eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

Synchronisation Workshop 2023

Persistent Identifiers

November 30th 2023

Josefine Nordling and Liisa Marjamaa-Mankinen, CSC

Funded by the European Union

Welcome to Session Three: Persistent Identifiers

Josefine Nordling (CSC) - chair
Liisa Marjamaa-Mankinen (CSC) - rapporteur

- The goal of this workshop is to discuss trust in and sustainability of PID systems, when defining PID Policy assessment criteria
- This is an interactive workshop; discussion and adding content is warmly welcomed.

Program for today

- Compliance Assessment Toolkit (CAT) - Wim Hugo (DANS)
- CAT in research workflows - workshop outcomes - Natascha van Lieshout (SURF)
- Presentation on PID use cases in FAIR-IMPACT and trust in PID infrastructures (KE report) - Josefine Nordling (CSC)
- Reflections from PID use case partners on trust and sustainability in PID systems - **Everyone - feel very free to join in on the discussion!**
  - Hervé l'Hours (CESSDA ERIC - PIDs in sensitive data)
  - Renato Juacaba Neto (EMBL-EBI - PIDs in data production workflows & sensitive data)
  - Parham Ramezani (LifeWatch ERIC - PIDs in complex data citation)
  - Nick Juty (UNIMAN - PIDs in data production workflows)
- Discussion about the three pre-asked questions

Some details for this session.

- The session will be recorded but only for internal use for the rapporteur and the report writing.
- Shared spreadsheet and note taking document for use in the session (and afterwards).
  - In the spreadsheet please keep information factual, short and include links wherever possible.
  - In the note taking document you may add more detail and background.
  - Please be careful when editing spreadsheet cells.
  - Please refrain from editing other people's information - even typos.

Some more details for this session.

- You may wish to add more information later, spreadsheet & note taking document will be available until **December 10th**
- Survey responses have been added to the spreadsheet
- 3 questions but you may not have answers for them all

The main questions

1. What does your project or initiative do to implement PIDs? Please provide any relevant links.
2. What are the guiding principles for you when choosing a PID system or service?
3. What does in your opinion constitute trust in a PID service?

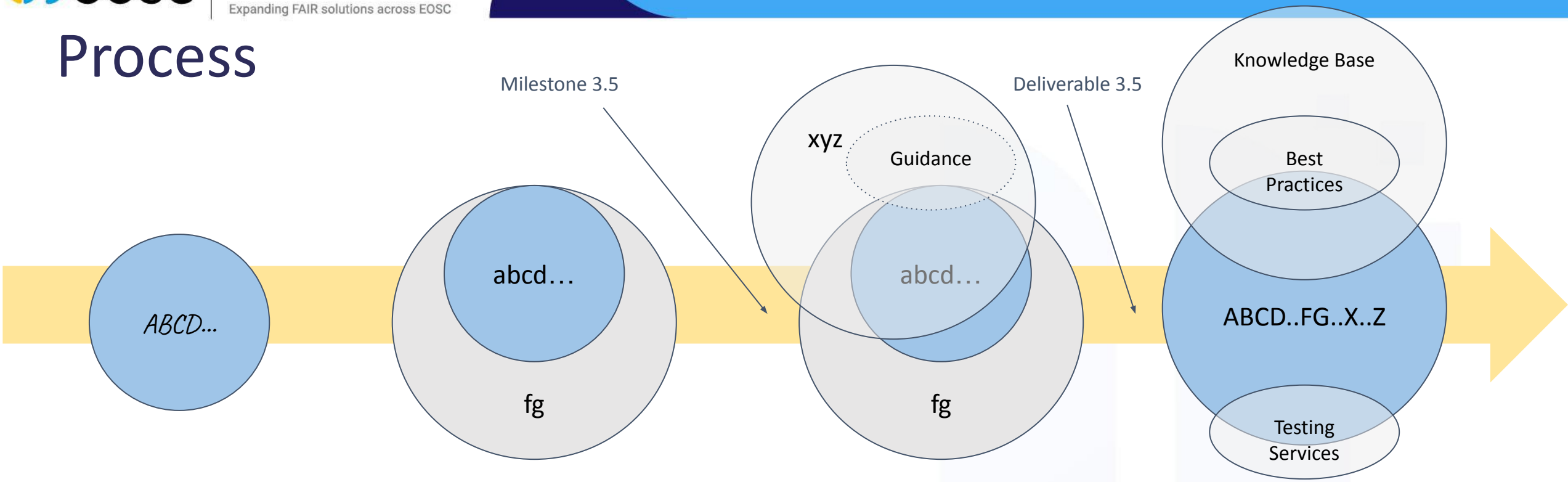
Compliance Assessment Toolkit

Wim Hugo (DANS)

Funded by the European Union

Context and Process

Process



European Commission, Directorate-General for Research and Innovation, Hellström, M., Heughebaert, A., Kotarski, R., et al., **A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC),** Publications Office, 2020, https://data.europa.eu/doi/10.2777/926037

Hugo, W., Steinhoff, W., Turner, D., Buys, M., & Zamani, T. (2023). D2.1 **Compliance Assessment Specification**. Zenodo. https://doi.org/10.5281/zenodo.10067253

FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

Work in Progress: FAIR-IMPACT WP3

Community Expectations
Use Cases
Workflows
PID Policies
Best Practices

FAIR-IMPACT
Expanding FAIR solutions across EOSC

Work in Progress: FAIR-IMPACT WP3

PID Knowledge Base integrated with CAT

FAIR-IMPACT
Expanding FAIR solutions across EOSC

FAIRCORE4EOSC
Core Components Supporting a FAIR EOSC

C1: Minimum Operations Are Available

Provider	SHOULD

#	Principle or Objective	Suggested Criterion	Description	Metric	Benchmark
C1	Preferred Unambiguous Interoperability	Minimum Operations Are Available	Service providers SHOULD provide a common Application Programming Interface to interact with PIDs, supporting a minimum set of operations (create, resolve and modify PID and PID Kernel Information)	Σ T1, n	=0 ⟶ 0 >1 ⟶ 1

#	Test	Description	Type	Method	Guidance
T1,1	CREATE	Create a PID and provide kernel information: API exists and evidence (URL) is available	Binary	Yes = 1 No = 0	G1
T1,2	UPDATE	Update kernel information for existing PID: API exists and evidence (URL) is available	Binary	Yes = 1 No = 0	G1
T1,3	Resolution Service	Resolution API (URL) or URI Pattern exists, evidence is provided	Binary	Yes = 1 No = 0	G1
G1	One may extend the tests to recognise typical and popular standards for API implementation, such as REST, SmartAPI, and the like.				

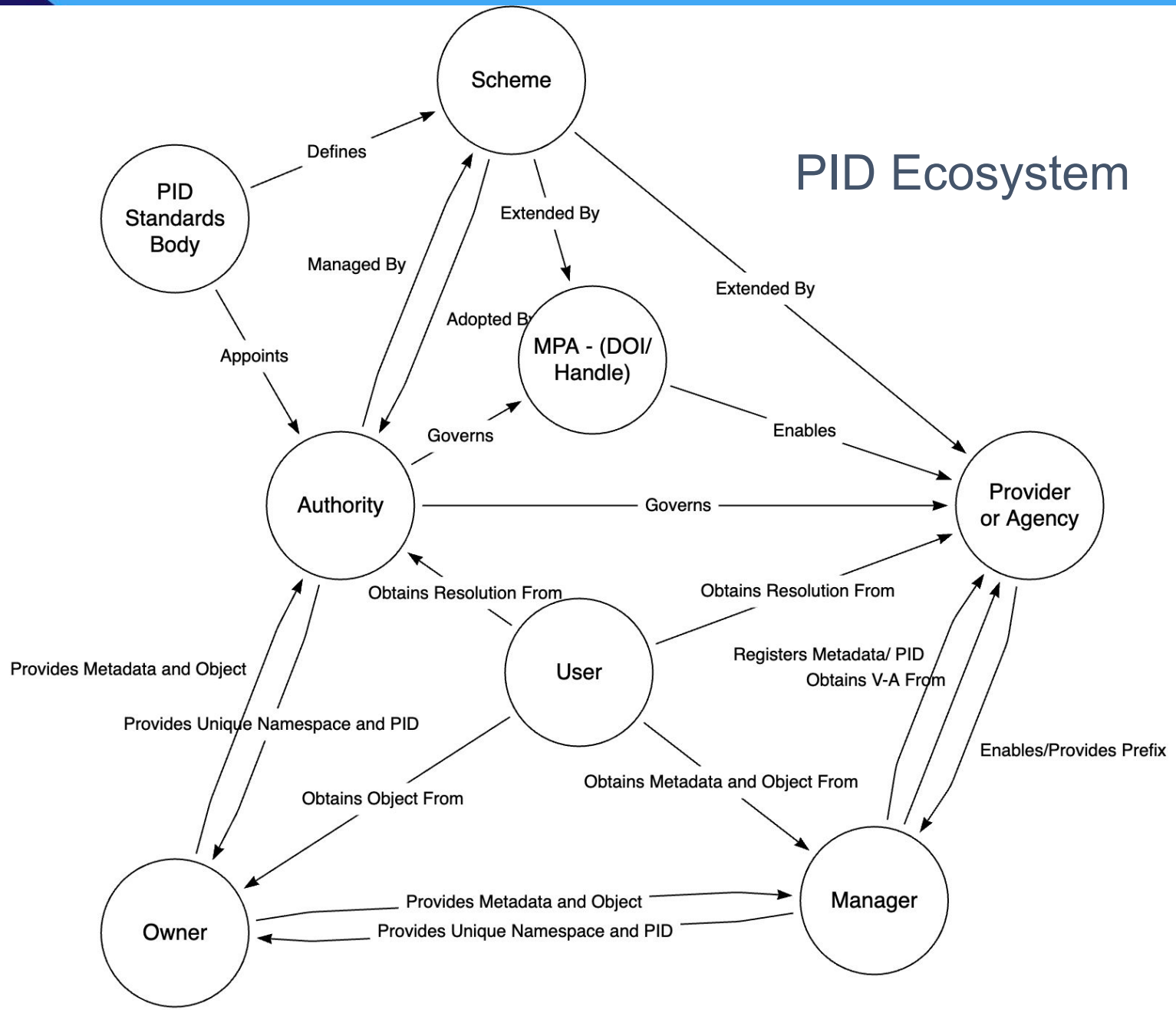
Resolvable Mutable

PID Stacks

PID Stacks

We have generalised the ecosystem that provides PID Stacks by looking at a number of PID families (Handle/ DOI, ARK, URNs of various kinds, ORCID, IGSN, and SWHID - and many others).

Note that an ***Owner***, in some cases, can obtain PIDs directly from an Authority (e.g. ARK, SWHID, PURL, …)



PID Stacks

PID Stacks are best described by looking at some examples. The table on the right shows a number of PID Stacks built from the Handle System. It includes popular stacks such as the **DataCIte** and **CrossRef** DOIs, as well as **ePIC**.

There are many other stacks in the ecosystem.

Scheme	Authority	MPA	Provider (Registration Agencies)
Handle System	DONA Foundation	Corporation for National Research Initiatives (CNRI)	Not investigated
Handle System	DONA Foundation	Coalition for Handle Services – China	Not investigated
Handle System	DONA Foundation	GDWG/ ePIC	See A.4.6
Handle System	DONA Foundation	CTIC	Not investigated
Handle System	DONA Foundation	MISADI	Not investigated
Handle System	DONA Foundation	Smart Africa Alliance	Not investigated
Handle System	DONA Foundation	Tunisian Internet Alliance	Not investigated
Handle System	DONA Foundation	RosTelecom	Not investigated
Handle System	DONA Foundation	International DOI Foundation	Airiti
Handle System	DONA Foundation	International DOI Foundation	BSI Identify
Handle System	DONA Foundation	International DOI Foundation	Chinese DOI
Handle System	DONA Foundation	International DOI Foundation	CNKI
Handle System	DONA Foundation	International DOI Foundation	CrossRef
Handle System	DONA Foundation	International DOI Foundation	DataCite
Handle System	DONA Foundation	International DOI Foundation	EIDR
Handle System	DONA Foundation	International DOI Foundation	HAND
Handle System	DONA Foundation	International DOI Foundation	JaLC
Handle System	DONA Foundation	International DOI Foundation	KISTI
Handle System	DONA Foundation	International DOI Foundation	mEDRA
Handle System	DONA Foundation	International DOI Foundation	EU-OP

CAT
EOSC Compliance Assessment Toolkit

eosc | FAIRCORE4EOSC

Enabling a FAIR EOSC ecosystem

Compliance Assessment Toolkit

EOSC PID Policy compliance for all actors in the ecosystem.

Workflows, Use Cases, Best Practices, Entities, …

(René van Horik, Wim Hugo, DANS)

The Value of EOSC PID Policy Compliance

For Authorities and Providers: It is important to back up public claims of EOSC Policy Compliance and of performance/ features with publicly available information

For Managers: Optional disclosure of policy compliance for purposes of trust and community acceptance

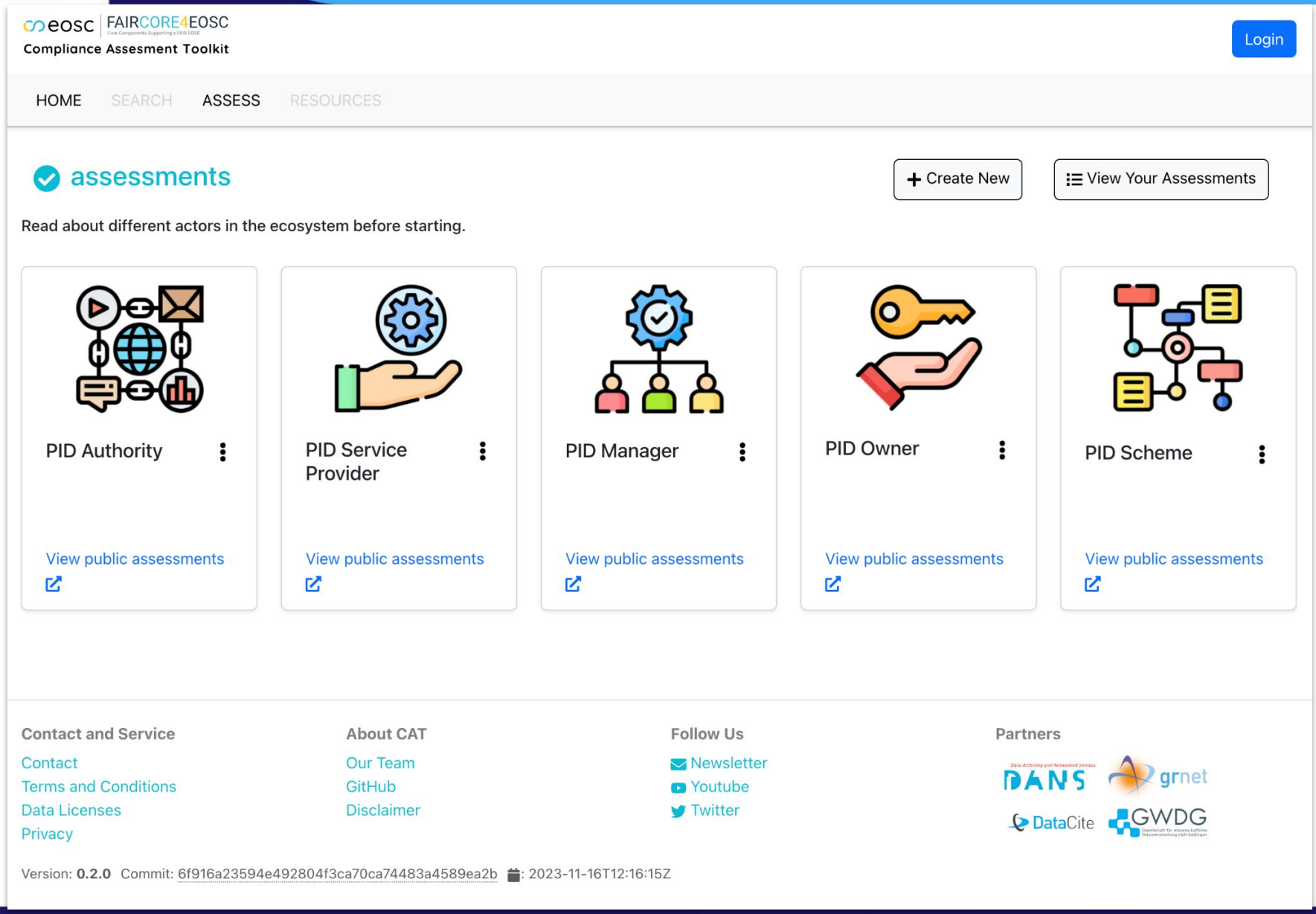
For Owners and Users: Matching PID services with use cases and expected benefits

All assessments are initially self-assessments - envisaged that we assist all interested PID Stacks with this. *The EOSC PID Policy envisages a compliance authority.*

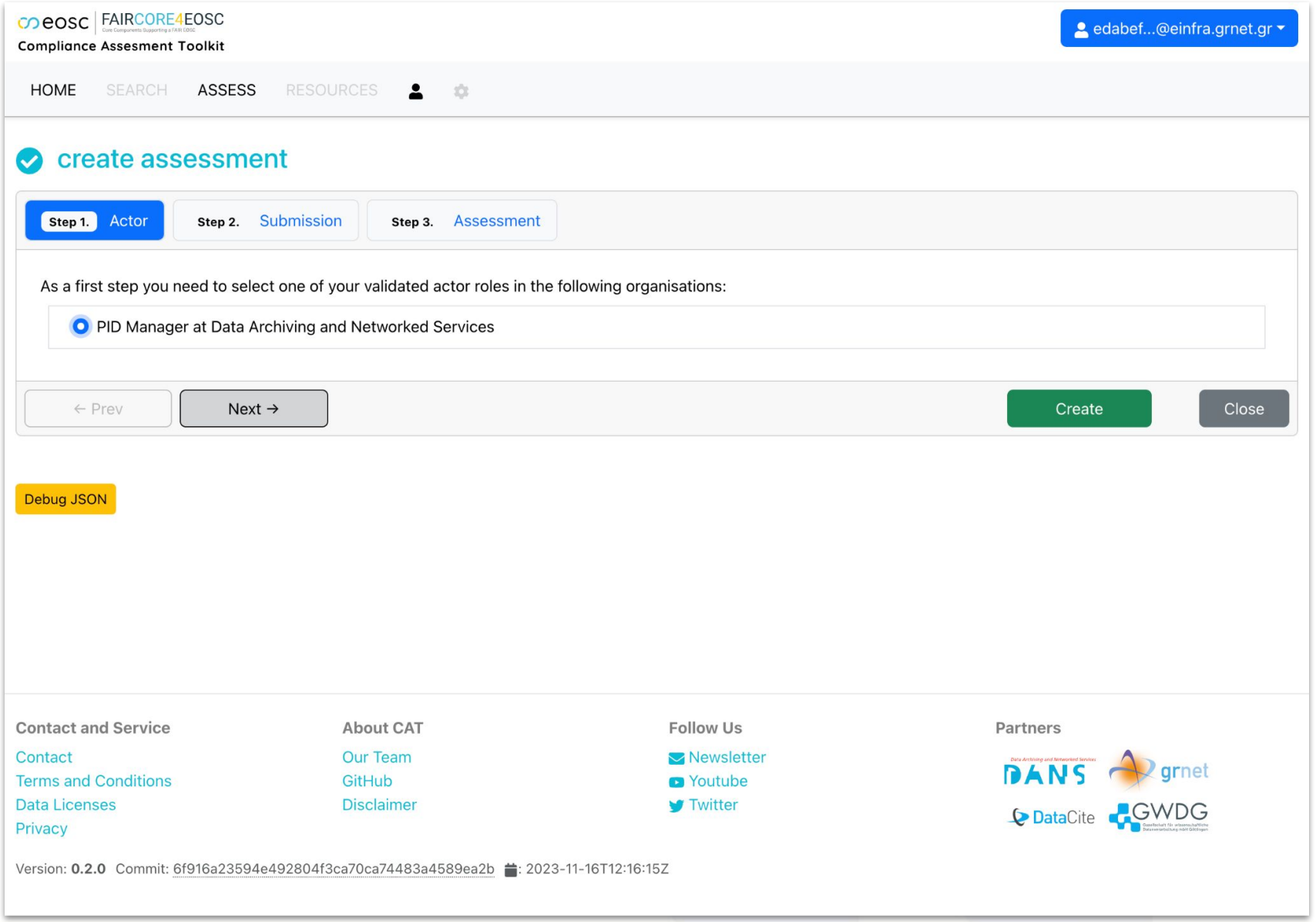
All assessments are initially self-assessments and ***there is no obligation to publish*** these assessments for the foreseeable future - publish if it is useful.

Find PID services and assess their characteristics, features, and how it matches use cases - whether they are policy compliant or not.

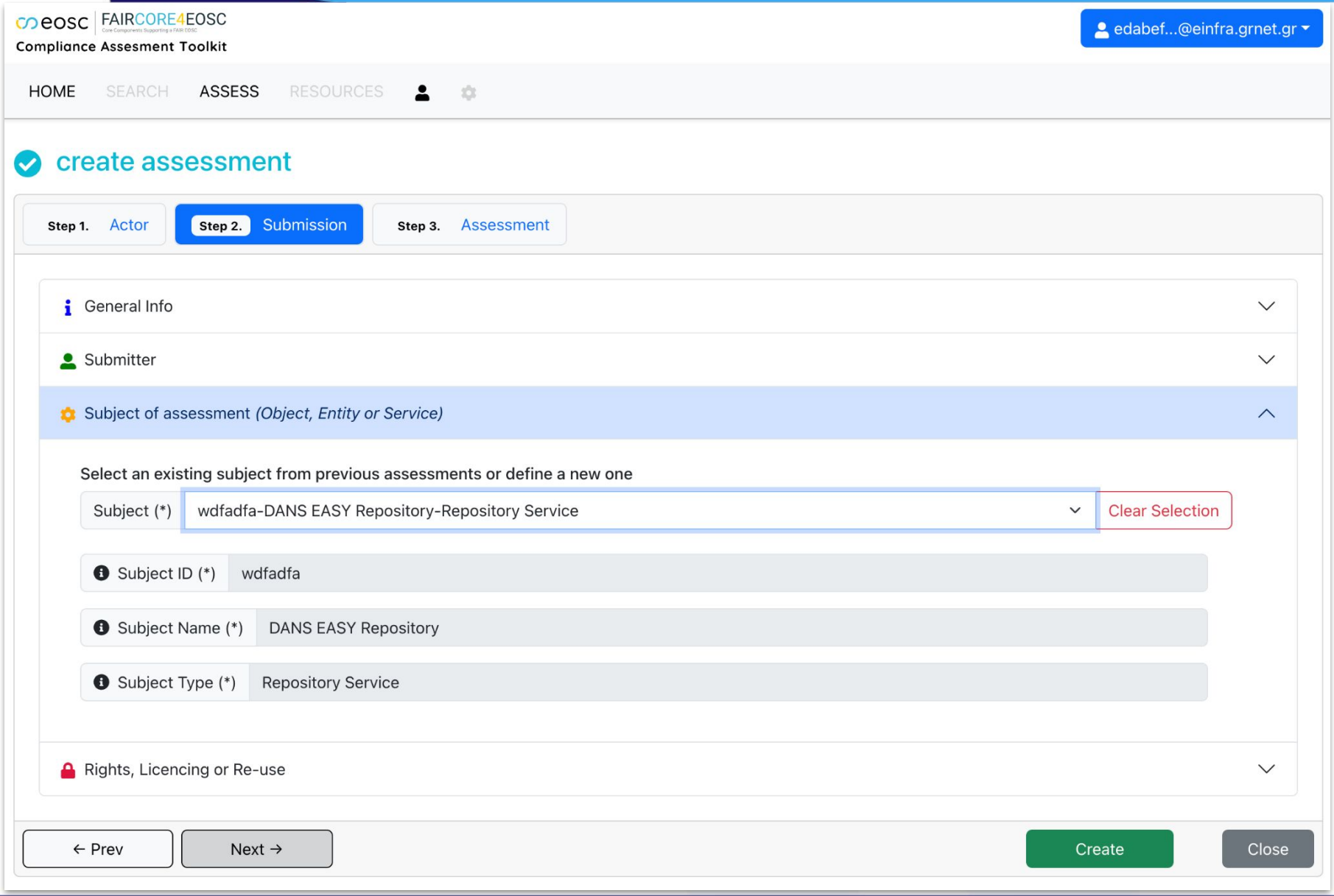
Assessments



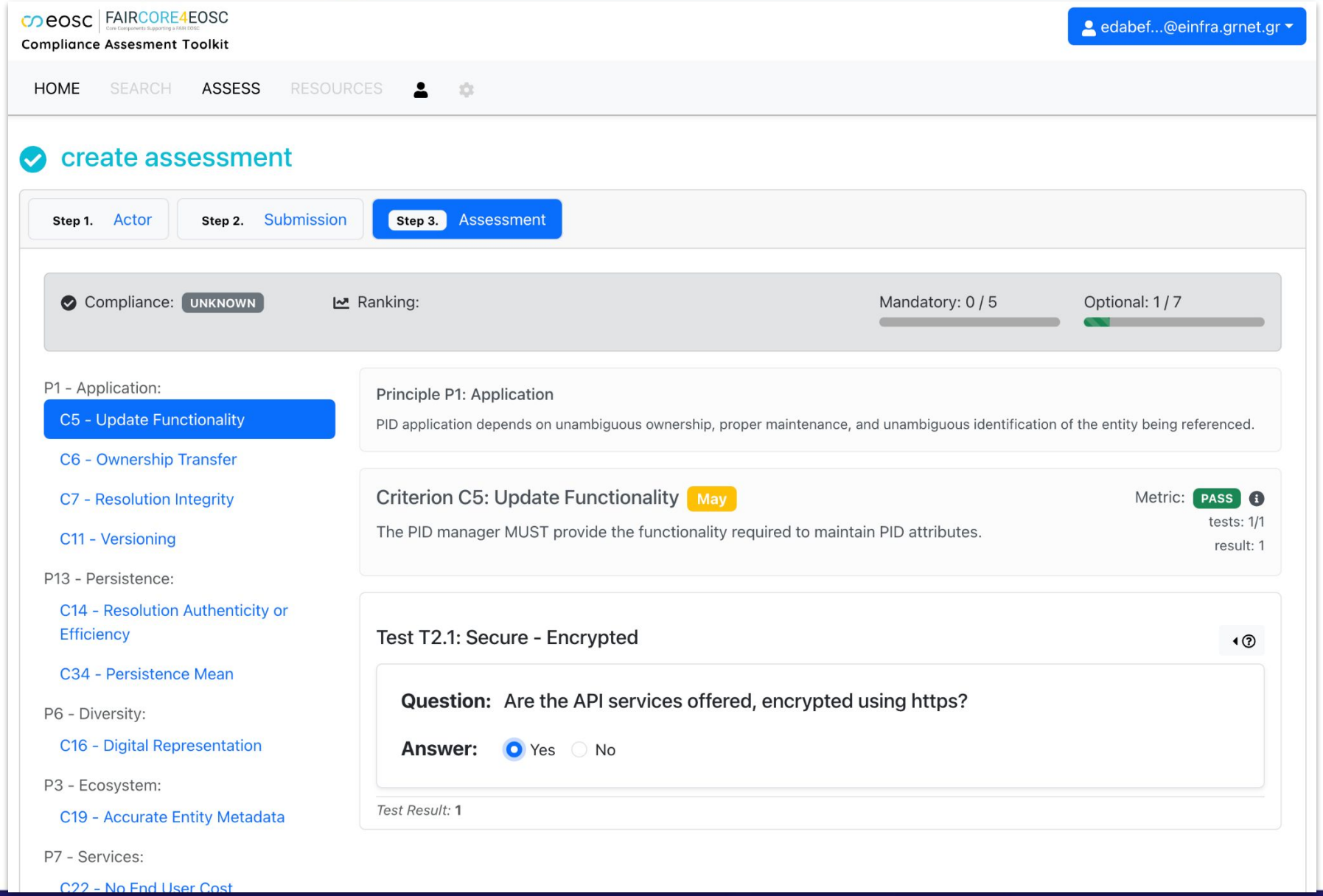
Create a New Assessment



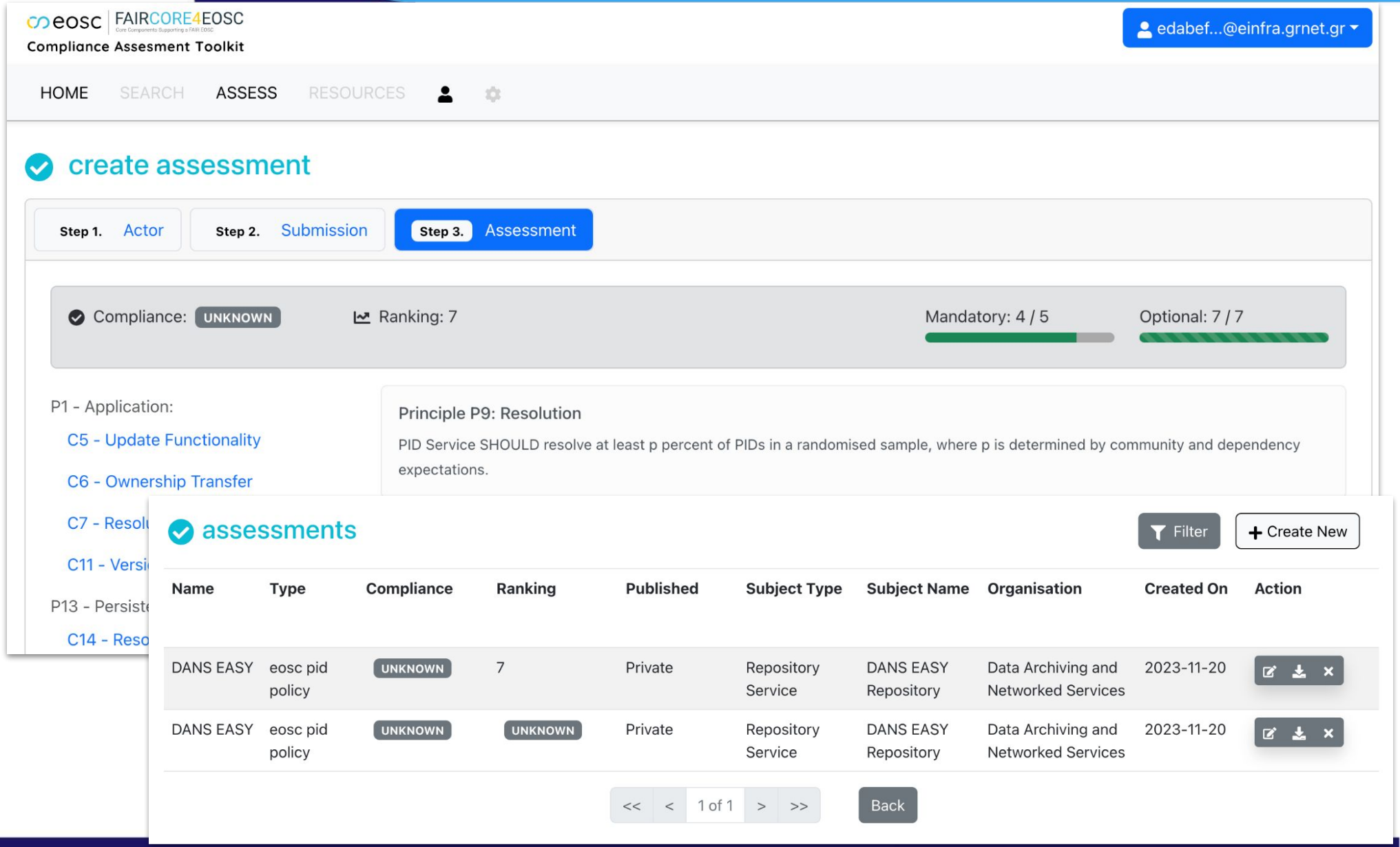
Object or Service Being Assessed



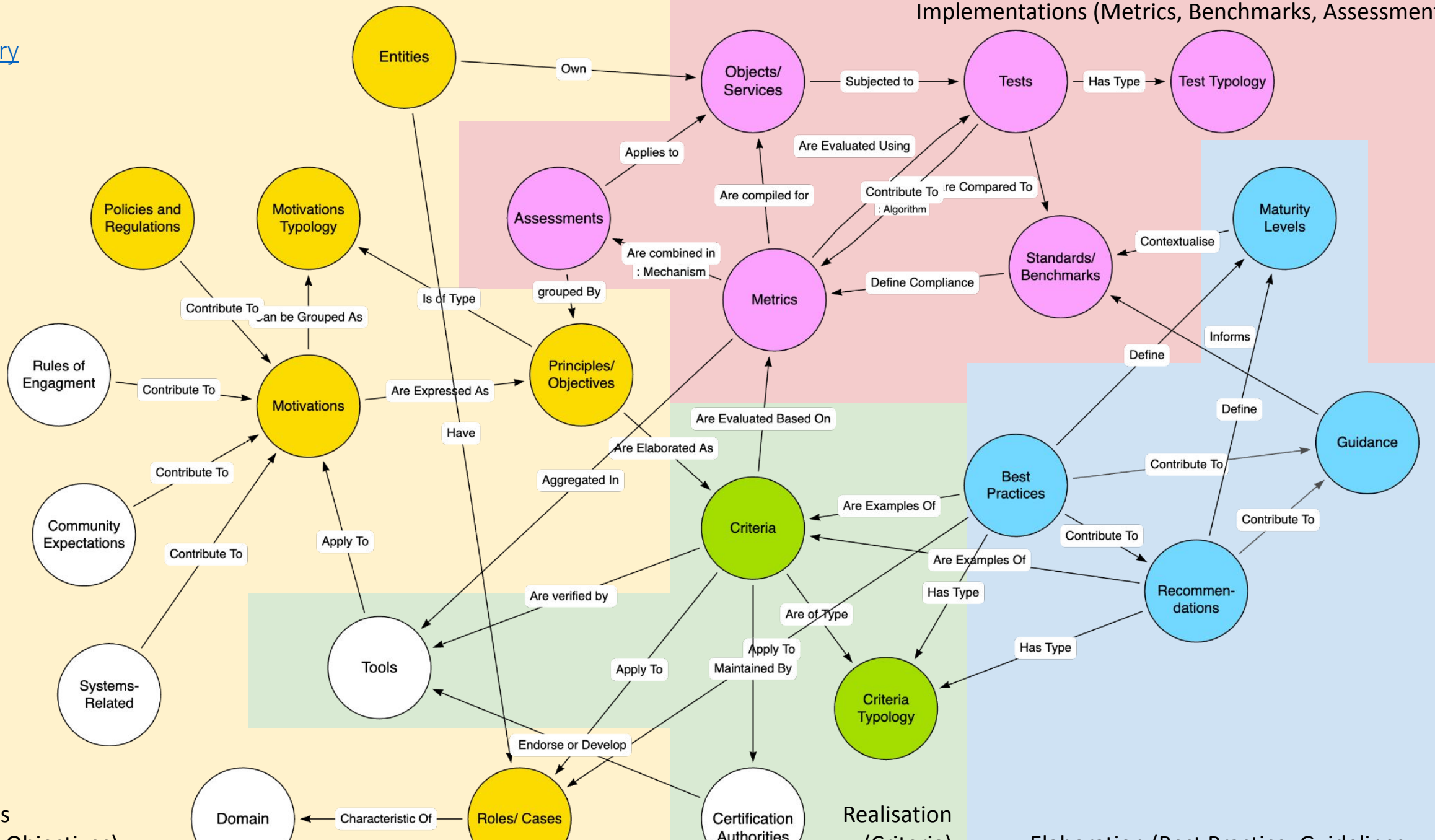
Assessment Criteria



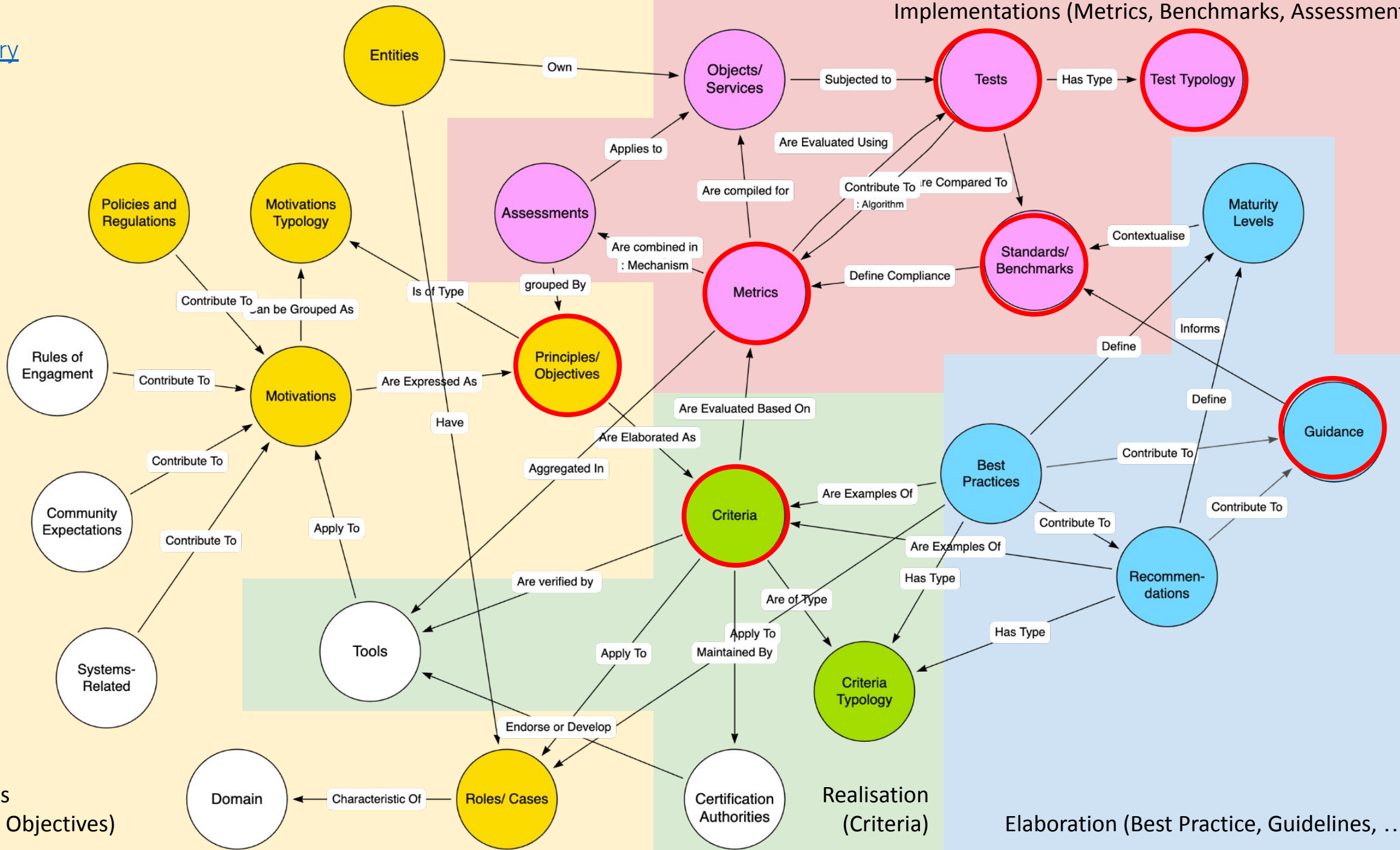
Mandatory and Optional Criteria



Full Story


Implementations (Metrics, Benchmarks, Assessment)
Entities
Own
Objects/ Services
Subjected to
Tests
Has Type
Test Typology
Applies to
Are Evaluated Using
Assessments
Are compiled for
Contribute To
: Algorithm
re Compared To
Policies and Regulations
Motivations Typology
Are combined in
: Mechanism
Metrics
Define Compliance
Standards/ Benchmarks
Contextualise
Maturity Levels
Contribute To
an be Grouped As
Is of Type
grouped By
Rules of Engagment
Contribute To
Motivations
Are Expressed As
Principles/ Objectives
Informs
Define
Have
Are Evaluated Based On
Are Elaborated As
Aggregated In
Guidance
Contribute To
Best Practices
Define
Contribute To
Community Expectations
Contribute To
Criteria
Are Examples Of
Contribute To
Contribute To
Apply To
Are Examples Of
Recommen-dations
Are verified by
Has Type
Are of Type
Systems-Related
Tools
Apply To
Maintained By
Apply To
Has Type
Criteria Typology
Endorse or Develop
Domain
Characteristic Of
Roles/ Cases
Certification Authorities
Motivations (Principles, Objectives)
Realisation (Criteria)
Elaboration (Best Practice, Guidelines, …)

Full Story


Implementations (Metrics, Benchmarks, Assessment)
Entities
Own
Objects/ Services
Subjected to
Tests
Has Type
Test Typology
Applies to
Are Evaluated Using
Assessments
Are compiled for
Contribute To
: Algorithm
re Compared To
Policies and Regulations
Motivations Typology
Are combined in
: Mechanism
Metrics
Standards/ Benchmarks
Contextualise
Maturity Levels
Define Compliance
Contribute To
an be Grouped As
Is of Type
grouped By
Rules of Engagment
Contribute To
Motivations
Are Expressed As
Principles/ Objectives
Define
Informs
Define
Guidance
Have
Are Evaluated Based On
Are Elaborated As
Aggregated In
Contribute To
Community Expectations
Contribute To
Apply To
Criteria
Are Examples Of
Best Practices
Contribute To
Contribute To
Contribute To
Are Examples Of
Recommen-dations
Are verified by
Has Type
Are of Type
Tools
Apply To
Maintained By
Apply To
Has Type
Systems-Related
Criteria Typology
Endorse or Develop
Domain
Characteristic Of
Roles/ Cases
Certification Authorities
Motivations (Principles, Objectives)
Realisation (Criteria)
Elaboration (Best Practice, Guidelines, …)

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

Outcomes from looking at PIDs in research workflows workshop

By: Natascha van Lieshout (SURF)

Funded by the European Union

EOSC Compliant PID Implementations - Practical Guidelines for Implementing Best Practices

09.00 - 13.00 CET
21 November 2023
Online

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

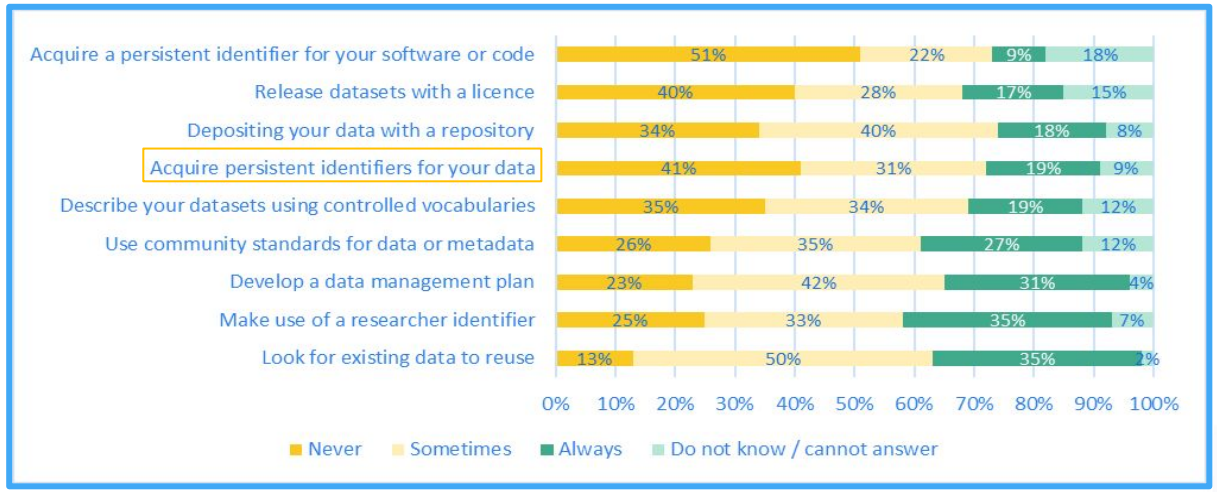
FAIR-IMPACT.eu

Funded by the European Union

Goal:

To demonstrate concrete tools, policies and developments within the realm of PIDs from the EOSC projects and use that to fuel further innovation through the FAIR IMPACT Open Support Call in January 2024.

Motivation



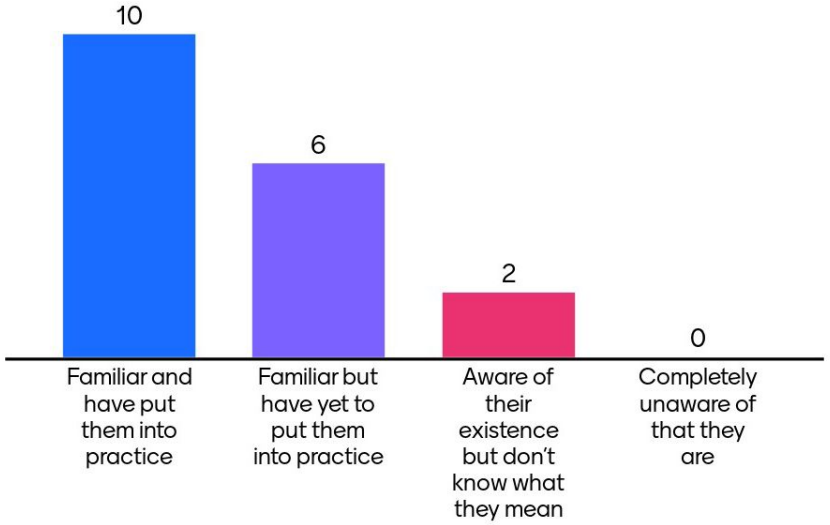
European Research Data Landscape Study

- Ran from 2021-2022
- 15,066 respondents

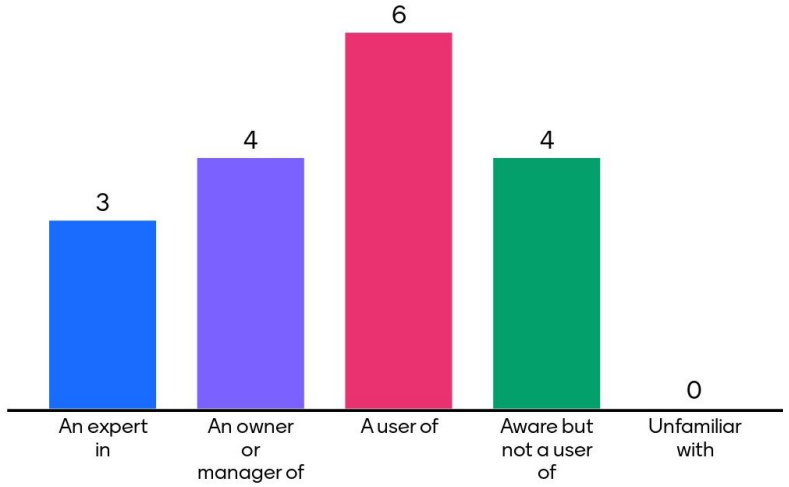
https://op.europa.eu/en/publication-detail/-/publication/03b5562d-6a35-11ed-b14f-01aa75ed71a1

https://zenodo.org/communities/erdl21

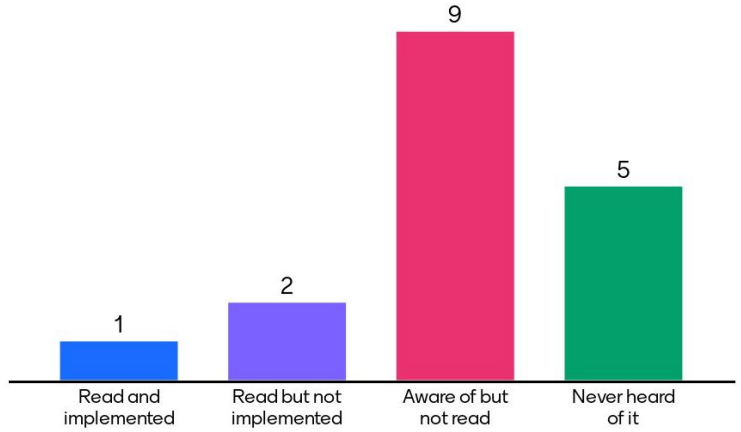
Familiarity with FAIR principles:



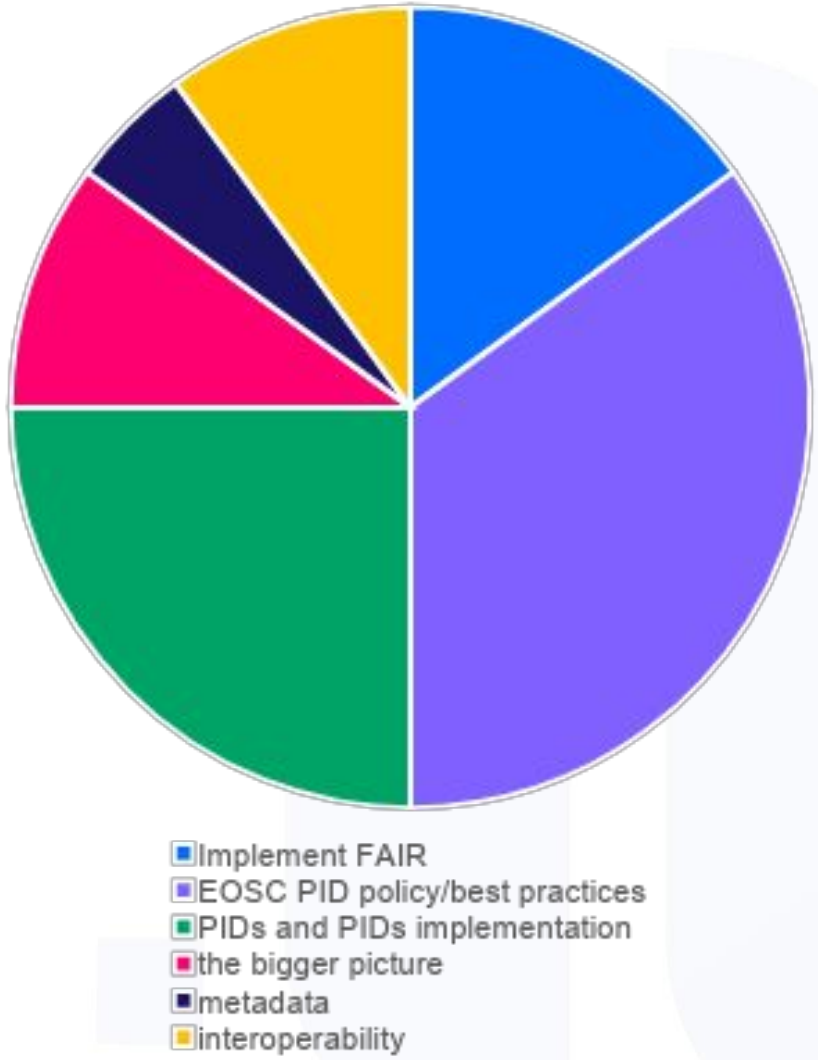
Familiarity with PIDs:



Familiarity with EOSC PID Policy:



I came to the EOSC Compliant PID Implementations workshop to learn about…



Open questions about PIDs:

Where to obtain or create them

What the open questions are in EOSC

why everyone thinks DOI equals PID...

Realtion between all PIds

not confused, but metadata (minimal) requirements are often not obvious

long term persistence guarantees

Uniqueness

the usage of PIDs is not obvious yet

How to use the same PID across very different domains

Who is the 'source' for governance

what the eosc pid policy means practically, especially if there is also a national pid policy in place.

Interoperability

Do we have a agreement about which pid to use for what (ORCID for authors, ROR for organisations, DOIs for paper, data and software, ..)?

why DOI is considered the only viable PID sometimes

Adressed

Developing

Breakout: Publications

- The perspective of owners and selection of appropriate PID services is also quite important but not really covered by EOSC PID Policy
- Reviewed considerations for owners to select an appropriate service
- PID Policy is ‘binary’ - there are some refinements necessary

Breakout: Software

- Intrinsic vs extrinsic PIDs
- A PID is an additional commitment and software licenses could potentially help define this commitment
- Human curation is necessary but we must standardize and track it through metadata
- Some things need to be left to specialists

Breakout: Datasets

- EOSC can define PID policies, best practices and guidelines and PID providers can enable compliance with these policies but this does not guarantee that end users will do so
- Interoperability between PIDs and other tools is crucial
- Data workflows and PID maintenance within those workflows remains complex

Thank you!

eosc | FAIR-IMPACT

@fairimpact_eu /company/fair-impact-eu-project

Funded by the European Union

Synchronisation Force

Use cases working on documenting best practices for PIDs

Synchronisation Force PID workshop
30 November, 2023

Josefine Nordling, CSC - IT Center for Science
WP lead on Persistent Identifiers

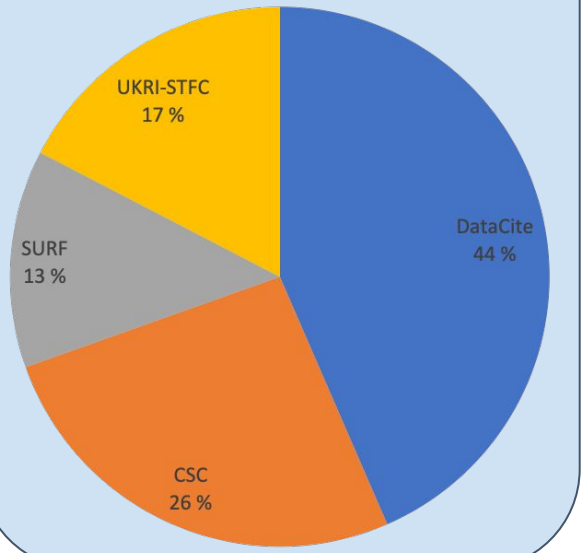
Funded by the European Union

Tasks on PIDs

Task 3.1

DataCite

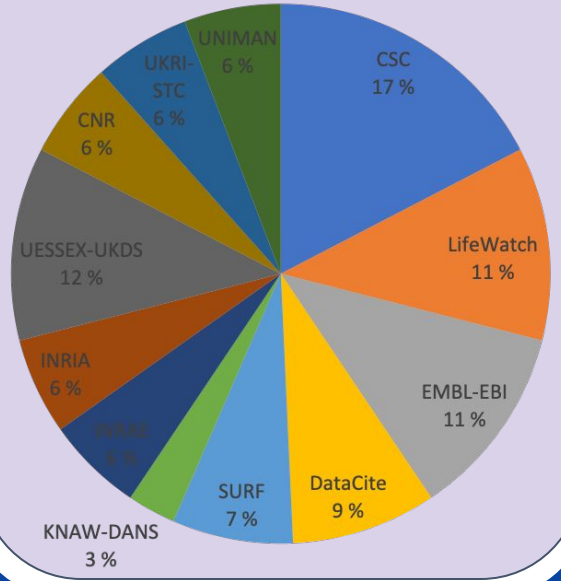
Setting up a coordination mechanism for EOSC PID service providers



Task 3.2

CSC

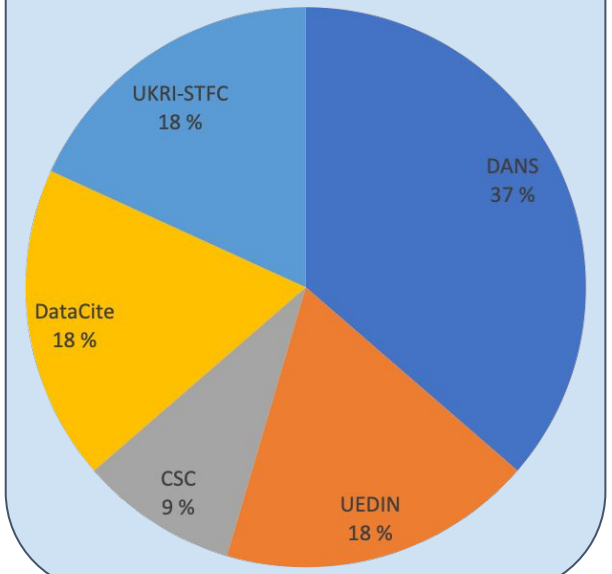
Integration of PID practices into FAIR data management



Task 3.3

DANS

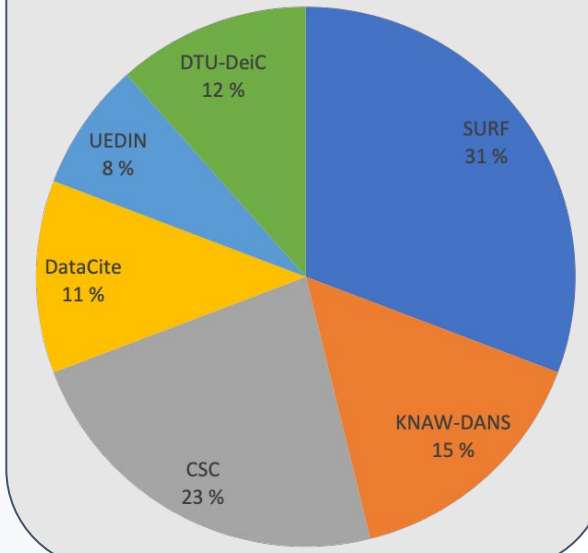
EOSC PID Policy alignment & support



Task 3.4

SURF

PID implementation programme



PID practices in FAIR data management

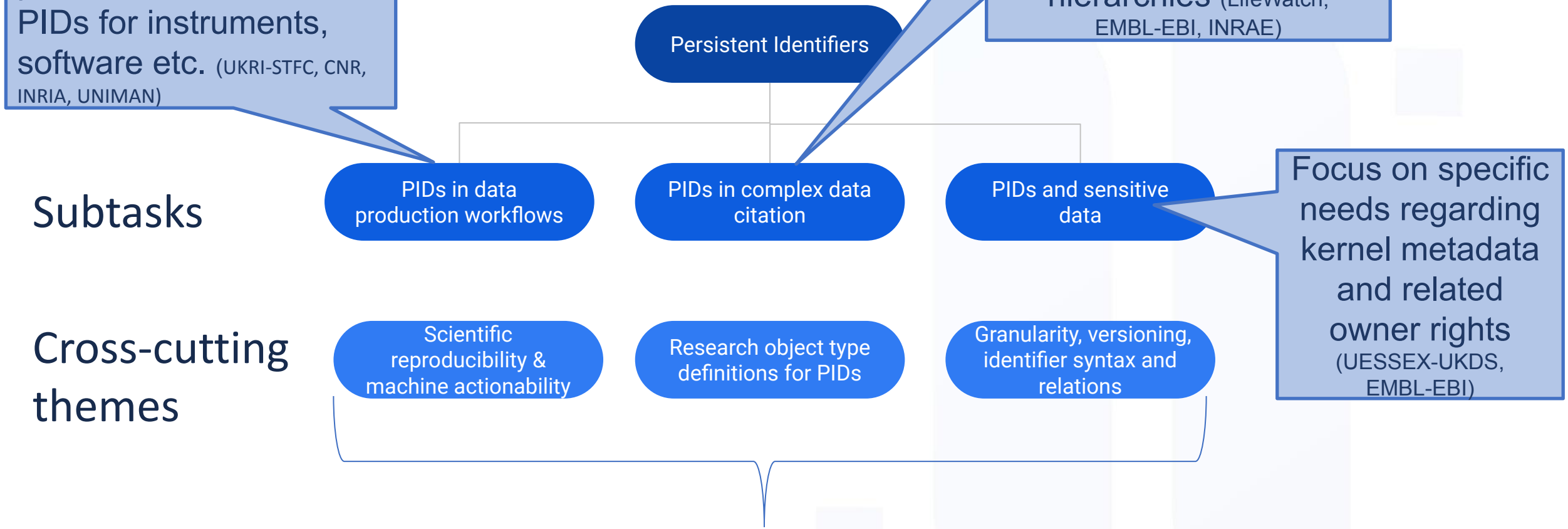
- Identifying and documenting best PID practices for; managing workflows, datasets and research objects
- Meeting user needs and a more coherent implementation of PIDs, leading to more exact data citation and a broader and more targeted use of PIDs
- Sharing the results across research communities to achieve alignment in PID practices
- Interacting with other communities through workshops

Source: Mindmeister.com

se case topics & structure

Focus on data products, automatic workflows, documenting data provenance in processes as well as PIDs for instruments, software etc. (UKRI-STFC, CNR, INRIA, UNIMAN)

Focus on PIDs for data citation with different types of entities, including versioning, collections and hierarchies (LifeWatch, EMBL-EBI, INRAE)



Focus on specific needs regarding kernel metadata and related owner rights (UESSEX-UKDS, EMBL-EBI)

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC

Knowledge Exchange report: PIDs & related risks and trust elements

Josefine Nordling, Senior Open Science Specialist
CSC – IT Center for Science
30 November, 2023

Funded by the European Union

Knowledge Exchange

Source: de Castro, P., Herb, U., Rothfritz, L., & Schöpfel, J. (2023). Building the plane as we fly it: the promise of Persistent Identifiers. Zenodo. https://doi.org/10.5281/zenodo.7258286

Risks related to PIDs

Risks related to PIDs - Political

Organisational change - Who is in control of the data? What about personal data transfer? - How to draw a consensus about the future of a given identifier? How to prevent increasing control from publishers? - If we want to build a reliable infrastructure, organisations need to figure out how to best involve all countries - Hard to argue for the importance of PIDs at a policy level.

misunderstood value
no community engagement
lack of funding
lack of commitment
non-conformity with gdpr
centralised solutions
lack of uptake
people dependencies
western-orientation
lack of strategic comms
sustainability
lack of support
discontinued services
poor scalability
commercial stakeholders
contingency funds
poor management
lack of human resources

Risks related to PIDs - Economic

No stable funding to ensure sustainability – the biggest risk! - Calculating the benefits for membership, difficult for smaller infrastructures and organisations to justify multiple memberships

misunderstood value
no community engagement
no interoperability
lack of funding
membership fees
lack of commitment
non-conformity with gdpr
centralised solutions
lack of uptake
people dependencies
western-orientation
lack of strategic comms
sustainability
lack of support
discontinued services
poor scalability
commercial stakeholders
contingency funds
poor management
lack of human resources

Risks related to PIDs - Socia

Awareness-raising crucial for commitment and showcasing value for supporting PID usage and uptake – People infrastructure - community agreement – Centralisation of PID infrastructure and people – Unrealiable data, lack of control and curation

Risks related to PIDs - Technologic

Risks concerning the metadata associated with a PID, such as quality, richness, completeness and risks around PID systems, especially their interoperability and scalability.

Figure 3. Framework of the analysed risk and trust variables

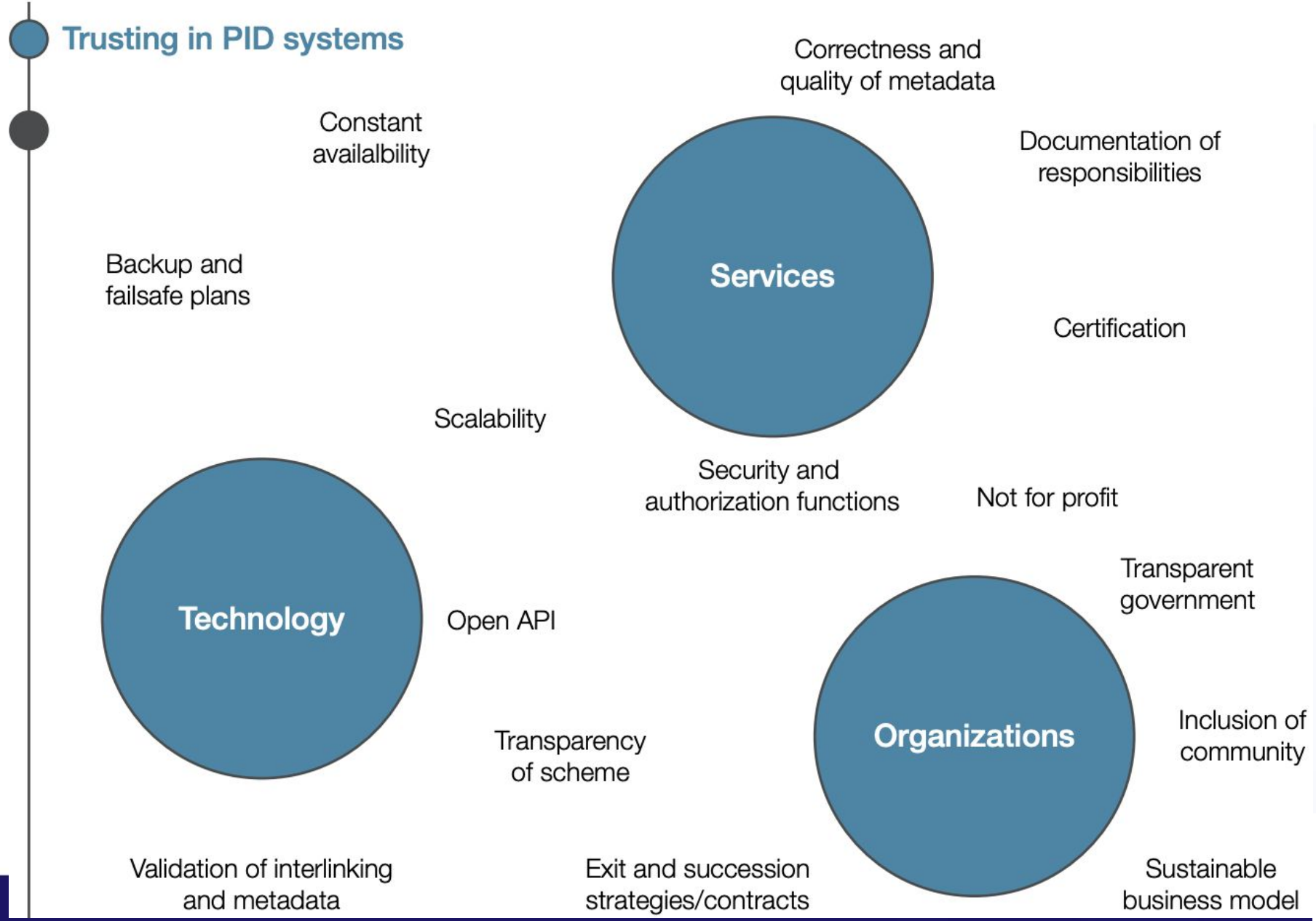
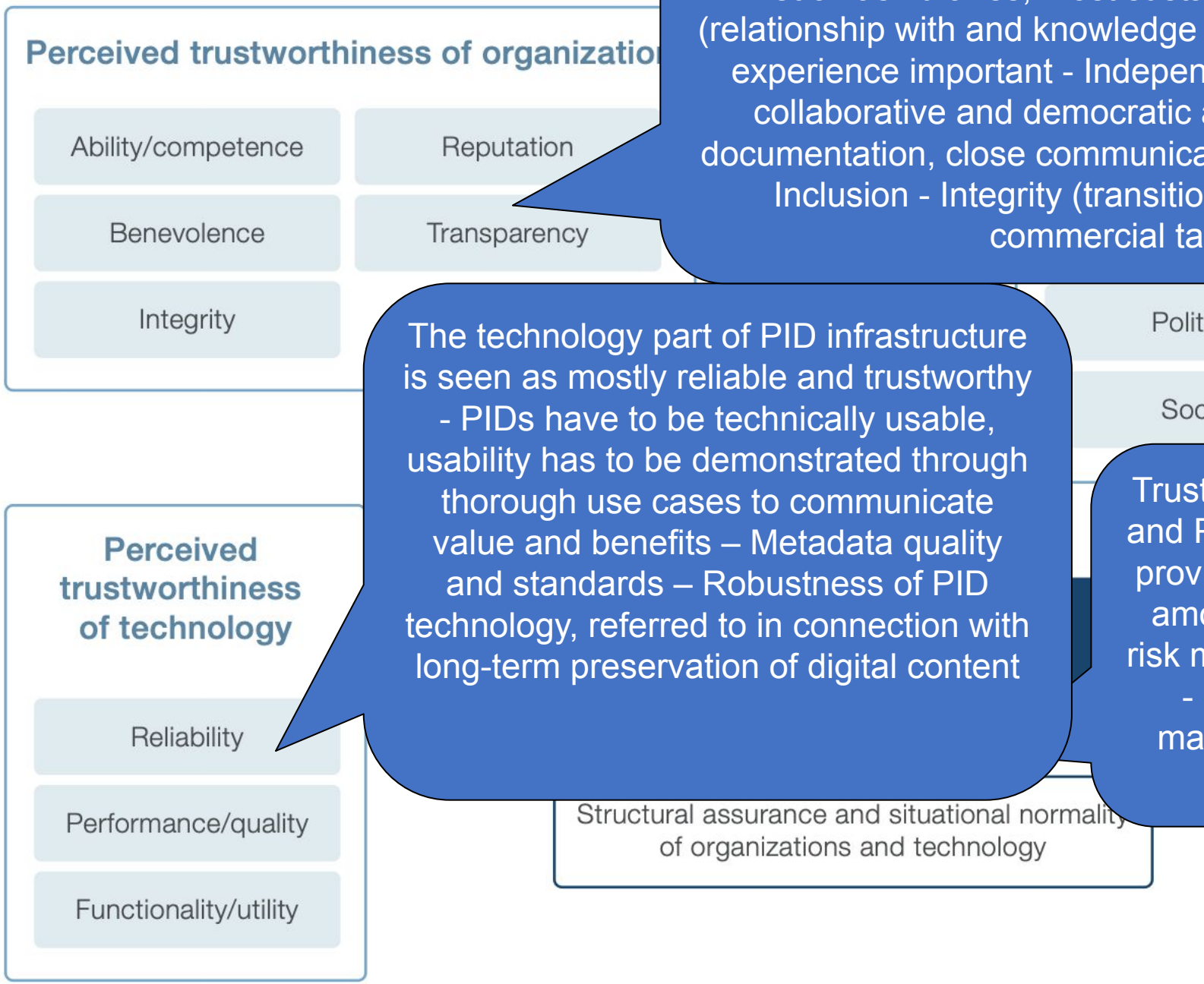


Figure 4. Framework of the analysed risk vari



State-funded, more traditional information management organisations, such as libraries, most sustainably reliable - Long existence (relationship with and knowledge of their users) - Agility and marketing experience important - Independent organisations (user-oriented, collaborative and democratic approach) – Transparency (open documentation, close communication, democratic decision-making) – Inclusion - Integrity (transition plans and protections against commercial takeovers in place)

The technology part of PID infrastructure is seen as mostly reliable and trustworthy - PIDs have to be technically usable, usability has to be demonstrated through thorough use cases to communicate value and benefits – Metadata quality and standards – Robustness of PID technology, referred to in connection with long-term preservation of digital content

Trusting relationships between PID users and PID providers - General trust into PID providers, uncommon with fallback plans among PID users - Contracts, policies, risk management workflows are important - Contingency plans needed - Risk management plans increase reliability

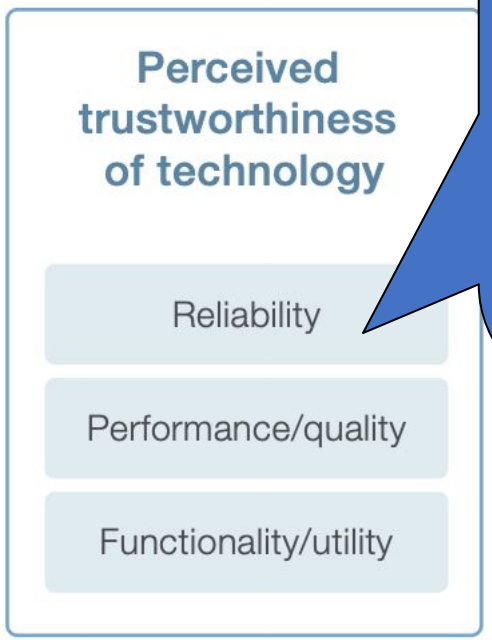
Knowledge Exchange

Figure 4. Framework of the analysed risk vari

State-funded, more traditional information management organisations, such as libraries, most sustainably reliable - Long existence (relationship with and knowledge of their users) - Agility and marketing experience important - Independent organisations (user-oriented, collaborative and democratic approach) – Transparency (open documentation, close communication, democratic decision-making) – Inclusion - Integrity (transition plans and protections against commercial takeovers in place)

Overall, there were very few doubts about the competence and ability of PID providers
Early and ideally favourable experiences with the PID are very important to establish trust!

usability has to be demonstrated through thorough use cases to communicate value and benefits – Metadata quality and standards – Robustness of PID technology, referred to in connection with long-term preservation of digital content



Structural assurance and situational normality of organizations and technology

Trusting relationships between PID users and PID providers - General trust into PID providers, uncommon with fallback plans among PID users - Contracts, policies, risk management workflows are important - Contingency plans needed - Risk management plans increase reliability

Knowledge Exchange

Thank you!

eosc | FAIR-IMPACT

@fairimpact_eu /company/fair-impact-eu-project

Funded by the European Union

Synchronisation Force