

# Commentary on ‘Data Repository Selection: Criteria that Matter’

## Science Europe’s response to a publishers’ approach

### 1. Introduction

Science Europe has been supporting developments towards Open Science in general, and towards sharing of FAIR<sup>1</sup> data in particular, for many years. It has carried out substantial work with its Member Organisations to align their approaches and to support researchers in their research data management (RDM). In its [Practical Guide to the International Alignment of Research Data Management](#), published for the first time in 2019 and updated in 2021, and developed in close collaboration with the broader research stakeholder community, Science Europe presents minimum requirements that a repository should meet to be considered as trustworthy repository. Regardless of the fields they cover and their size, data repositories have to guarantee the quality of the data preserved.

Science Europe welcomes the opportunity to comment on the [draft criteria for trustworthy repositories](#) developed by a group of publishers who have come together in the FAIRsharing community and would like to raise the following points:

### 2. Response to the Draft Paper

#### 2.1. General Comments

Science Europe welcomes initiatives that increase consistency and improve support for researchers needing to comply with RDM requirements. At European level, the European Open Science Cloud (EOSC) is a central undertaking. It aims to federate data services and infrastructures to make research data interoperable and allow researchers to access, (re-)use, and share data. Stakeholders involved in the development of EOSC have already made significant progress in defining services, tools, and procedures that will set new standards at international level. The success of EOSC will, to a large extent, depend on best possible RDM standards. Selecting a repository to share data is an important part of the RDM tasks required of researchers.

There are several thousand repositories in Europe, ranging from small to large; some are general, some are discipline-specific, and others are institutional. There is currently no unique accepted list of trustworthy repositories, and their levels of maturity, trustworthiness and sustainability of

---

<sup>1</sup> Data that is Findable, Accessible, Interoperable and Re-usable.

service are difficult to assess. Discipline-specific repositories have certain policies and standards in place that meet the needs of the specific research community that they serve. Other repositories that serve the general research community have policies and standards that are necessarily more generic.

Due to the differing nature of these repositories, they are difficult to compare. Too strict and too detailed criteria risk excluding repositories that can offer valuable services to a dedicated scientific user group. Some repositories have been certified as 'trustworthy' by one or several acknowledged certification bodies; however, small, institutional, or discipline-specific repositories might not (yet) have the means to seek such certification.

Science Europe [recommends](#) that researchers should refer to certified repositories or discipline-specific repositories that are broadly recognised as trustworthy by their respective community where and when possible. But there are cases in which no such repository can be identified. Researchers should then be supported in their choice by a minimum selection of core criteria. Any supporting tool should not be prescriptive, overly complicated or exclude important repositories of research communities that may be in active usage for already quite some time, but do not meet formal certification criteria

Science Europe acknowledges that the criteria developed by a group of publishers within the FAIRsharing community are intended to support researchers who wish to publish the data underlying their research findings and publishers in providing adequate guidance. It is understandable that publishers require access to data, for example for the purpose of providing a high quality peer review. However, based on its experience and broad consultations when developing its own criteria, Science Europe would like to point out a number of areas where it has considerable concerns with the suggested criteria as they currently stand.

## 2.2. Comments on the Draft Criteria

In general, Science Europe shares the same concerns raised by [COAR in its response to the draft criteria](#), and would like to stress the following points in particular:

- **The criteria should focus on the researchers' needs.** The decision on where to deposit data should be taken by the researcher based on community standards, and if applicable supported by the researcher's home institution and/or funding agency. Researchers should have the freedom to choose the repository that is considered best suited for the respective datasets. It is not the role of publishers or any commercial entity to influence the researcher's choice and define the primary location for datasets for the purpose of sharing, description, curation, and preservation. While the criteria presented by the group of publishers are not inherently incorrect, they are excessively focused on the desires of publishers to link and peer review the data, and do not take into account other important considerations a researcher may have when choosing a repository. For example, researchers may need to deposit data according to applicable legal context. They may also want to deposit data in a discipline-specific repository regularly used by their peers or their institution offering better interoperability as they use discipline-specific standards and metadata.
- **The criteria should support the whole research system.** The draft criteria are a mix of different requirements that do not entirely reflect current standards and practices in the

research system. By setting too high standards too early, they risk to deepen the inequality and limiting the diversity of repositories in the research ecosystem by inadvertently restricting compliance to a handful of well-resourced repositories. Criteria for trustworthy repositories should support research and the development of a suitable repository system as a whole, taking into consideration sustainability, integration and interoperability issues. In a system where more and more alignment is sought among the different actors, especially in light of the development of EOSC, criteria that are not supportive and aligned with sector standards will cause confusion and disruption.

- **The suggested tool excludes a significant number of repositories.** The draft criteria proposed explicitly exclude institutional repositories, which in some cases can be the most suitable choice for the deposition of a dataset and sometimes a long time in use choice for researchers. Access to such repositories will become possible even for researchers from outside the institution in the future, with an increasing number of repositories connecting to EOSC. Furthermore, many repositories currently do not comply with the extremely large number of criteria and do not have the resources to adapt to them without some lead time. A step-by-step approach should be taken, with a focus on repository-level criteria in the first place. The idea of creating a database of available repositories with a filtering tool can be very helpful for researchers to select a trustworthy repository. However, the code of such a tool and the underlying data including documentation (to understand why publishers may attribute more importance to some criteria over others) need to adhere to the FAIR principles and should be openly shared with the research community to ensure transparency. The proposed criteria can lead to many repositories falling through the net when using the filtering tool. Again, such an approach would favour a small number of well-resourced repositories to the detriment of diversity in the research system.

Science Europe would like to offer dialogue with the authors of the draft criteria and FAIRsharing representatives, who would curate the data depository descriptions in the future, to discuss and jointly find solutions for our common goal to support researchers in their RDM and in the identification of trustworthy repositories for high-quality data in an independent manner.

## **About Science Europe**

Science Europe represents major public organisations that fund or perform excellent, ground-breaking research in Europe. It brings together the expertise of some of the largest and most respected European research organisations to jointly push the frontiers of how scientific research is produced and delivers benefits to society. Science Europe's 36 members manage a large variety of national and international funding programmes, from bottom-up schemes to mission-oriented research. They collectively invest €18.5 billion in 27 countries each year.