The Importance of FAIR Interoperability from a RI Perspective

Keith G Jeffery

Scenario

- A researcher wishes to understand the relationships between a volcanic eruption and its effect on human and animal life, health, water supply, power supply, transport, agriculture….
- This requires interoperation across geosciences, environmental sciences, health sciences, civil sciences, agricultural science…..
- But the relevant assets are hard to **F**ind, **A**ccess, **I**nteroperate and **R**e-use
  - Location, description, rights, formats, language…….

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

FAIR Principles

FORCE11
The Future of Research Communications and e-Scholarship

https://www.force11.org/group/fairgroup/fairprinciples

- **TO BE FINDABLE:**
- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
  F2. data are described with rich metadata.
  F3. (meta)data are registered or indexed in a searchable resource.
  F4. metadata specify the data identifier.
- **TO BE ACCESSIBLE:**
- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
  A1.1 the protocol is open, free, and universally implementable.
  A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
  A2 metadata are accessible, even when the data are no longer available.
- **TO BE INTEROPERABLE:**
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
  I2. (meta)data use vocabularies that follow FAIR principles.
  I3. (meta)data include qualified references to other (meta)data.
- **TO BE RE-USABLE:**
- R1. meta(data) have a plurality of accurate and relevant attributes.
  R1.1. (meta)data are released with a clear and accessible data usage license.
  R1.2. (meta)data are associated with their provenance.
  R1.3. (meta)data meet domain-relevant community standards.

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Implications: Findability

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
  F2. data are described with rich metadata.
  F3. (meta)data are registered or indexed in a searchable resource.
  F4. metadata specify the data identifier.

- F1: Universally Unique Resolvable Persistent Identifier
- F2: multiple elements to support Findability – note elements may have structure of attributes
- F3: use of a repository – ideally with CoreTrustSeal
- F4: one element in the metadata points to the data by use of an identifier

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Implications: Accessibility

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2 metadata are accessible, even when the data are no longer available.

- A1: either directly using an ID which is also an address (deprecated) or by an ID which is resolvable to an address
- A1.1: in essence based on W3C standards
- A1.2: this is much more difficult and involves tokens – GEANT trying to standardise
- A2: so-called 'tombstone' metadata – necessary for provenance

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Implications: Interoperability

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles.

I3. (meta)data include qualified references to other (meta)data.

- I1: implies a logic language – e.g. RDF triples or relational n-tuples that can handle deduction and induction
- I2: values of attributes in metadata elements refer to vocabularies (ideally ontologies) that are F,A,I,R
- I3: if a metadata record refers to another digital object the reference (i.e. the link between the two objects with addresses) should be qualified (e.g. with role, temporal duration, authentication based on licensing…)

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Implications: Reusability

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage license.

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

- R1: relates to F2
- R1.1: reference to licence – but problem is how to encode the licence as authorisation properties that can be processed
- R1.2: provenance may be inbuilt in the qualified references (I3) or refer to external provenance system (e.g. PROV)
- R1.3: metadata standard is used commonly

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Importance of Metadata for FAIR

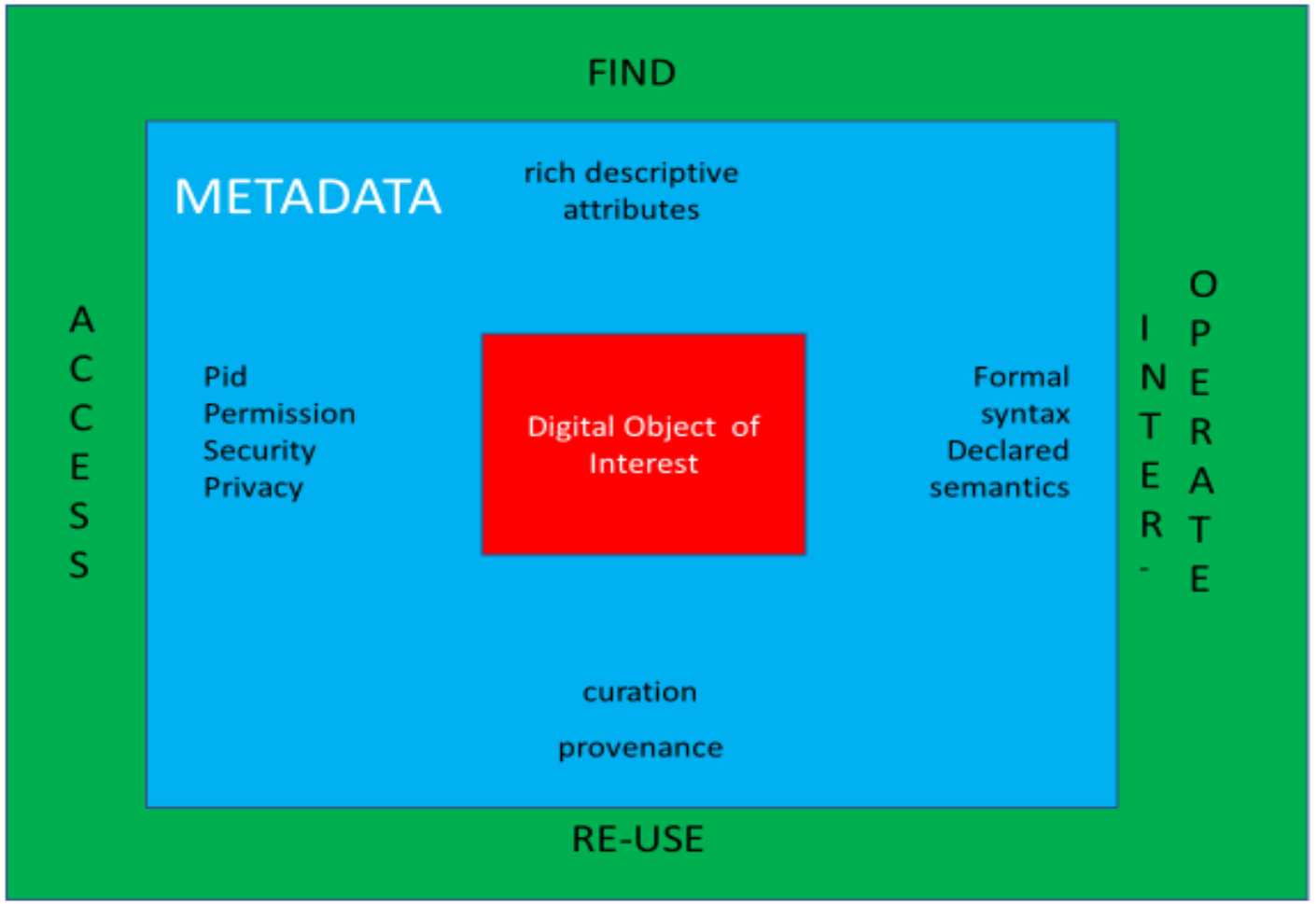


Figure 1: Metadata and FAIR

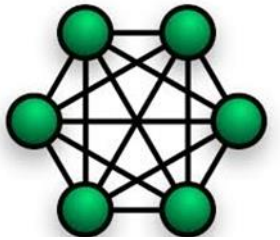
THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

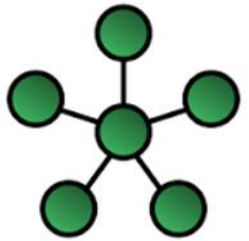
FAIR is not easy

Acknowledgement Wikipedia

- Asset descriptions in many formats, languages
- Same asset described differently multiple times
- Many assets not described adequately (or at all)
- Asset descriptions with different and more-or-less formal syntax
- Asset descriptions with different and more-or-less formal semantics

- 2 approaches:
- Broker between any pair of asset metadata descriptions
  - means n*(n-1) brokers
- A canonical rich metadata format and convert to it from each asset description
  - means n brokers

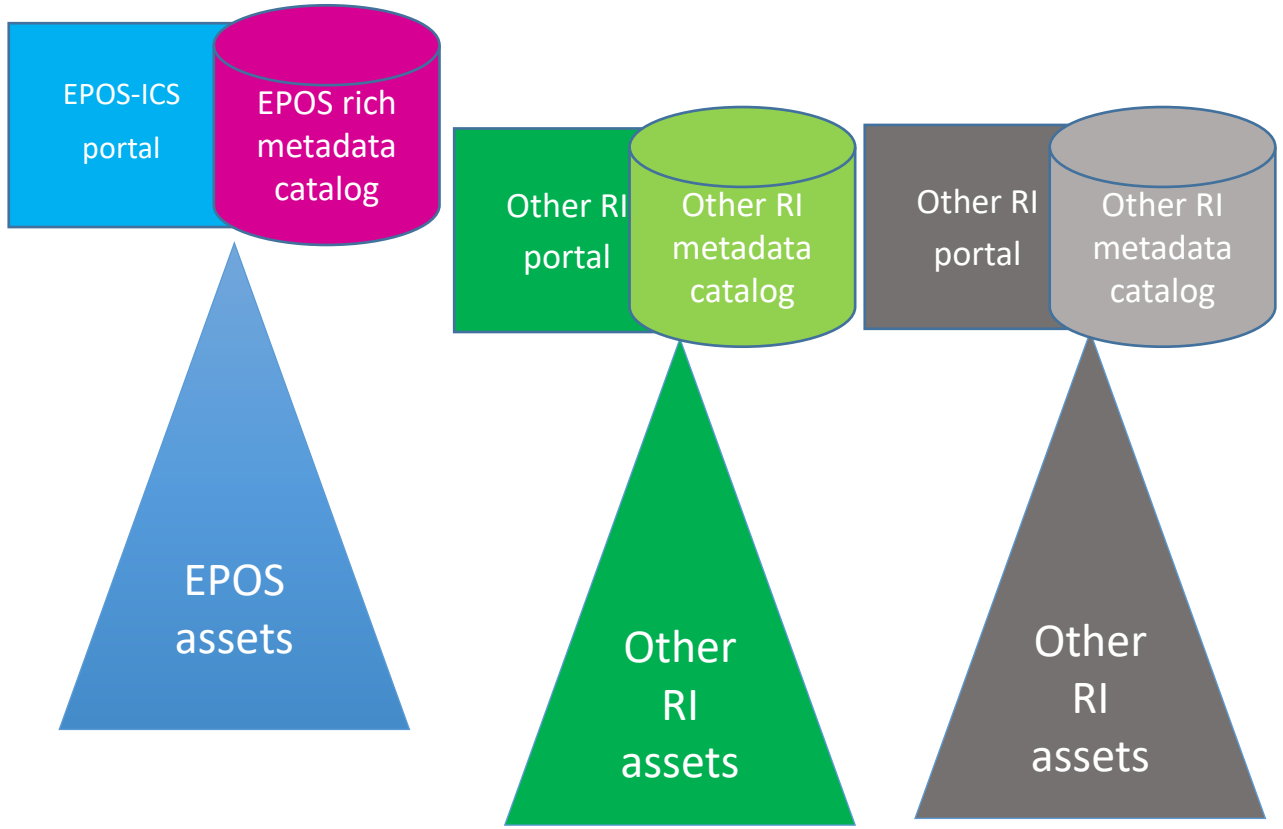




THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

The Requirement from a RI point of view

- A user working in the community of one RI (e.g. EPOS)
- Not only requires EPOS assets to be FAIR
  - for interoperation over heterogeneity within the RI
- But wishes also to (re-)use assets of one or more other RIs
  - Incorporating in her workflow
- To accomplish a research objective
- Requires FAIR metadata catalogs providing FAIR utilisation of RI assets



THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

The catalogs must have

- Compliance with FAIR principles
  - Work on how to achieve FAIRness from GO-FAIR, FAIR'sFAIR … and domain-specific projects e.g. FAIR4Health, ENVRIFAIR ….
  - Indicators of FAIRness developed by RDA FAIR Data Maturity WG
- Sufficient information for intended use
  - Discovery, contextualisation (relevance, quality, permissions), action
- Formal syntax and declared semantics
  - for autonomic processing
- Referential and functional integrity
  - For reliable processing

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

The catalogs must have

- Compliance with FAIR principles
  - Work on how to achieve FAIRness from GO-FAIR, FAIR'sFAIR … and domain-specific projects e.g. FAIR4Health, ENVRIFAIR ….
  - Indicators of FAIRness developed by RDA FAIR Data Maturity WG
- Sufficient information for intended use
  - Discovery, contextualisation (relevance, quality, permissions), action
- Formal syntax and declared semantics
  - for autonomic processing
- Referential and functional integrity
  - For reliable processing

And - of course - the EPOS catalog has these characteristics

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

FAIR and SERVICES

- FAIR originally designed for DATA not SERVICES
  - Problem: Implicit download
    - compare taking out a library book
    - Increasingly impractical (size, network latency)
  - Problem: processing associated with the data
    - Locality of data relative to computing resources, processing software and user
    - Resources required (computing, sensor networks, lab equipment)
    - Legalistics (permissions, security, privacy, liability)
- FAIR for SERVICES
  - If FAIR is applied to services which provide the data
    - Services intrinsically offer resources required
    - Services may be moved to the data (lower network latency)
    - Services may reduce the data (data management or analytics) for network transport
  - But…Metadata for services different from metadata for data

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

FAIR research workflows

- This is the real challenge
- Making research workflows – the research process – FAIR end-to-end
  - Initiation: the hypothesis

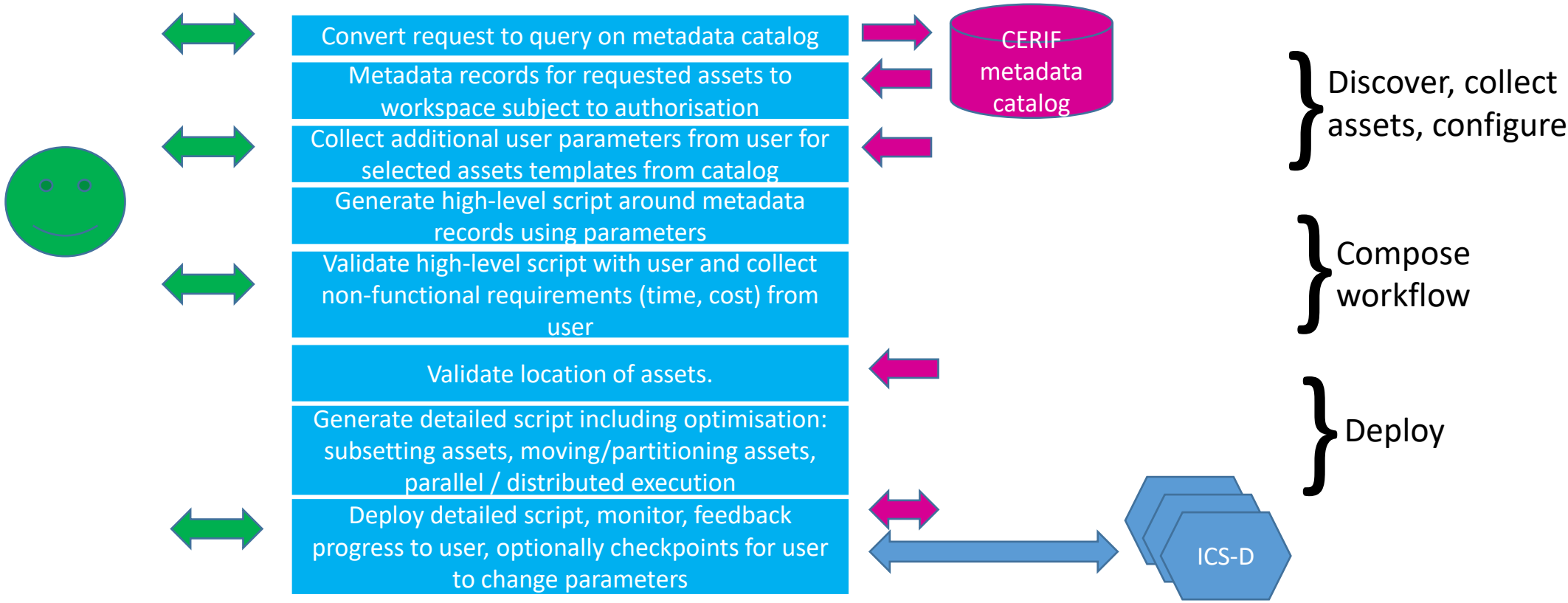
EPOS ICS-C	• Method: how to prove/disprove the hypothesis (observation / experiment / simulation) • Results (raw, calibrated, processed)
EPOS ICS-D	• Analysis (analytics, mining, simulation) • Visualisation
EPOS ICS-C	• Publication (finalised text, datasets, processes) • Reuse for reproducibility or re-purposing

Incremental accretion of rich FAIR metadata

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

EPOS Workflow Design

EPOS ICS-C



Not yet implemented fully: work in progress

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564

Thank You

WebSite

www.epos-eu.org

Newsletter

www.epos-eu.org/newsletter

Social

THIS PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME UNDER GRANT AGREEMENT N° 676564