scale research infrastructures. As stated by the newly founded Science Europe organisation in its reply to the consultation on the European Research Area Framework (Science Europe 2011), National and European efforts need to be developed in a complementary way in order to be successful. While European support enables transnational sharing of resources, national funding can provide underpinning investment and provide for national access. The different sources of funding target the same research infrastructure in a mutually reinforcing way. Synergies between European and national actions are therefore considered as very important. Management and training of scientists, as well as technology transfer at Pan-European research infrastructures is often of a very high quality. Emerging RIs can benefit from these experiences and appropriate measures to improve the management of research infrastructures are welcomed. Only well-managed research infrastructures will be in a position to provide professional support to the users including on issues like training or data analysis.

Taking into account these ambitions, the main challenge will be how to increase the accessibility of the data. The euroCRIS community could use this opportunity by setting the scene for optimisation of the interoperability of generated data. The Research Infrastructure as main actor in the generation of knowledge and innovation should be smartly linked to the scientific output they yield. The global context should definitely be kept in mind in this effort.

The MERIL portal will be developed into a reliable and up to date source of information about Research Infrastructure metadata as a first step. However, this provides the basis for an even wider use. Future needs and potential services should therefore continuously be further explored in support of the wider community of RI users, both in science and policy. How to create smart links between the RI metadata and scientific output remains to be discussed.

## References

- ESF/ EUROHORCs Report (2009): EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions
- European Commission Communication, SEC (2010): 1161 Europe 2020 Flagship Initiative Innovation Union
- Expert Group on Research Infrastructures (2010): A vision for strengthening world-class research infrastructures in the ERA, Luxembourg: Publications Office of the European Union, ISBN: 978-92-79-14214-7
- Hoellrigl, T.; Dinger, J.; Hartenstein, H. (2010): FedWare: Middleware Services to Cope with Information Consistency in Federated Identity Management, In: *Proceedings of the Fifth International Conference on Availability, Reliability and Security (ARES 2010)*, Krakow, Poland, February, 2010, pp. 228-235.
- Jörg, B.; Ruiz-Rube, I.; Sicilia, M.-A.; Dvorak, J.; Jeffery, K.; Hoellrigl, T.; Rasmussen, H. S.; Vestdam, T.; Garcia Barriocanal, E.; Engfer, A. (2012): Connecting Closed World Research Information Systems through the Linked Open Data Web. In: *International Journal of Software Engineering and Knowledge Engineering (IJSEKE)*, Volume 22, Consuming and Producing Linked Data on Real World Applications. June 2012.
- Geoghegan-Quinn M. (2012): European Commissioner for Research, Innovation and Science ‘The importance of research infrastructures for Europe’ Speech at ICRI2012, Copenhagen, 21 March 2012
- Science Europe (2011): Science Europe Contribution to the Public Consultation on the European Research Area Framework

## Contact Information

Paul Beckers  
European Science Foundation  
1 quai Lezay-Marnésia  
67000 Strasbourg  
France  
[pbeckers@esf.org](mailto:pbeckers@esf.org)

Martin Jägerhorn  
AVEDAS AG  
Waldstr. 65  
76133 Karlsruhe  
Germany  
[m.jagerhorn@avedas.com](mailto:m.jagerhorn@avedas.com)

Thorsten Höllrigl  
AVEDAS AG  
Waldstr. 65  
76133 Karlsruhe  
Germany  
[t.hoellrigl@avedas.com](mailto:t.hoellrigl@avedas.com)