EOSC-hub

A Service Provider's Guide to Data Sharing Policies

These reference cards convey data sharing policy recommendations to be adopted by data and service providers within the EOSC-hub consortium.

Our recommendations contribute to the developing field of data sharing policies in the EOSC at large.

The Why | The What | The How

Why should you care about implementing data sharing policies? Three reasons.

For more details, consult the D2.8 and D2.9 deliverables via: https://bit.ly/2zDAAnM

DATA INTEGRITY & AUTHENTICITY

Information about provenance of scientific data is crucial to assess data integrity and authenticity.

EOSC-hub should consider the logging and tracking of scientific provenance data as an element of service integration design.

Good practice example: extending standard provenance modelling frameworks to include "workflow" structures[1].

Another EOSC-hub example is the PID-based provenance support through the integration with specific services like B2HANDLE as adopted by ENES[2].

CROSS-DOMAIN COLLABORATION

A wide variety of stakeholders broadens the engagement and facilitates cross-domain collaboration.

EOSC-hub should engage with a broader set of stakeholders, including social science and statistical data service providers, in supporting the design of a Europe-wide framework for research with sensitive data.

Examples of such engagements are EOSC-hub partners contributing to new projects like SoBigData++[3].

TRUST AND CONFIDENCE

Adopting formal data sharing policies is an excellent step towards service accreditation. For providers of data and data services, external accreditation is becoming highly desirable, and sometimes essential, in building the necessary trust with important user communities, partner service providers, or both.

Whether the right path for you is CoreTrustSeal, ISO27001, FitSM, or even ISO16363, formalising your policies on data sharing is a key first step, and of itself a great way to build trust with your existing user base. Formal policies will also help you create networks with your partners to support emerging research data codes of conduct across shared user communities.

EOSC-hub and for example the CORBEL-project could mutually benefit from service accreditation[4].

NOTES

1] P.Missier et al, D-PROV: extending the PROV provenance model with workflow structure. In: TaPP; 2013, https://bit.ly/3c5Dvml

2] https://bit.ly/37ZWTkc

3] https://bit.ly/2VhwJEx

4] In the framework of the CORBEL project: EOSC-hub partner ECRIN has developed principles and recommendations https://bit.ly/2Z8PdZ2

EOSC-hub

A Service Provider's Guide to Data Sharing Policies

These reference cards convey data sharing policy recommendations to be adopted by data and service providers within the EOSC-hub consortium.

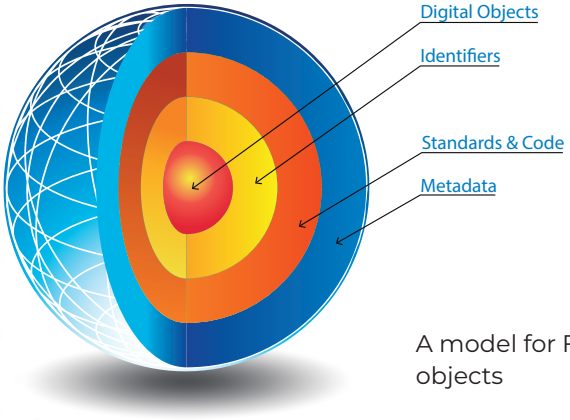
Our recommendations contribute to the developing field of data sharing policies in the EOSC at large.

The Why | **The What** | The How

What are essential concepts in the context of data sharing policies in the EOSC-hub ecosystem?

For more details, consult the D2.8 and D2.9 deliverables via: https://bit.ly/2zDAAnM

A FAIR ECOSYSTEM SUPPORTING FAIR DIGITAL OBJECTS

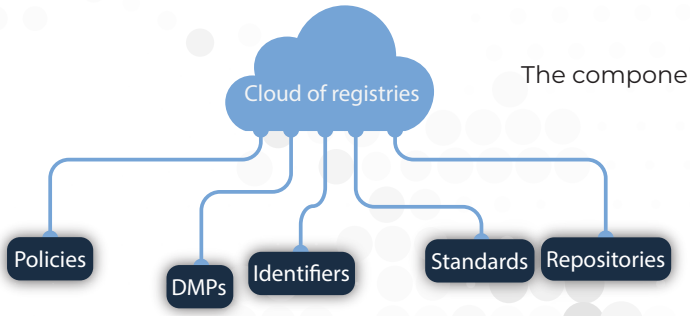


A model for FAIR digital objects

All shareable data objects in the EOSC-hub ecosystem should be FAIR Digital Objects.

FAIR Data Objects are Findable, Accessible, Interoperable and Reusable. A FAIR ecosystem ensures a number of data services and components to be in place that enable FAIR data[5].

Offered through EOSC-hub is a variety of services that implement these concepts such as EGI DataHub[6] and the EUDAT B2 Service Suite[7].



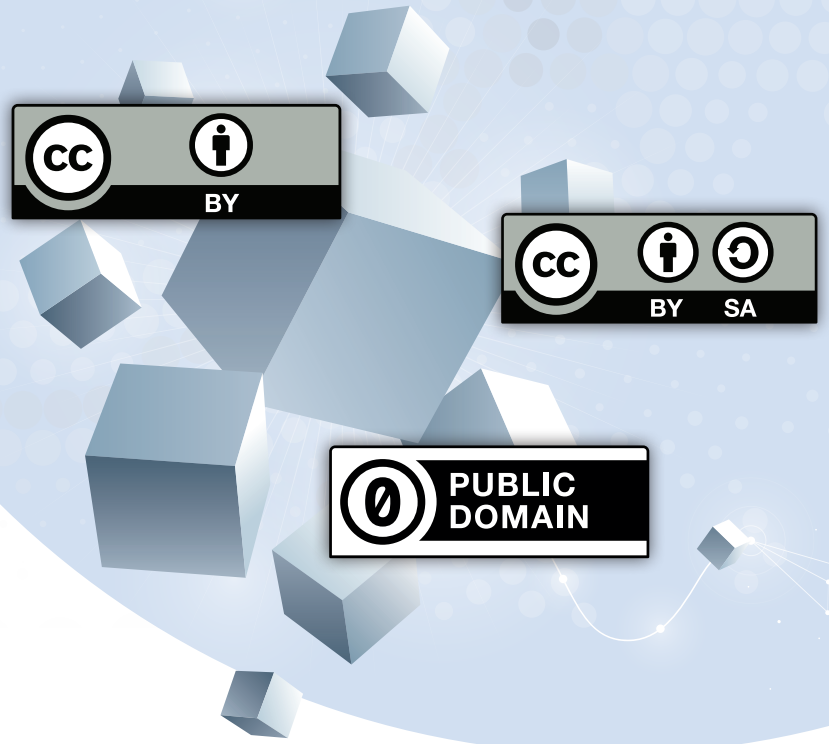
The components of a FAIR ecosystem

LICENSING

Data objects in the EOSC-hub ecosystem should adopt licences from the Creative Commons 4.0 licence suite.

Where data are openly shareable, these should be one of: CC BY 4.0 (attribution), CC BY SA 4.0 (attribution with onward propagation), CC0 (public domain or rights waiver)[8].

The EOSC-hub B2SHARE service, for example, hosts a tool to help the user choose the correct licence for their data[9].



NOTES

5] S.Hodson, S.Jones et al, Turning FAIR into reality, European Commission Expert Group on FAIR Data, November 2018, https://bit.ly/2YvDJPX

6] https://bit.ly/2CHkURB

7] https://bit.ly/2CzM24Q

8] https://bit.ly/2VbCund

9] https://bit.ly/2YxP1Dv

This card is provided by Parties of the EOSC-hub Consortium (grant agreement 7775360)

EOSC-hub

A Service Provider's Guide to Data Sharing Policies

These reference cards convey data sharing policy recommendations to be adopted by data and service providers within the EOSC-hub consortium. Our recommendations contribute to the developing field of data sharing policies in the EOSC at large.

The Why | The What | **The How**

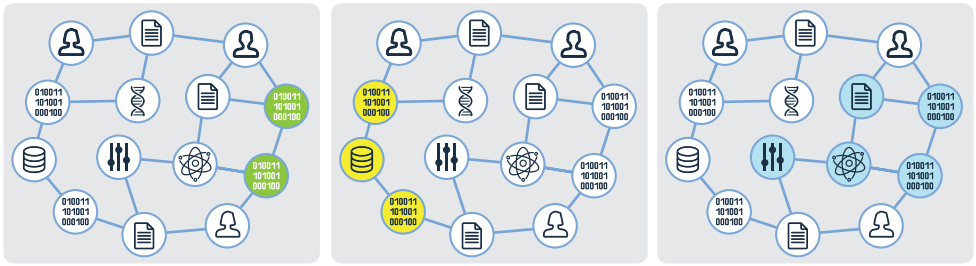
How can you implement data sharing policies practically? Some examples.

For more details, consult the D2.8 and D2.9 deliverables via: https://bit.ly/2zDAAnM

PERSISTENT IDENTIFIERS

Persistent Identifiers (PIDs) are long-lasting references to a data object. By using PIDs, you ensure findability and accessibility. EOSC-hub should initiate a programme of technical research for direct retrieval of data objects by PID.

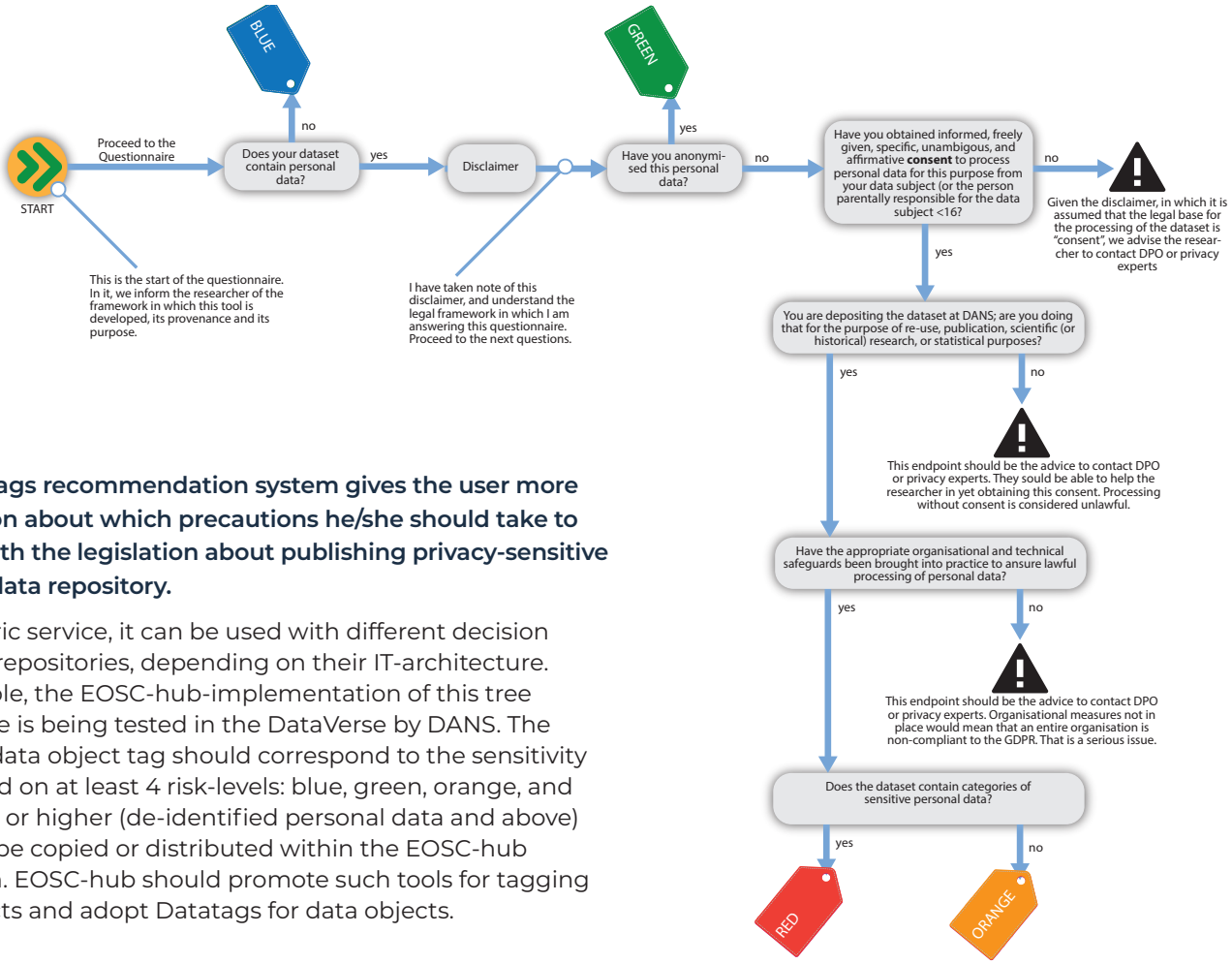
Additionally, EOSC-hub should track the work of the FREYA project and adopt best practices in PID resolution as they emerge. One of the outputs of the FREYA project, is the PID-graph.



Source: PID Chart, Fenner & Aryani, FREYA

Figure: FREYA-project PID-graph of three use cases with digital objects connected by PIDS: different versions of software code (left), datasets hosted by a particular repository (middle) and all digital objects connected to a research object (right).

DATATAGS RECOMMENDATION SYSTEM



The Datatags recommendation system gives the user more information about which precautions he/she should take to comply with the legislation about publishing privacy-sensitive data to a data repository.

As a generic service, it can be used with different decision trees and repositories, depending on their IT-architecture. For example, the EOSC-hub-implementation of this tree as a service is being tested in the DataVerse by DANS. The resulting data object tag should correspond to the sensitivity level, based on at least 4 risk-levels: blue, green, orange, and red. Green or higher (de-identified personal data and above) shouldn't be copied or distributed within the EOSC-hub ecosystem. EOSC-hub should promote such tools for tagging data objects and adopt Datatags for data objects.

EOSC-hub

A Service Provider's Guide to Data Sharing Policies

These reference cards convey data sharing policy recommendations to be adopted by data and service providers within the EOSC-hub consortium. Our recommendations contribute to the developing field of data sharing policies in the EOSC at large.

The Why | The What | **The How**

How can you implement data sharing policies practically? Some examples.

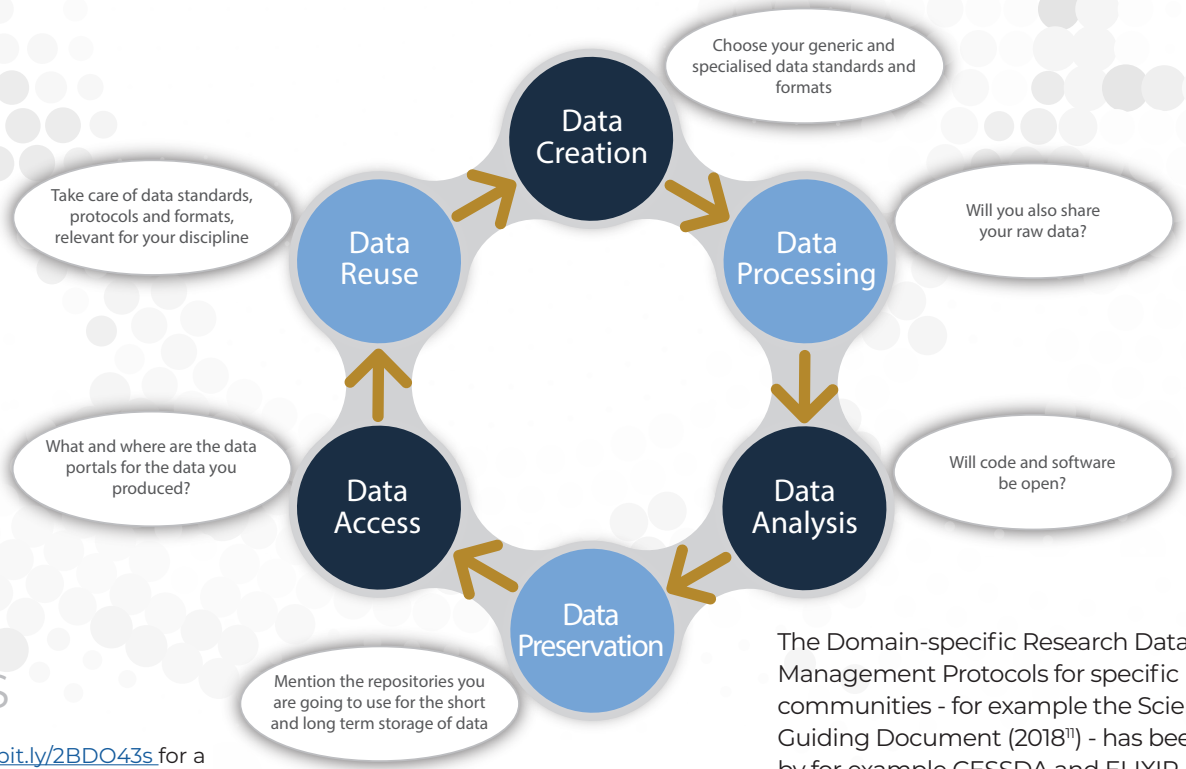
For more details, consult the D2.8 and D2.9 deliverables via: https://bit.ly/2zDAAnM

ADOPT GOOD DOMAIN PRACTICES

The Five Safes Principles

Within Medical and Health data services, a good domains practice is that of the Five Safes. EOSC-hub should adopt these Five Safes principles as guidance[10] for the management and handling of sensitive data in the EOSC-hub ecosystem.

Domain-specific RDM Protocols



The Domain-specific Research Data Management Protocols for specific communities - for example the Science Europe Guiding Document (2018[11]) - has been adopted by for example CESSDA and ELIXIR. EOSC-hub stimulates the application of data domain protocols.

NOTES

10] See https://bit.ly/2BDO43s for a good discussion of the principles and current applications

11] via https://bit.ly/3d4qhaR