D2.9 - Final Data Policy recommendations

Lead Partner:	DANS (T2.4 lead)
Version:	1 (NOT YET APPROVED BY THE EUROPEAN COMMISSION)
Status:	
Dissemination Level:	Public
Document Link:	https://docs.google.com/document/d/1K8OWeF9smFIg502NZcBHP1Lrv41e3QfUo5imuKuM2q8/edit

Deliverable Abstract
This deliverable conveys data sharing policy recommendations to the EOSC-hub data and service providers. It strongly builds on EOSC-hub D2.8 (First Data policy recommendations) by translating these first policy recommendations into four reference cards that together form the Service Provider's Guide to Data Sharing Policies. D2.9 sets out to contribute to the developing field of data sharing policies in the EOSC.

NOT YET APPROVED BY THE EUROPEAN COMMISSION

DELIVERY SLIP

	Name	*Partner/Activity*	*Date*
From:	Frans Huigen, Ilona von Stein	DANS/WP2	30 June 2020
Moderated by:			
Reviewed by:			
Approved by:			

T2.4 WORKING GROUP

I. von Stein, M. Wittenberg, E. Dijk & F. Huigen (DANS); R. Baxter (EPCC); J. Nordling (CSC); C. Ohmann & M. Matei, M. Panagiotopoulou (ECRIN); A. Manzi (EGI.eu); T Wildish (EMBL-EBI); R. Carrillo & G Savini (TRUST-IT).

DOCUMENT LOG

Issue	*Date*	*Comment*	*Author*
V0.1	29/04 - 29/05 2020	First draft delivered for internal review on May 29th 2020. Various feedback iterations by all working group members on textual content, reasoning, design choices and deliverable quality. At the same time, various design iterations with TRUST-IT.	F Huigen, I von Stein, M Wittenberg; R Baxter ; J Nordling; C Ohmann & M Matei; A Manzi; T Wildish; R Carrillo & G Savini;
	29/05 - 18/06 2020	First draft reviewed by internal reviewers of the EOSC-hub, and by the members of the T2.4 working group. Feedback concerned content of the deliverable, particularly the phrasing and framing of EOSC-hub recommendations, and on the design of the reference cards: density of information was a concern.	E Dijk; M Panagiotopoulou & C Ohmann; A Manzi; R Carrillo & G Savini
V1.0	18/06 - 30/06 2020	Comments, suggestions, corrections incorporated. In parallel: design iterations with TRUST-IT. Integrated deliverable text with reference cards. Final version delivered June 30th, 2020.	F Huigen & I von Stein All authors

NOT YET APPROVED BY THE EUROPEAN COMMISSION

TERMINOLOGY

https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary

Terminology/Acronym	***Definition***
CORBEL	Coordinated Research Infrastructures Building Enduring Life-science Services, an initiative of thirteen new biological and medical research infrastructures (BMS RIs). https://www.corbel-project.eu/home.html.
CoreTrustSeal	A core certification of digital repositories, based on the DSA–WDS Core Trustworthy Data Repositories Requirements and procedures. https://doi.org/10.5281/zenodo.3632533
DataTags	A system of human-readable and machine-actionable labels that express conditions under which datasets can be stored, transmitted, or used; https://techscience.org/a/2015101601/
DOI	Digital Object Identifier, a well-recognised form of PID (qv); http://www.doi.org/
FAIR principles	Principles of best practice in open research data management, an acronym of Findability, Accessibility, Interoperability and Reusability; https://www.force11.org/group/fairgroup/fairprinciples
FitSM	FitSM is the name for a family of standards for lightweight IT service management (ITSM). https://www.egi.eu/services/fitsm-training/
GDPR	The General Data Protection Regulation came into force as of 28 May, 2018. It is a regulation valid for the whole of the European Union, concerning privacy and the protection of privacy and personal data.
ISO27001	ISO 27001 (formally known as ISO/IEC 27001:2005) is a specification for an information security management system (ISMS). https://www.iso.org/isoiec-27001-information-security.html
ISO16363	A system for the "audit and certification of trustworthy digital repositories", by the Consultative Committee for Space Data Systems (CCSDS).
PID	Persistent identifier, for example a DOI or accession number.
RDA (WG)	Research Data Alliance (Working Group)
Sensitive data	Data which, for whatever reason, cannot be openly shared without the risk of disclosure of legally or ethically sensitive information.
Service	Way to provide *value* to customers through bringing about results that they want to achieve. In the context of the FitSM standard series, when referring to services, usually IT services are meant.
Service Provider	Organisation or federation (or part of an organisation or federation) that manages and delivers a service or services to customers

Contents

Scope and Context

This document represents D2.9 Final Data Policy recommendations: a set of policies to facilitate the sharing and safe processing of data from across European research infrastructures.

The main objective of this deliverable is to convey policy recommendations to the EOSC-hub data and service providers in a visually attractive and pragmatic way. This work strongly builds on the First Data Policy recommendations D2.8 report as published in December 2018 [1]. In deliverable D2.9 we translated the First Policy Data recommendations into four reference cards that together form the Service Provider's Guide to Data Sharing Policies. An additional benefit of these cards is the wider raising of awareness and the dissemination of existing good practice in relation to data sharing policies from within the EOSC-hub consortium.

The development and implementation of data sharing policies in the EOSC is a developing field. EOSC builds in iterations, and building the EOSC requires Europe's research institutions and infrastructures to align and co-develop. These policy recommendations of D2.9 are part of this developing field. It might well be the case that adaption and revision is necessary in response to EOSC governance mechanisms or other developments.

Structure of the report

The first section considers the relation between the First Data policy recommendations (D2.8) and the Final Data policy recommendations (D2.9). This is followed by a section that sets forward why and how practical recommendations from D2.8 were translated into The Why, The What, and The How reference cards of D2.9. Afterwards, the Service Provider's Reference Guide to Data Sharing Policies, that consists of four reference cards, is displayed. We conclude this deliverable with proposed next steps.

From First (D2.8) to Final (D2.9) Data policy recommendations

D2.8 (First Data policy recommendations) recommends 22 practical steps bridging general policy recommendations and future technical implementation of data sharing within the EOSC-hub service ecosystem.

The recommendations fall under three broad headings:

1. Implement FAIR
2. Build technical expertise in 'safe data' and 'safe settings'
3. Support the wider development of ethical and information governance frameworks

The deliverable at hand, the Final Data policy recommendations (EOSC-hub D2.9), specifically respects and reuses the existing outcomes of the First Data policy recommendations (EOSC-hub D2.8). The First Data policy recommendations in itself did actively reuse other community best practices as well, e.g. the overall policy framework for EOSC as developed by the EOSCpilot project [2] and the EC Expert Group report Turning FAIR into reality [3]. With such a vast knowledge base, for the Final Data policy recommendations, we made the same choice to acknowledge prior work and reuse the existing recommendations of D2.8 as a starting point for D2.9. This starting point is supported by the perceived high quality of D2.8, since it was very well-received both by internal reviewers as well as by the European Commission.

Obviously, since the release of EOSC-hub's First Data policy recommendations in 2018, progress has been made in the field. Among other things, the set up of EOSC Governance has moved forward, as well as a wealth of practical projects and initiatives that develop and implement activities related to data sharing policies. Examples (only) of such efforts are the work of the RDA FAIR data Maturity Model WG on developing a common set of core assessment criteria on FAIRness [4], developing best practices in PID resolutions as part of the FREYA project [5] and the work of the FAIRsFAIR project on supplying practical solutions for the use of the FAIR data principles throughout the research data life cycle [6].

Nevertheless, we consider the essence of the recommendations in D2.8 to be still valid. For D2.9, instead of reinventing the wheel, we translated the first Policy Data Recommendations into hands-on guidance that we recommend to be adopted by data and service providers within the EOSC-hub Consortium. One particular section on the Why-card does not find its origin in the first data policy recommendations from D2.8. The "Trust and Confidence" section on the Why-card addresses the importance of service accreditation and formalising policies according to certain standards (CoreTrustSeal, ISO27001 and ISO16363). Nevertheless we've decided to make it part of D2.9 since this section on the adoption of formal data sharing policies as a step towards service accreditation provides a solid answer to the question "Why should you care about implementing data sharing policies?".

Design methodology and mapping

To assist the service providers engaged in the EOSC-hub ecosystem as practically as possible, we mapped the three broad headings of D2.8 (see above) to three questions: "Why?", "What?", and "How?":

1. "**Why** should [providers] care about implementing data sharing policies?" maps to: "Support the wider development of ethical and information governance."
2. "**What** are essential concepts that [service providers] should be aware of in the context of data sharing policies in the EOSC-hub ecosystem?" This question maps to: "Implement FAIR".

3. "**How** should providers of data and services bridge the gap between future technical implementation and policy recommendations for data sharing pragmatically?" This question maps to: "Build technical expertise in 'safe data' and 'safe settings'".

The Why and the What questions have their own reference card. The How question has two cards. For a more detailed mapping of the broad headings of D2.8 into the categorisation of the Service Provider's Guide to Data Sharing Policies, see Appendix I.

The reference cards below provide reference to sources - international initiatives, concepts, outputs - that might be able to help the data and service providers answer the main three questions: why, what, and how.

We thought of a way to comprehensively translate the recommendations from D2.8 into practical questions for providers in the EOSC-hub consortium. To group them the way as it is now, was a pragmatic choice in a similar sense as the design choice. With the reference cards, we offer data and service providers in the EOSC-hub consortium the 'low hanging fruit'. Without going too much into detail, they are offered a palette of information, aiming to provide them at least the three 'major' recommendations from D2.8 (implement FAIR, build technical expertise, support wider ethical governance development frameworks).

Underlying all this is the vast knowledge, created by parties involved, and we deemed it impossible to offer the providers granularity and good reference at the same time, without losing the conciseness of the reference cards. This perhaps results in a condensed deliverable making it understandable only by those who already know the presented aspects.

The Service Provider's Guide to Data Sharing Policies: four reference cards

The four reference cards have been developed in close collaboration with TRUST-IT. On the next four pages, you will find the reference cards in the following order:

1. A Service Provider's Guide to Data Sharing Policies: **The Why** / What / How
2. A Service Provider's Guide to Data Sharing Policies: Why / **The What** / How
3. A Service Provider's Guide to Data Sharing Policies: Why / What / **The How**

   Note that there are two **How** reference cards.

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.

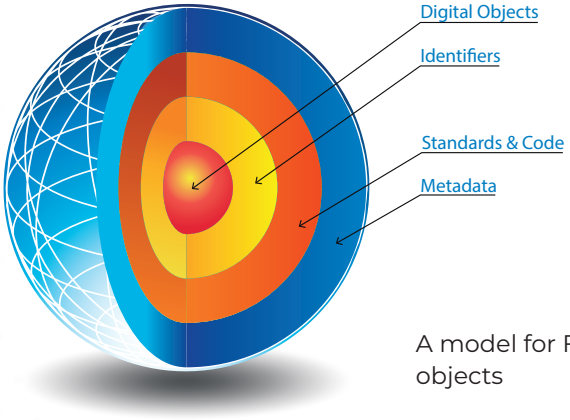
Our recommendations contribute to the developing field of data sharing policies in the EOSC at large.

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

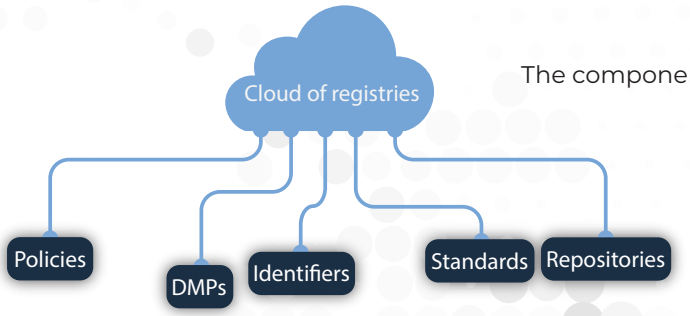


A model for FAIR digital objects

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].



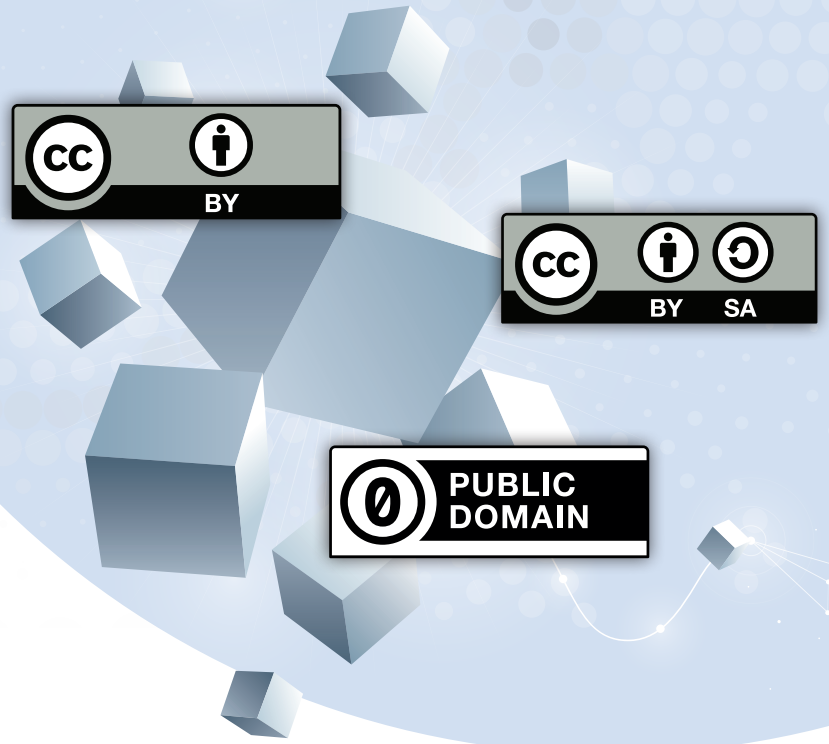
The components of a FAIR ecosystem

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

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**

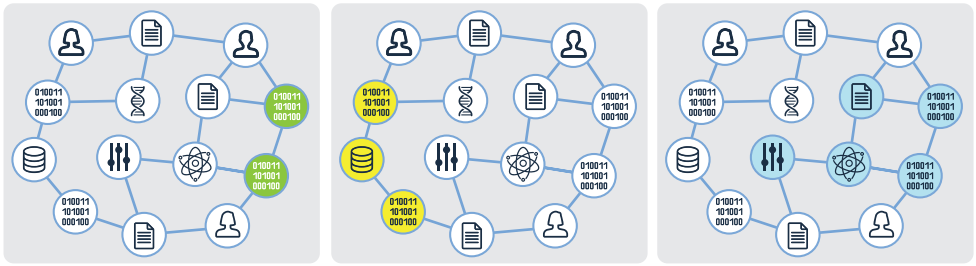
How can you implement data sharing policies practically? Some examples.

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.

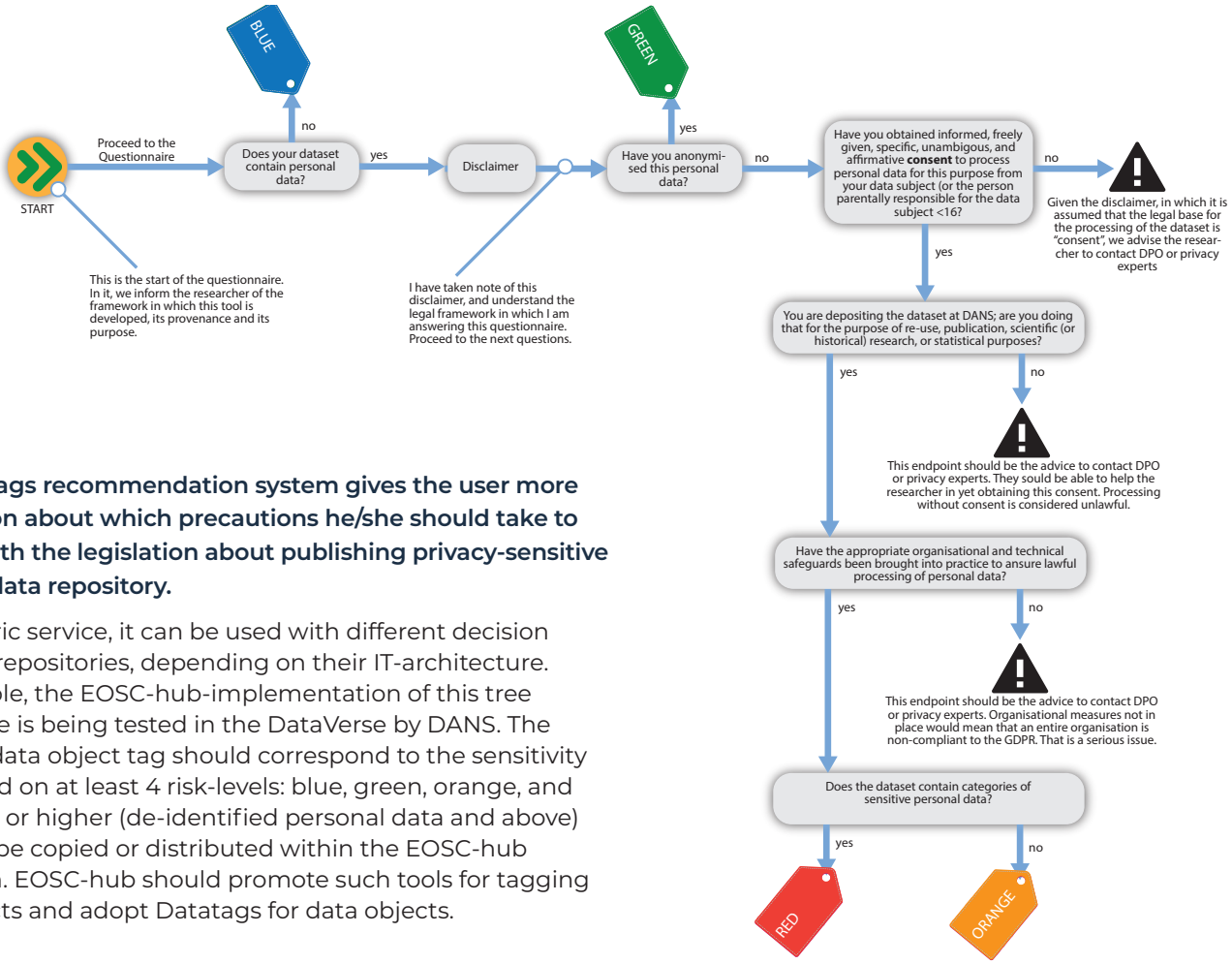
Additionally, EOSC-hub should track the work of the FREYA project and adopt best practices in PID resolution as they emerge. One of the outputs of the FREYA project, is the PID-graph.



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.

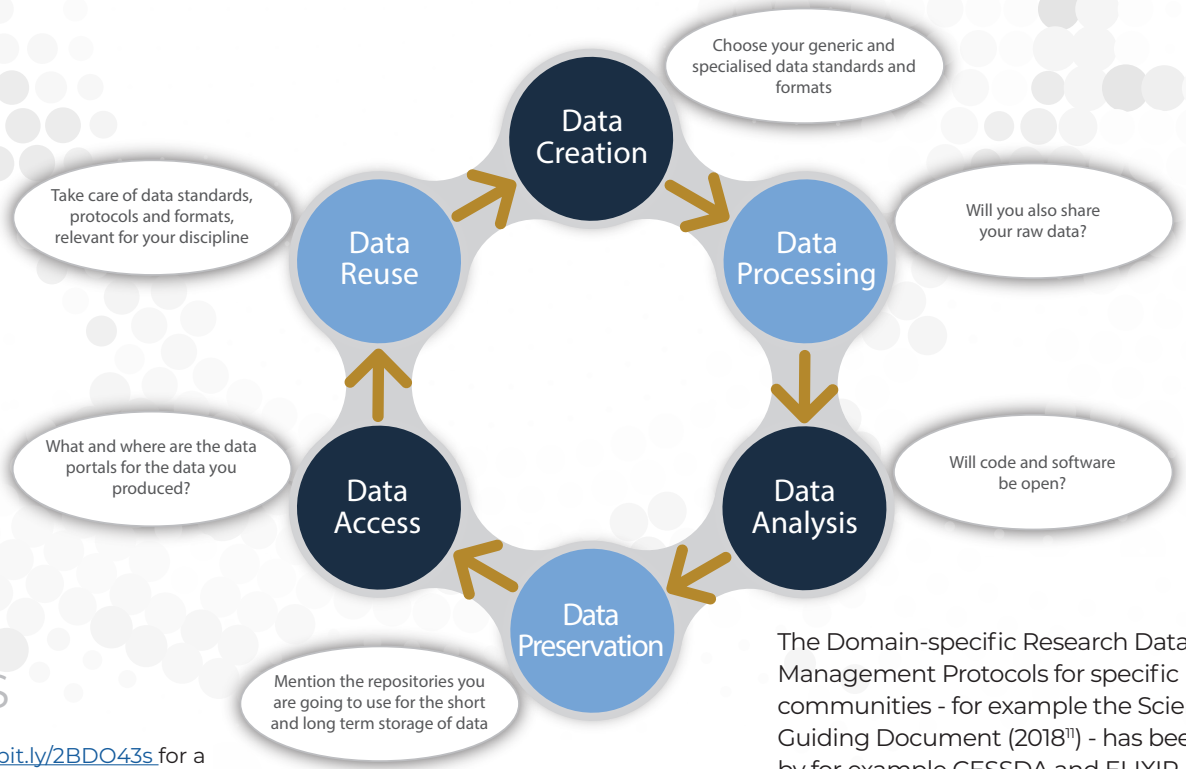
For more details, consult the D2.8 and D2.9 deliverables via: https://bit.ly/2zDAAnM

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

Next steps

The First Policy recommendations (EOSC-hub D2.8) have been translated into four reference cards that together form the Service Provider's Guide to Data Sharing Policies. These cards (The Why, The What and The How) convey data sharing recommendations to be adopted by data and service providers within the EOSC-hub consortium.

We suggest an online dissemination strategy for these cards, although they are also designed in such a way that they can be printed on (A4) paper. We will showcase the cards at a variety of EOSC and FAIR related (online) activities and events, providing appropriate visibility to all stakeholders. Also, further online dissemination (e.g. social media presence, webinars, video interviews) will be carried out to facilitate adoption and to obtain user feedback. Another possibility is to organize and roll-out an online focused workshop. Details of the communication plan of these cards are to be worked out in accordance with EOSC-hub project governance and management.

References

[1] R Baxter et al, D2.8: First Data policy recommendations, EOSC-hub, December 2018, https://documents.egi.eu/document/3419

[2] S Battaglia et al, D3.3: Draft Policy Recommendations, EOSCpilot, August 2018, https://eoscpilot.eu/content/d33-Draft-Policy-Recommendations

[3] S Hodson, S Jones et al, Turning FAIR into reality, European Commission Expert Group on FAIR Data, November 2018, https://doi.org/10.2777/1524

[4] RDA FAIR Data Maturity Model WG: https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

[5] The FREYA project: https://www.project-freya.eu

[6] The FAIRsFAIR project: https://www.fairsfair.eu

Appendix I: Table mapping of the recommendations and reference cards

EOSC-hub D2.8: Support the wider development of ethical and information governance frameworks	
EOSC-hub D2.9: Reference Card "The Why"	
Rec 16	EOSC-hub should consider the logging and tracking of scientific provenance data as an element of service integration design.
Rec 17	EOSC-hub should consider convening a technical working group on the topic of recording provenance across the EOSC-hub service ecosystem.
Rec 21	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.
Rec 22	EOSC-hub should consider the feasibility and desirability of developing a code of conduct for sensitive data research.

EOSC-hub D2.8: Implement FAIR	
EOSC-hub D2.9 Reference Card "The What"	
Rec 3	All shareable data objects in the EOSC-hub ecosystem should be FAIR Digital Objects.
Rec 4	Data objects should be minimally described by a metadata record that follows the recommended schema in Appendix I.
Rec 5	A data object's metadata record should be the minimally required description of it in a data catalogue.
Rec 6	A data object should have a unique persistent identifier.
Rec 7	A data object's persistent identifier must form part of its metadata record.
Rec 8	An http GET request on a data object's persistent identifier should return an HTML landing page that can be rendered in a standard Web browser.
Rec 9	A data object's HTML landing page should encode its metadata record according to the schema.org approach.

Rec 10	An http GET request on a data object's persistent identifier accepting a different return format (e.g. XML or JSON) should return the data object's metadata record in that format (content negotiation).
Rec 11	EOSC-hub should track the work of the FREYA project and adopt best practices in PID resolution as they emerge.
Rec 12	EOSC-hub should initiate a programme of technical research for metadata discovery that builds on the Elixir Beacon approach for cases where metadata records may themselves contain sensitive data.
Rec 13	EOSC-hub should initiate a programme of technical research for direct retrieval of data objects by PID.
Rec 14	Data objects should be published on the Web in an open, non- proprietary format chosen to suit its content or subject.
Rec 15	Data objects in the EOSC-hub ecosystem should adopt licenses from the Creative Commons 4.0 license suite. Where data are openly sharable, 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)

EOSC-hub D2.8: Build technical expertise in 'safe data' and 'safe settings'	
EOSC-hub D2.9 Reference Card "The How"	
Rec 1	EOSC-hub should adopt a DataTag system for data objects based on at least four risk levels: blue, green, orange, and red.
Rec 2	EOSC-hub should promote the use and development of decision-tree based tools for tagging data objects. A GDPR tool should be based on existing work at DANS; other tools should be developed in collaboration with relevant scientific and research communities.
Rec 18	Data objects with tags of green or higher (de-identified personal data or above) should not be freely copied or distributed within the EOSC-hub ecosystem.
Rec 19	EOSC-hub should adopt the Five Safes principles as guidance for the management and handling of sensitive data in the EOSC-hub ecosystem.
Rec 20	EOSC-hub should develop a technical design framework for Safe Haven services to support the 'safe data' and 'safe settings' dimensions of the Five Safes Principles.