FAIRSFAIR

Fostering Fair Data Practices in Europe

FAIR assessment framework for data services

Patricia Herterich (DCC) and Hylke Koers (SURF)

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Agenda

- Welcome & introduction (10 min)
- Methodology & findings (15)
- Basic framework (10)
- Q&A (5)
- Your feedback (15; interactive)
- Wrap up (5)

Patricia Herterich, DCC

Hylke Koers, SURF

FAIRsFAIR in a nutshell

Call: H2020-INFRAEOSC-5c

Budget: 10 million euro

Length: 36 months
Starting date: March 1 2019

6 core partners/WP leads

FAIRSFAIR
Fostering Fair Data Practices in Europe

FAIRsFAIR partners

Data Archiving and Networked Services
DANS

Trust-IT Services
Communicating ICT to markets

Universiteit van Amsterdam

DCC

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

eua EUROPEAN UNIVERSITY ASSOCIATION

uc3m | Universidad Carlos III de Madrid

CSC

Inria

Universidade do Minho

SURF

SPARC Europe

University of Essex

DataCite
FIND, ACCESS, AND REUSE DATA

escience data FACTORY

UKRI Science and Technology Facilities Council

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN

INES
Centre Informatique National de l'Enseignement Supérieur

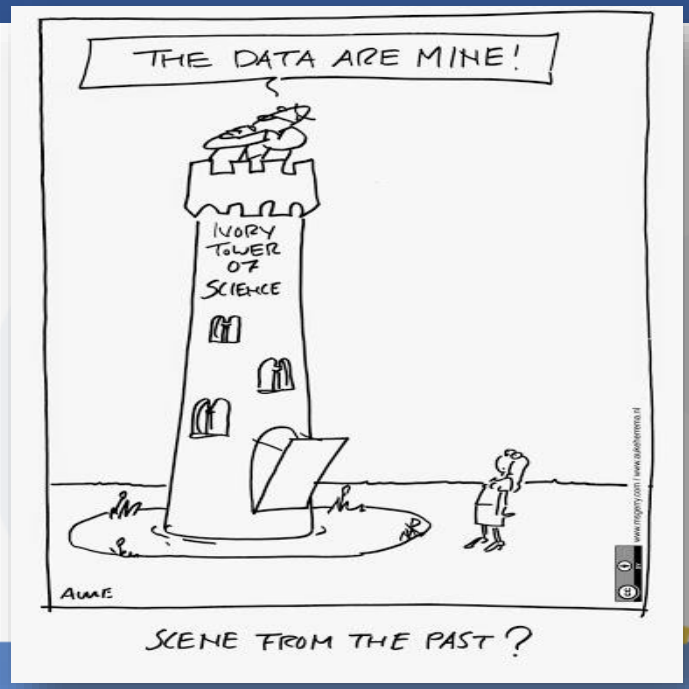
RDA
RESEARCH DATA ALLIANCE

DTL

Universität Bremen

Our objective

To supply practical solutions for the use of the FAIR data principles throughout the research data life cycle. Emphasis is on fostering FAIR data culture and the uptake of good practices in making data FAIR.



The FAIR symphony

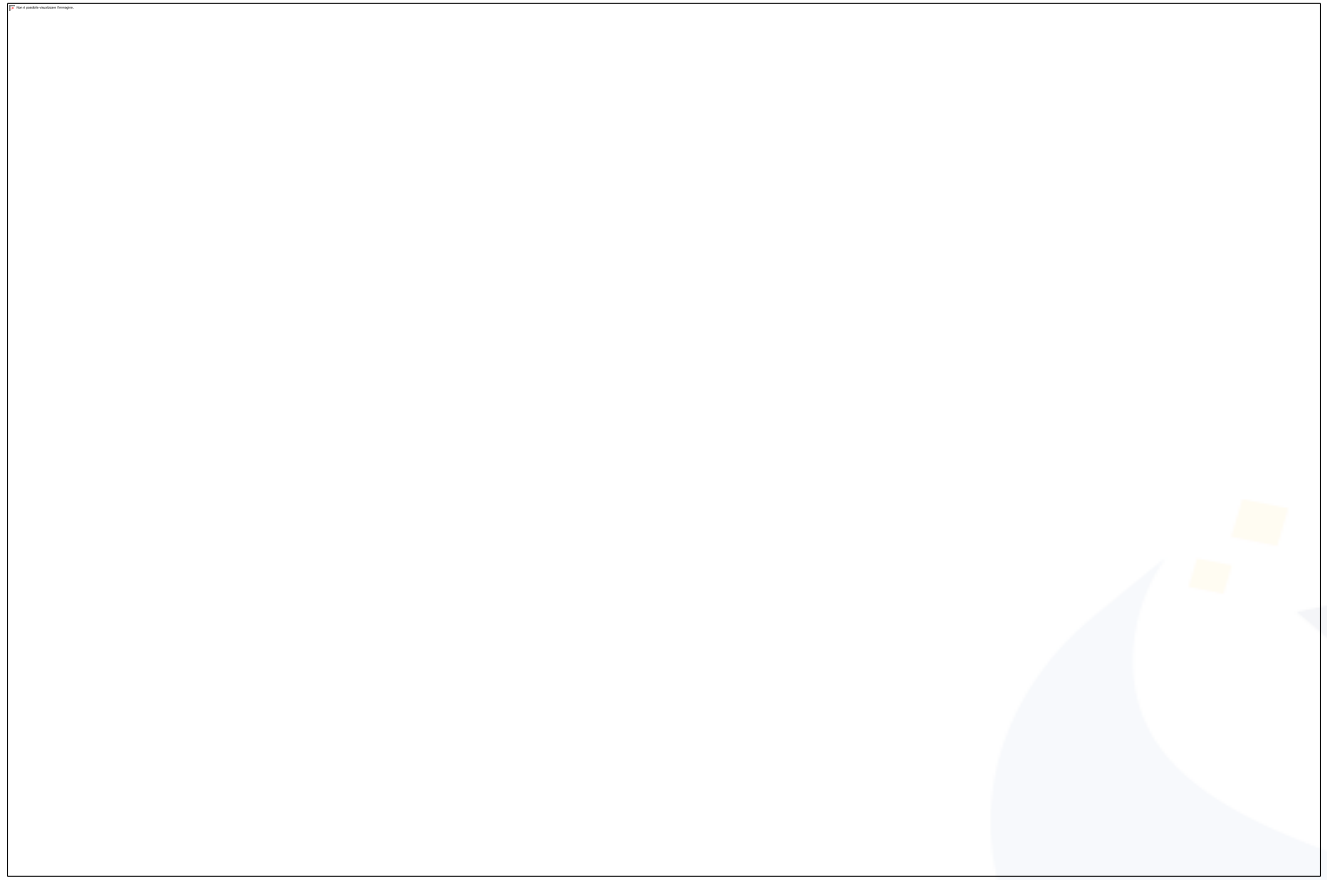


Image: European Union Youth Orchestra, euyo.eu

The FAIR symphony needs FAIR services

Priority Recommendation

Rec. 13: "Develop metrics to certify FAIR services": More work is needed to extend the FAIR data principles for application to a wide range of data services, including registries, Data Management Planning tools, metadata standards and vocabulary bodies, identifier providers, software libraries and other cloud services…

(Also recently underlined by EOSC FAIR WG)

Guidance for service owners to enable FAIR

- FAIR is not an absolute, but rather a set of guiding principles that need further interpretation and definition
- A lot of work on this has been done for datasets and other digital objects
- For service owners, there is currently little guidance on how to make their service fit in the FAIR data ecosystem

(illustration from vecteezy)

Objective

To deliver an assessment framework for data services that will help service owners to incrementally improve their services

→ *stimulating an optimal interplay between digital objects and services*

→ *help realize the full potential of a truly FAIR ecosystem*

Output 1: Case studies and methodology for ‘FAIR enablement’

4.2. Case Study 1: B2FIND

Service Summary

B2FIND[21] is a metadata aggregator. The service harvests metadata from different community repositories and harmonises them such that users and services can search through the combined metadata. B2 FIND offers a rich faceted graphical search interface and a HTTP REST API that has been implemented in python for EUDAT’s B2FIND Training[22]

URL: http://b2find.eudat.eu/ **EOSC:** https://marketplace.eosc-portal.eu/services/b2find

Users

The service targets two types of user groups:

- Scientific communities that can provide their metadata and integrate via the B2FIND service with other metadata
- Scientists who can employ the service to search for interesting research data across different communities simultaneously.

Services

- Metadata harvesting and harmonisation to communities with a tool to search across the metadata for scientists.
- The relevant metadata of a DO is shown and a link to the metadata provenance is provided.

Target Digital Objects

- Metadata entries

Examples

- B2FIND entry (KONTROL 1984[23])
- OAI-PMH dataset’s metadata[24]

Purpose

B2FIND is a metadata aggregator. It gathers metadata from communities and repositories and integrates the different types of metadata. It provides a graphical user interface and an API to present the metadata and allows faceted searches across the metadata corpus.

Adoption

By now B2FIND hosts 824566 metadata entries harvested from 22 communities. We were unable to establish from the documentation how many users use B2FIND.

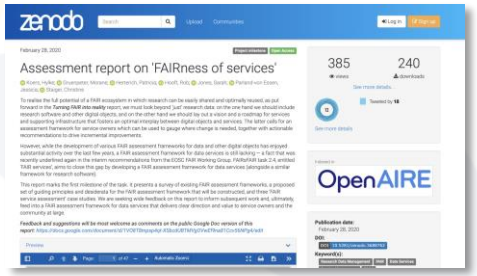
Documentation

EUDAT provides guidelines on how to use the B2FIND services[25] as well as detailed guidelines for harvesting and mapping metadata[26]

FAIR enablement mapping (see Annex C for details)

F1	F2	F3	F4	A1	A1.1	A1.2	A2	I1	I2	I3	R1	R1.1	R1.2	R1.3

FAIR enablement mapping:
Enable / Respect / Reduce



“M2.7 Assessment report on 'FAIRness of services'”, available at: https://doi.org/10.5281/zenodo.3688762

Output 2: Basic framework for FAIR service assessment

(intentionally left blank… for now...)

FAIRSFAIR

Fostering Fair Data Practices in Europe

FAIRness of Services: Methodology and Findings

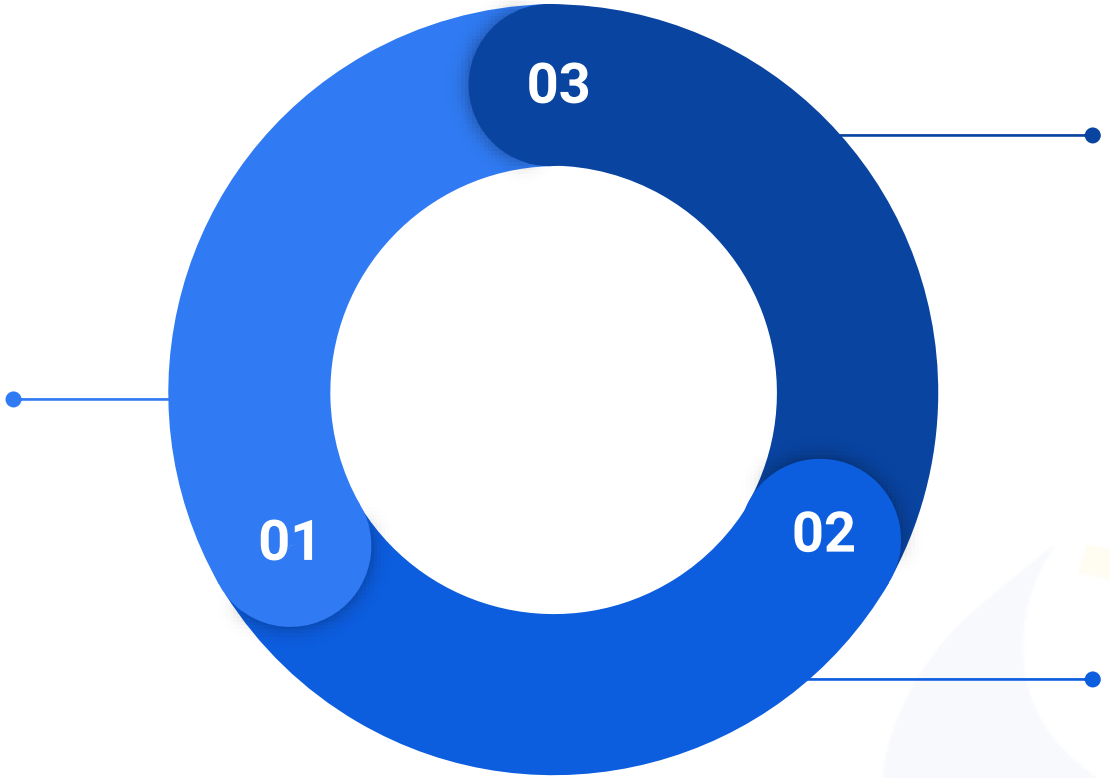
Patricia Herterich

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Building blocks

Structured literature review

Covering earlier work on service assessment such as CoreTrustSeal, the TRUST principles - but also recent work from EOSC-synergy, EOSC-nordic and other related projects.



EOSC-hub week workshop

Feedback from FAIRsFAIR workshop on FAIR certification of repositories and other data services during EOSC-hub week in May 2020

Interviews with service owners

Semi-structured interviews with a range of service owners

Literature review

- Covering 18 papers and reports
- Analysis
  - Extracting statements (249 in total)
  - Coding by type of statement (recommendation, requirement, principle, action, etc.) and applicability to a wider range of data services

	A	B	C	D	E	F
1	Source	Identifier	In scope	If 'no', why?	Where addressed?	Statement
2	EOSC-Portal	EOSC-Portal: 1	Yes		Trustworthiness	The service is accessible by users outside its original community.
3	EOSC-Portal	EOSC-Portal: 2	Yes		Trustworthiness	The service is described through a common template focused on value proposition and functional capabilities.
4	EOSC-Portal	EOSC-Portal: 3	Yes		(Part of scope)	At least one service instance is running in a production environment available to the user community.
5	EOSC-Portal	EOSC-Portal: 4	Yes		FAIR-enablement	Publish Research data is Findable, Accessible, Interoperable and Reusable [reference to FAIR].

Dataset: Koers, H., Herterich, P., Hooft, R., Gruenpeter, M., & Aalto, T. (2020). Collected recommendations and requirements for FAIR-enabling services (Version 1.0). https://doi.org/10.5281/ZENODO.4293778

Findings from literature review

- Various projects working on FAIR data object assessment, less focus on FAIR-enabling services
- Work that has been done is characterized by:
  - Broad diversity in type of statements → difficult to compare and harmonize
  - Some resources focus on certain aspects potentially resulting in an unbalanced view
  - Unclear process to implement recommendations and requirements
  - Lack of case studies reporting benefits and challenges
  - Unclear or abstract relation to EOSC work (e.g. EOSC rules of participation or service management onboarding requirements)

Interviews

- 5 interviews with service owners, approx. 60 minutes each
- Semi-structured interviews covering the following parts:
    1. Understanding the service, its users and context;
    2. Understanding the service maturity;
    3. Understanding affinity and familiarity with FAIR;
    4. Soliciting viewpoints on a FAIR assessment framework.

Analysis

- Key statements from each interview
- Coded by

Type	Applicability	Progress	In Scope	If No, why?
What do they offer - technical	Generic	unknown	Yes	Applies only to certain services
What do they offer - social	F Findability	vision	No	Not directly for service owners
What do they want - technical	F1 PIDs	in progress		Not defined enough
What do they want - social	F2 Findability metadata	in production		Too detailed
	F3 Searchable resource	none		
	F4 Links between (meta)data			
	A Accessibility			

Dataset: Herterich, P., Aalto, T., & Koers, H. (2020). Analysed data from interviews on FAIR-enabling services (Version 1.0). https://doi.org/10.5281/ZENODO.4293788

Findings from interviews

- Services have been supporting aspects of FAIR before the term was coined
- They have a good understanding of their communities and awareness of the wider policy context
- Most find that automated FAIRness assessment tools could be helpful
- Training for users will be needed
- Sustainable funding will be crucial
- External and community-endorsed assessment framework is preferred over self-assessment

Fin

"It's like a driving instructor. They are allowed to pass things, you know, they're allowed to grant to FAIR badges. But it's not the car company giving you FAIR. It's not the car company saying here's a car, and we'll give you a driver's licence. It's: here's a car, you can pass your driver's licence using this because it passed its MOT and it's all the level that it's safe to drive on the roads. We need the FAIR data equivalent of driving instructors". (M. Hahnel)

FAIR before the term was

ommunities and

essment tools could

- Training
- Sustainable funding will be crucial
- External and community-endorsed a preferred over self-assessment

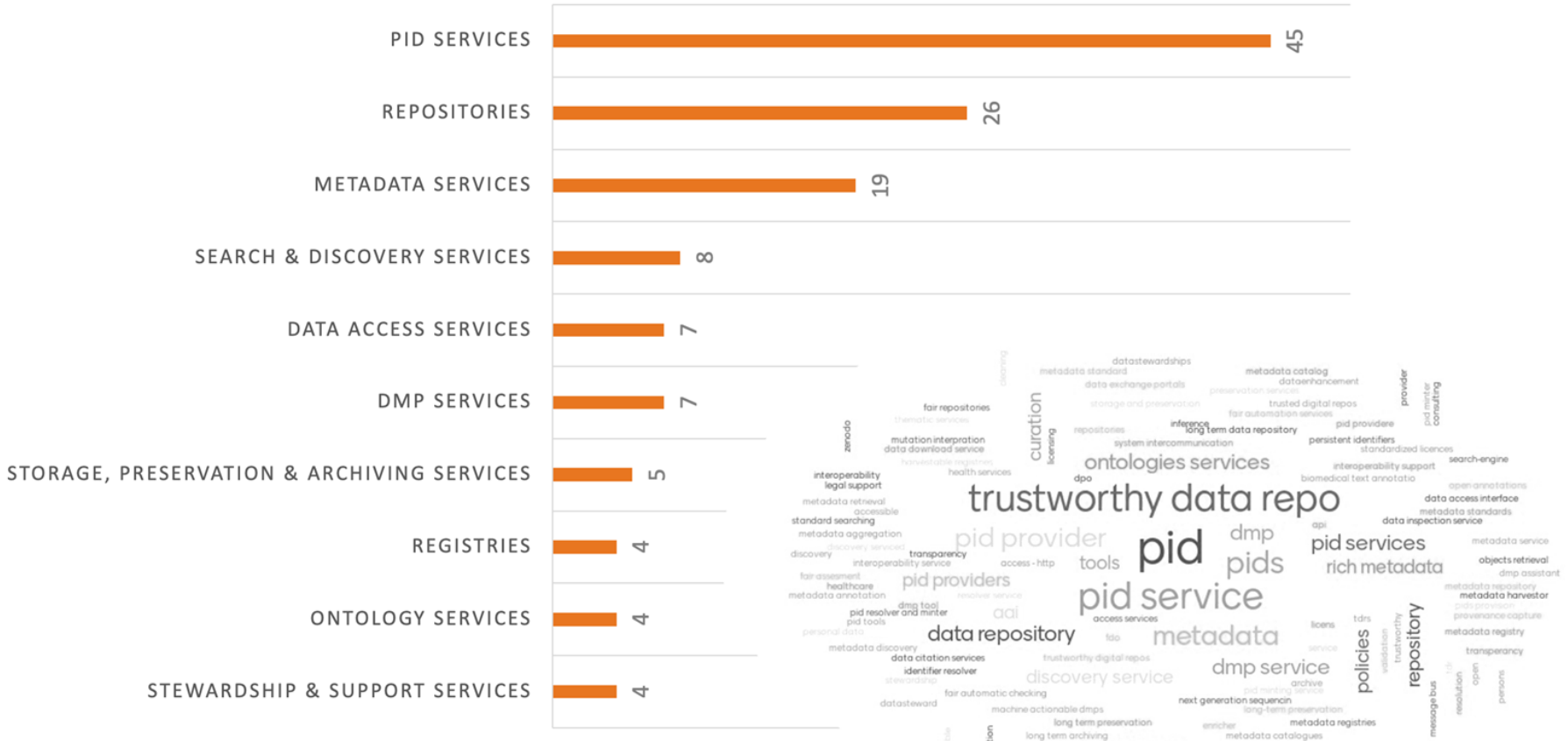
"We want a test that we are able to pass, but we nevertheless want it to be quite serious. Nothing is more horrible to an honest operator than a test that doesn't actually test. Self-assessments with free form questions [...] tend to select for people who know how to advertise themselves well." (G. Aben)

EOSC-hub session

- Feedback on:
  - Scope, i.e. which services are seen to be essential in enabling FAIR data and should thus be in scope for the assessment framework;
  - Desired qualities, i.e. what are seen to be important properties for a data service to be an enabler for FAIR data;
  - Form of the assessment framework, specifically on the continuum from descriptive (sharing recommendations and good practices) to prescriptive (formal certification)
- 90 active contributors, 44 % of which identified as service providers

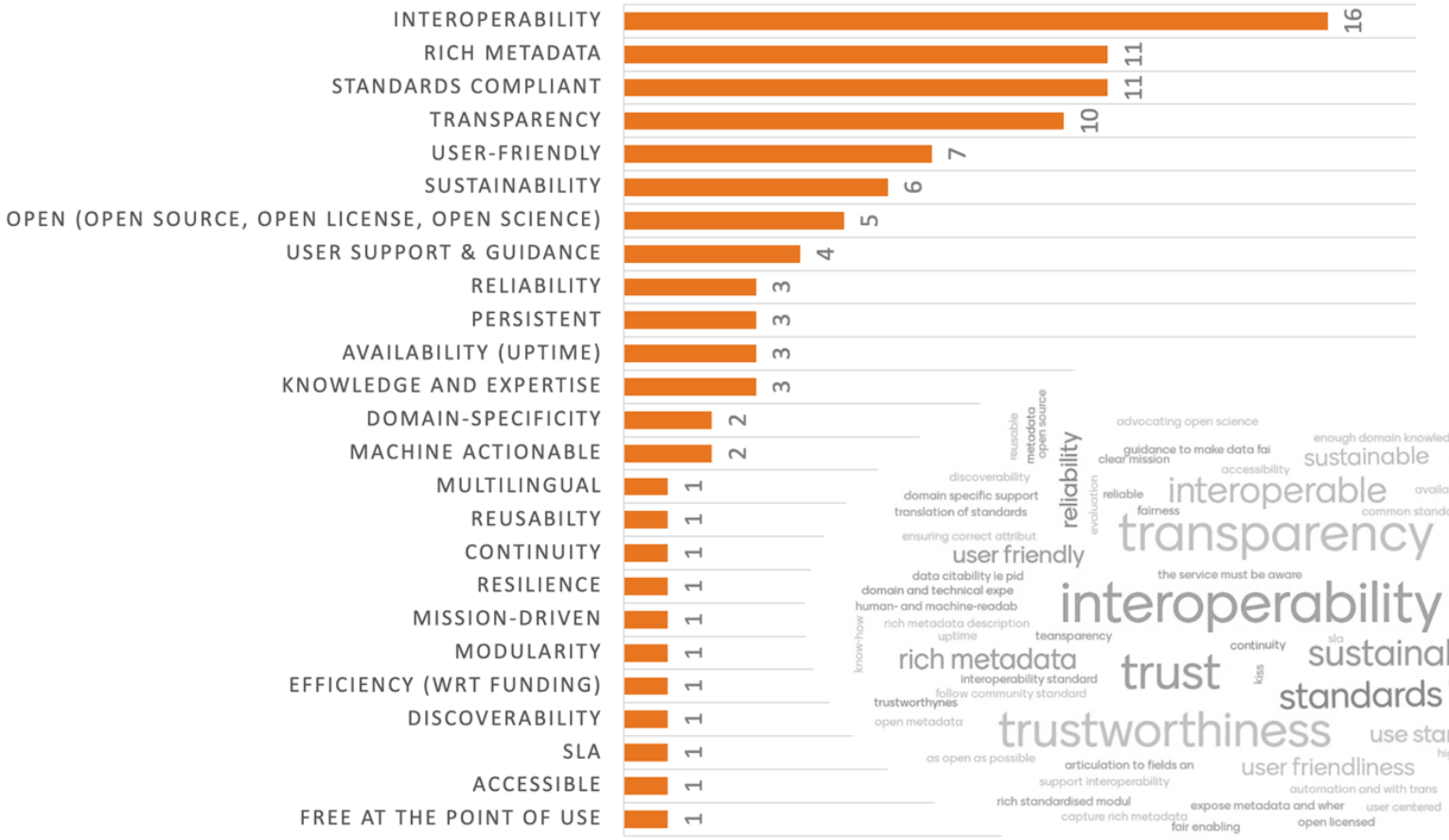
Findings - scope

Please name three types of data services that you consider essential to enable FAIR



Findings - desired qualities

What do you consider to be the most important qualities for a data service to enable FAIR data?



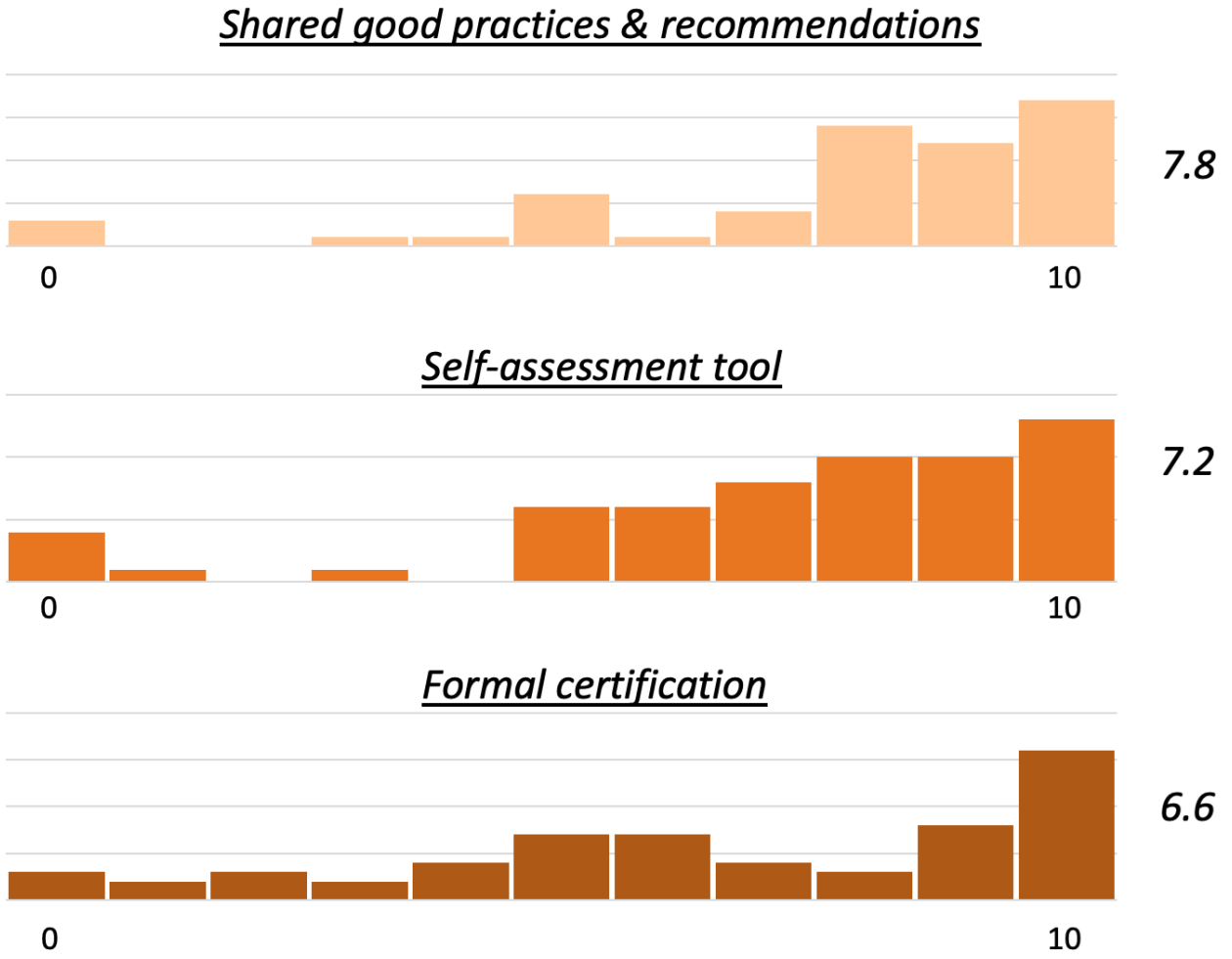
Findings - form

How important is it for you that

(i) there are shared 'good practices' and recommendation for FAIR-enabling services;

(ii) There is a self-assessment tool for FAIR-enabling services;

(iii) There is a formal certification process for FAIR-enabling services



FAIRSFAIR

Fostering Fair Data Practices in Europe

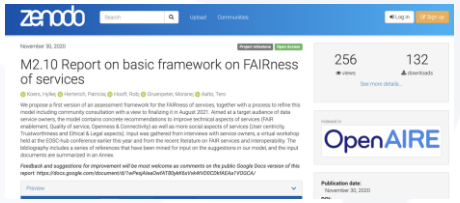
FAIRness of Services: Basic Assessment Framework

Hylke Koers

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Proposing: A basic framework for FAIR service assessment

Technically-oriented	*Socially-oriented*
FAIR enablement	User centricity
Quality of service	Trustworthiness
Open & Connected	Ethical & Legal



"M2.10 Report on basic framework on FAIRness of services", available at: https://doi.org/10.5281/zenodo.4292599

Each aspect has a high-level objective with actionable recommendations

- High-level objective
- Actionable, detailed recommendations

FAIR enablement

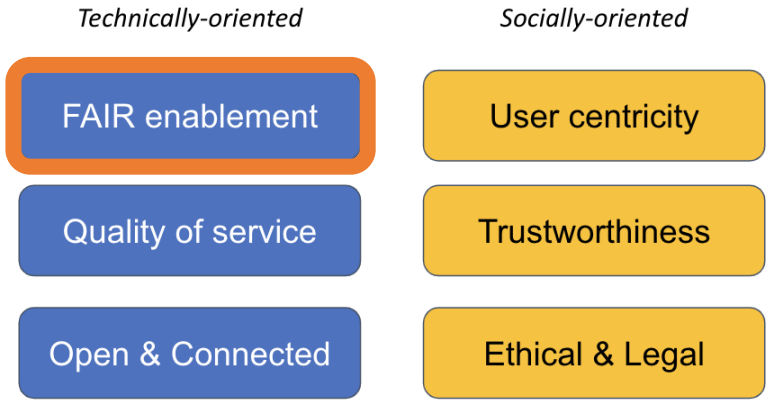
Objective:

The service enables FAIR data by elevating the FAIRness of digital objects and/or supporting the FAIRification process. FAIR enablement is actively driven through the implementation of community-supported standards and interoperability frameworks.

Recommendations:

- Perform a self-assessment on how the function(s) of the service *enable, respect* or *reduce* each of the FAIR principles for the data that it operates on.[12] Make the results of the self-assessment publicly available, together with an outlook on the desired state for the service (including a cost/benefit analysis).[13]
- Use automated tests that show how the service increments FAIRness of digital objects in a verifiable, measurable, repeatable and scalable way. Root such tests in community-supported methodologies that measure the FAIRness of digital objects in an objective way.
- In consultation with the target community (or communities), identify which metadata

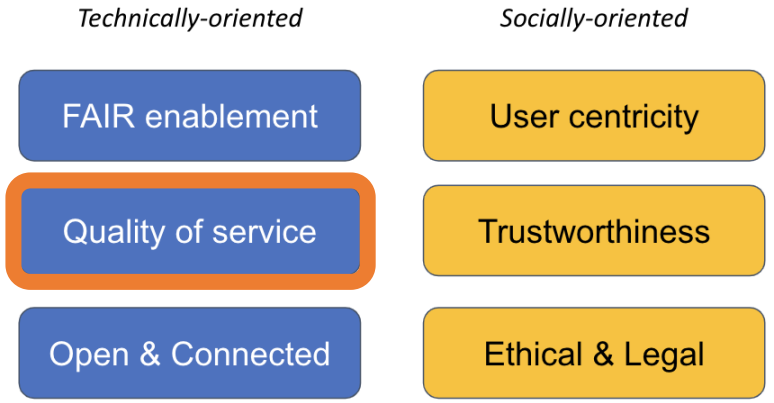
FAIR enablement



Objective:

- *The service **enables FAIR data** by elevating the FAIRness of digital objects and/or supporting the FAIRification process. FAIR enablement is actively driven through the implementation of **community-supported standards and interoperability frameworks**.*

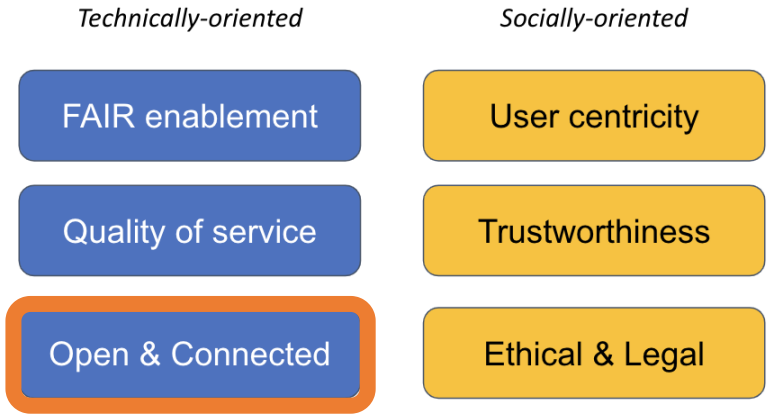
Quality of service



Objective:

- *The service is delivered in a **reliable, secure, high-quality way**, consistent with its specifications.*

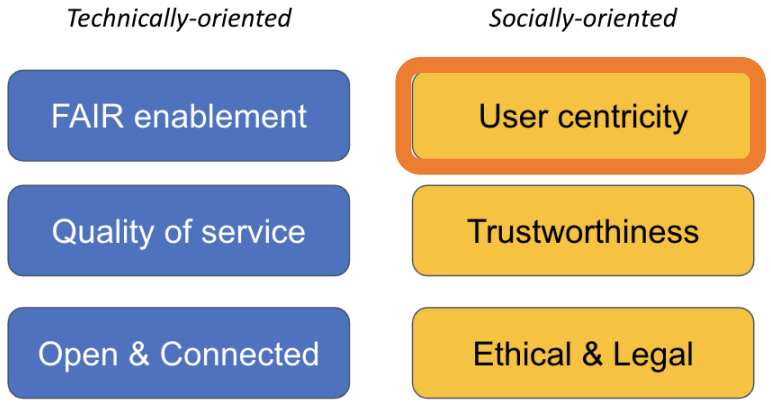
Open & Connected



Objective:

- *The service is operated in a **transparent**, **low-barrier** and **inclusive** way; seeking **integrations** and **connections** with other services; and championing principles of openness consistent with **Open Science** and **Open Research**.*

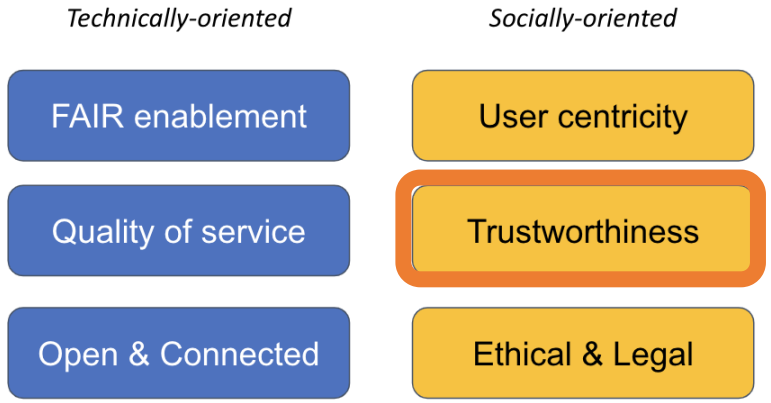
User centricity



Objective:

- *The service is managed such that it serves the (possibly evolving) goals of the* ***user community****, and* ***maximises usability while minimizing burden.***

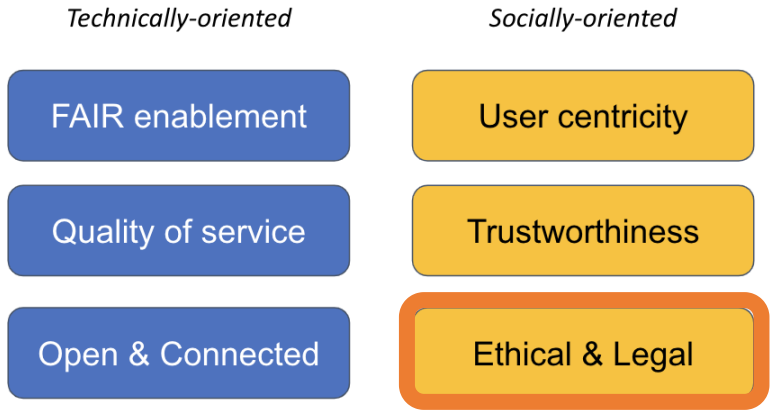
Trustworthiness



Objective:

- *The service is perceived by the user community as **reliable** and **trustworthy**, both in terms of its utility and its warranties, **now and in the future**.*

Ethical & Legal

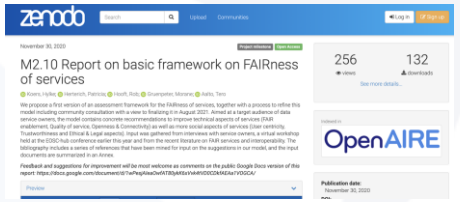


Objective:

- *The service complies with all applicable **legal** and **ethical guidelines**, in a **transparent** and **auditable** way.*

A basic framework for FAIR service assessment

Technically-oriented	*Socially-oriented*
FAIR enablement	User centricity
Quality of service	Trustworthiness
Open & Connected	Ethical & Legal



"M2.10 Report on basic framework on FAIRness of services", available at: https://doi.org/10.5281/zenodo.4292599

Next steps

Dec – Jan 2020	Public consultation on assessment framework as proposed in this work – including webinar to kick off the public consultation (perhaps with RDA FAIR maturity WG)
Feb 2021	Workshop with different stakeholder groups to validate completeness
March 2021	Dissemination of workshop findings, including next iteration of the assessment framework
April 2021	Second workshop for service owners to validate utility and actionability
May 2021	Draw up final framework for assessing FAIR services
June 2021	"Last chance" external expert review
June 2021	Internal review within FAIRsFAIR
July – Aug 2021	Finalization and publication of framework for assessing FAIR services (deliverable D2.7)

Next steps

Dec – Jan 2020	Public consultation on assessment framework as proposed in this work – including webinar to kick off the public consultation (perhaps with RDA FAIR maturity WG)
Feb 2021	Workshop with different stakeholder groups to validate completeness
March 2021	Dissemination of workshop findings, including next iteration of the assessment framework
	Second workshop for service owners to validate utility and actionability
	Draw up final framework for assessing FAIR services
	"Last chance" external expert review
	Internal review within FAIRsFAIR
July – Aug 2021	Finalization and publication of framework for assessing FAIR services (deliverable D2.7)

February 19

Let us know (e.g. via chat) if you're interested to join!

FAIRSFAIR

Fostering Fair Data Practices in Europe

Questions?

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

FAIRSFAIR

Fostering Fair Data Practices in Europe

Your Feedback

Hylke Koers

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Please visit www.menti.com → use code 92 94 76 6

FAIRSFAIR

Fostering Fair Data Practices in Europe

FAIRness of Services: Next Steps

Hylke Koers

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558

Next steps

Dec – Jan 2020	Public consultation on assessment framework as proposed in this work – including webinar to kick off the public consultation (perhaps with RDA FAIR maturity WG)
Feb 2021	Workshop with different stakeholder groups to validate completeness
March 2021	Dissemination of workshop findings, including next iteration of the assessment framework
April 2021	Second workshop for service owners to validate utility and actionability
May 2021	Draw up final framework for assessing FAIR services
June 2021	"Last chance" external expert review
June 2021	Internal review within FAIRsFAIR
July – Aug 2021	Finalization and publication of framework for assessing FAIR services (deliverable D2.7)

Finally, thanks to “team 2.4” !

- Tero Alto (CSC)
- Morane Gruenpeter (INRIA)
- Patricia Herterich (DCC)
- Rob Hooft (DTL)
- Hylke Koers (SURF; task lead)

And earlier contributions from:

- Christine Staiger (DTL)
- Roberto Di Cosmo (INRIA)
- Sarah Jones (DCC)
- Jessica Parland-von Essen (CSC; work package lead)
- Jonas Tana (CSC)