

What is a research Infrastructure?

Commonalities and differences across science and policy, from mega-science to shared facilities

Benedetto Lepori* and Marco Cavallaro**

*benedetto.lepori@usi.ch

<https://orcid.org/0000-0002-4178-4687>

Institute for Communication and Public Policy, Faculty of Communication, Culture and Society Department,
Università della Svizzera italiana, Switzerland

** marco.cavallaro@usi.ch

<https://orcid.org/0000-0003-4810-5525>

Institute for Communication and Public Policy, Faculty of Communication, Culture and Society Department,
Università della Svizzera italiana, Switzerland

1. Introduction

Since the beginning of the 21st century, the concept of *Research Infrastructure* (RI) has gained popularity in the research policy literature (Hallonsten and Cramer 2020), and in the political discussion about research funding (Franssen 2020). The emergence of the concept has been promoted by the European Union (EU) through the establishment of the European Strategic Forum for Research Infrastructures (ESFRI), while most developed countries have integrated RIs and their funding within their set of research policy instruments (Bolliger and Hallonsten 2020).

Despite its prominence, the notion of RI remains relatively ambiguous and subject to debate. Frequently, definitions are of operational nature and are driven by the interests of actors to receive funding for their own infrastructure. Some scholars even argue that the RI concept represents a purely political construct to fund initiatives that have little in common besides being awarded the RI label (Hallonsten 2020).

In our work, we investigate the different meanings of the concept of ‘research infrastructure’ in the scholarly literature and in the political process at the EU and national levels. First, we propose to challenge the presumption that the concept is purely procedural and look into meanings attributed by science and policy actors and what they have in common. Second, we explore differences in meanings across disciplinary communities, and between science and policy. Third, we identify some ideas that seem to constitute the core of the notion of research infrastructure that is widely shared by different scientific communities and by policy, such as the idea of open access, the one of sharing tools for research and the existence of a community of users.

While struggles over resources are an unavoidable and even necessary dimension of research funding policies, the lack of clarity about the concept and its different extensions makes an informed discussion and prioritization more difficult and might affect the legitimacy and accountability of funding decisions. Our work will contribute to national and international debates on RIs, by inquiring into commonalities between their different conceptions across disciplines and sectors.

2. Literature review

There is a small number of works dealing specifically with the origin of the concept of RIs, showing how it was historically linked to the concepts of “big science” or “mega-science” (Cramer & Hallonsten 2020), where RIs were mostly considered to be large and single-sited facilities such as particle accelerators and telescopes, and to the emergence of European research policy and its goal of coordinating national initiatives (Ulnicane 2020).

This literature emphasizes the political and processual nature of the RI definition. As put forward by Hallonsten (2020), the criteria proposed in the ESFRI definition of RIs as ‘facilities, resources or services of a unique nature that have been identified by European research communities to conduct top-level activities in all fields (ESFRI Forum 2018) are not easily applicable to the European RI landscape. No RI included in the ESFRI roadmap is unique worldwide, and most of them have ‘competitors’ even at the national level, and the criterion of top-level activities hardly applies to most of them. In Hallonsten’s perspective, the core of the definition is in fact the identification of a process through which RIs are, first, identified by research communities and, second, prioritized in a political process involving countries, which are in principle willing to commit resources to their establishment and maintenance. The RI concept as used in the political process of roadmapping represents therefore a way of labelling and prioritizing some initiatives to channel them European and national funding. Accordingly, the (political) RI definition is closely associated with the establishment of roadmaps and related funding instruments (Bolliger and Hallonsten 2020).

According to the literature, the labelling of RIs is, therefore, by and large, a tool for the governance and funding of research by public authorities (Franssen 2020). Its emergence can be seen as a response by public authorities to two emerging issues. On the one hand, structural changes in public research funding implied a reduction of baseline funding to universities and public research organizations, through which many RIs were funded in the past (Lepori, Jongbloed and Hicks 2023). This generated the need for a specific funding channel for RIs beyond the few very large infrastructures, which always required specific arrangements because of their size. On the other hand, current research policy increasingly emphasizes the need for coordination, achieving a critical mass and avoiding duplications (Elzinga 2012); accordingly, it has become less acceptable to finance, in parallel, similar initiatives in a decentralized way. As shown by the case of digital humanities in the Netherlands, funding for RIs can be used as an incentive for scholarly communities, particularly in traditionally fragmented fields such as social sciences and humanities, to develop stronger forms of cooperation (Franssen 2020). Conversely, success in putting RIs on national and European roadmaps largely depends on the ability of the related communities to organize their activities jointly.

The heterogeneity of RIs has led to efforts to develop typologies or classifications of RIs to create some order and to identify common patterns. A typology in terms of the RIs’ functions has been proposed by Hallonsten (2020). He distinguished between systems to perform measurements (*instruments*), facilities to observe the real world (*observatories*), collections of data to be used in research (*repositories*) and, finally, support that allows research at remote sites, such as aircraft (*vessels*).

Applying these classifications to the 60 RIs included in the ESFRI roadmaps, he was able to identify some patterns. Expectedly, the single-sited RIs are mostly instruments and observatories and concentrated in sciences (astronomy, physics, material sciences, engineering). On the contrary, the multiple-site category is very heterogeneous in terms of functions and organization; this applies particularly to the multiple-sited and multiple-purpose RIs, where it is hardly possible to find any commonalities – some of them being simple collections of national facilities. In Hallonsten’s view, this shows how flexible the RI concept is, but also questions whether overstretching it to this extent makes the concept useless. From

a slightly different perspective, this analysis suggests that the concept of RIs might still be rather clear when dealing with specific and localized infrastructures, where concentrating facilities in one place allows the construction of more powerful telescopes, accelerators or test facilities; as soon as the single-sited constraint is lifted, it becomes more and more difficult to distinguish in practice between RIs and networks of laboratories or researchers sharing some facilities.

Clearly, the previous analysis shows that the mega-science concept does not any more adequately describe RIs. Has the concept been so overstretched to become virtually meaningless or can we identify emerging constitutive dimensions of a new concept?

In our work, we propose to delve further into the meaning of the RI concept as understood by scientific and policy communities and how it is associated with other patterns in science policy. Thereby, we aim to identify commonalities that could contribute to a core definition of RIs.

3. Methodology

To this aim, we used a mixed-method cross-sectional research design, using both quantitative and qualitative approaches (Schoonenboom & Johnson, 2017).

Firstly, to trace the use of the RI concept throughout the years and across scientific disciplines, we conducted a search query in November 2022 on the Scopus database (www.scopus.com) for the publications including the sentence ‘research infrastructure’ in the title or the keywords. To identify specific sets of meanings associated with RIs, we analyzed the co-occurrence of words used in the title and abstract of the publications and defined clusters through the association of neighboring words. The analysis was performed using the VoS viewer software (Van Eck and Waltman 2010).

Secondly, to analyze patterns in the use of the RI concept in policymaking, we considered a body of policy documents from ESFRI, the InRoad project (<https://www.inroad.eu/>) - which notably collected a set of national definitions of RIs - supranational entities (EU, OECD) and national RI roadmaps in Europe. Based on these documents, we compare definitions across countries and supranational entities, and conduct a linguistic analysis, by looking into the prototypicality of words used in the definitions of RIs. For the latter, we will use the ProtAnt software which can be used to analyse neighboring words to a core concept and identify its meaning (Anthony & Baker 2015). For this analysis, European countries were selected based on the size of their research ecosystem and on available information.

4. Preliminary results

4.1. RI definition across scientific disciplines

Our search query retrieved 1,578 documents, most of them being journal papers (655) or conference papers (650). As shown in Figure 1, the term was very rarely used before 2005, while the number of documents exceeded 150 in 2020 (data for the years 2021 and 2022 are still incomplete). The introduction of the term in European research policy, with the foundation of ESFRI in 2002, therefore pre-dates its scholarly usage, which started with Papon’s paper on European research cooperation (Papon 2004). Most disciplinary papers deal with specific instances of RIs and take for granted the label of ‘research infrastructure’ without attempting to elaborate on its definition.

Several works analyzed the establishment of entities, which are today included among research infrastructures, such as CERN or EURATOM. However, they were usually subsumed under concepts such as ‘big science’ or ‘mega-science’, i.e., with the idea that some forms of scientific inquiry required a large scale of investments (in terms of funding, personnel, and political

process), which required coordinated action at the country and/or international level. This original idea of ‘big science’ (and related criteria of scale and uniqueness) is still present in many political definitions of Ris, but hardly fits the current usage of the term.

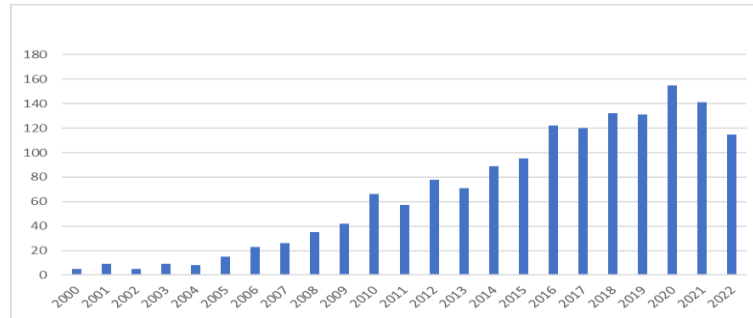


Figure 1: Publications in Scopus using the word ‘research infrastructure’

Retrieved publications are distributed over all research areas, with the largest numbers in computer science (19%), engineering (13%), social sciences (10%), medicine (8%), physics and astronomy (7%). While the term ‘big science’ was mostly used for facilities in natural sciences and engineering, the term ‘research infrastructure’ has therefore become widespread in all subject domains. Given the differences between scientific domains in how research is conducted and in the type of facilities required, this is expected to translate into high heterogeneity of the entities labelled as RIs.

A fine-grained view of the concepts associated with the RI term is provided by the analysis of the words used in the title and abstract of the publications, grouped by co-occurrences. In this analysis, neighboring words in the map occur together frequently in the publications, and, accordingly, clusters of words identify specific sets of meanings associated with RIs.

Four main clusters can be identified (Figure 2); their main feature is to be mostly associated with specific disciplinary contexts, suggesting indeed that RI definitions are largely discipline-specific. More specifically, we distinguish between:

- A *research data* cluster (blue), where the focus is on repositories, data architecture, ontologies and open data. Expectedly, this cluster is also associated with social sciences and humanities.
- An *IT* cluster (yellow), including advanced computing facilities, but also software platforms and testbeds.
- A cluster dealing with facilities and research instruments in *physics, engineering and environmental sciences*, such as telescopes, and accelerators (green); expectedly, this cluster includes also the sentence ‘large research infrastructure’, as well as the terms associated with European policies such as ESFRI.
- A *health-related* cluster (red), which can be broadly divided into two dimensions: on the one hand, clinical medicine such as clinical trials and patients’ data, and on the other hand (basic and translational) medical research, such as biobanks.

This definition considers several aspects linked to RIs, such as their purpose (“achieve excellence in RI”, “use beyond research”), their form (single-sited, virtual, or distributed), and some wide categories and examples. Although this definition serves as a reference for EU Member States, the set of national definitions collected within the InRoad project shows variations across national contexts.

For example, the Austrian, French, Dutch, and Spanish definitions, among others, state that, to be considered as such, RIs must offer “unique capabilities”. Some countries defined specific threshold values, e.g., at least € 50 million in construction costs and ten years of service life in Germany, the range of € 3-14 million for the construction and/or implementation of RIs in Denmark, or the minimum of five years of service life in the Netherlands. In Germany, the use of RIs is regulated based on scientific quality standards, while in Sweden, RIs must be easily accessible to researchers, industry, and other stakeholders (InRoad 2018).

Differences can also be found in the RIs’ target groups. While most countries largely consider researchers as their main target groups, some countries emphasized the industry relevance of the RIs. This includes the UK, which in its last roadmap to date, sees RIs as enablers for the development, demonstration, and delivery of new “innovative processes, products, and services” (UKRI 2020). The evaluation of Spanish potential RIs takes into account the potential industrial return and cooperation with other types of stakeholders.

Table 1 shows an overview of aspects covered in the EU and national definitions of RIs.

Table 1: Aspects covered in national definitions of RIs

	AT	CH	DE	DK	EU	FI	FR	IL	NL	SE	UK
Costs and size	•	•	•		•			•	•		•
Governance		•	•				•		•		
Industry relevance				•						•	•
International	•	•		•		•	•				•
National interest		•	•	•		•	•			•	•
Open Access policy		•	•	•		•	•	•	•	•	•
Scientific excellence					•	•	•	•	•	•	•
Service life, sustainability		•	•	•		•	•		•	•	•
Societal relevance	•			•							
Uniqueness	•				•		•		•		

5. Preliminary conclusions

Through this preliminary analysis, we identified two usages of the term ‘research infrastructure’.

Within the scientific sector, the term RI is generically used for entities or facilities, or tools shared by a research community to organize joint research activities. The nature of these entities varies between scientific fields, but they have in common two basic ideas: the existence of research communities sharing them and the fact that what is shared is not just research ideas or people, but some kind of material (or electronic) artefact. In practice, it might be sometimes difficult to distinguish between research cooperation and RIs.

In policy, the term RI is used in the research policy process (at the national and international level) to designate entities that are awarded a certain label and, by this, are facilitated in the

search for research funding from different sources. In such a perspective, RIs are a tool in the governance and funding of research, which allows for prioritizing scientific programs and structuring research communities around a specific stream of resources. By its nature, the subset of RIs in this second meaning is much smaller than in the first; and some labelled RIs do not fully correspond to the first definition as they are more collections of independent activities than shared facilities.

The linguistic analysis of the policy documents will further test our preliminary findings and possibly reinforce the identification of core elements for defining RIs.

Open science practices

In this paper, we used data from Scopus for the quantitative analysis along with a body of publicly available documents a body of policy documents from ESFRI, the InRoad project (<https://www.inroad.eu/>), supranational entities (European Commission, OECD) and national RI roadmaps in Europe.

Author contributions

BL conceived the original idea, conducted the literature review and developed the theoretical framework with support from MC. BL prepared datasets and carried out the analysis of the use of the RI concept across disciplines. MC compiled the official sources and carried out the analysis of RI definitions in research policy with BL's advice. BL and MC contributed equally to the interpretation of results.

Competing interests

Not applicable.

Funding information

Not applicable.

References

- Anthony, L. & Baker, P. (2015). ProtAnt: A tool for analysing the prototypicality of texts. *International Journal of Corpus Linguistics*, Volume 20, Issue 3, 273 – 292.
- Bolliger, I. & Hallonsten, O. (2020). The introduction of ESFRI and the rise of national research infrastructure roadmaps in Europe. In K. Cramer & O. Hallonsten (Eds.) *Big Science and Research Infrastructures in Europe* Edward Elgar Publishing.
- Bonaccorsi, A. (2008). Search Regimes and the Industrial Dynamics of Science. *Minerva*, 46(3), 285-315.
- Cramer, K. & Hallonsten, O. (Eds.) (2020). *Big Science and Research Infrastructures in Europe*, Edward Elgar Publishing.
- Elzinga, A. (2012). Features of the current science policy regime: viewed in historical perspective. *Science and Public Policy*, 39(4), 416-428.
- ESFRI (2017). *Sustainable European Research Infrastructures, a call for action*. EC Publications Office, Luxembourg.
- Franssen, T. (2020). Research infrastructure funding as a tool for science governance in the humanities: A country case study of the Netherlands. In K. Cramer & O. Hallonsten (Eds.) *Big Science and Research Infrastructures in Europe*, Edward Elgar Publishing.
- GSO (2019). *Framework for Global Research Infrastructures*. GSO.
- Hallonsten, O. (2020). Research infrastructures in Europe: The hype and the field. *European Review*, 28(4), 617-635.

- Hallonsten, O. & Cramer, K. (2020). Big science and research infrastructures in Europe: Conclusions and outlook. In K. Cramer & O. Hallonsten (Eds.), *Big Science and Research Infrastructures in Europe*, Edward Elgar Publishing.
- InRoad (2018). Good practices and common trends of national research infrastructure roadmapping procedures and evaluation mechanisms.
- Lepori B., Jongbloed B., & Hicks D. (2023). *Handbook of Public Research Funding*. Edward Elgar Publishing Ltd., Cheltenham.
- OECD (2010). *Establishing Large International Research Infrastructures: Issues and Options*. OECD, Paris.
- Papon, P. (2004). European scientific cooperation and research infrastructures: Past tendencies and future prospects. *Minerva*, 42(1), 61-76.
- Schoonenboom, J., & Johnson, R.B. (2017). How to Construct a Mixed Methods Research Design. *Köln Z Soziol*, 69(2), 107–131.
- Ulnicane, I. (2020). Ever-changing Big Science and Research Infrastructures: Evolving European Union policy. In K. Cramer & O. Hallonsten (Eds.), *Big Science and Research Infrastructures in Europe*, Edward Elgar Publishing.
- Van Eck, N. & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.