

euroCRIS
Workshop
for
CRIS
Developers
and
Managers

Topics

- Research Information Interoperability
 - OpenAIRE Guidelines for CRIS Managers
 - Experience from Implementors
 - Processing in the OpenAIRE Validator & Aggregator

Why?

- Information **producers**:
What is going to happen with our metadata?
- Information **consumers**:
What metadata is (going to be) available?
Where does it come from?
- **euroCRIS**:
What more should to be done to make it maximum useful?
What are the most painful points of research information interoperability?
What can be generalized?

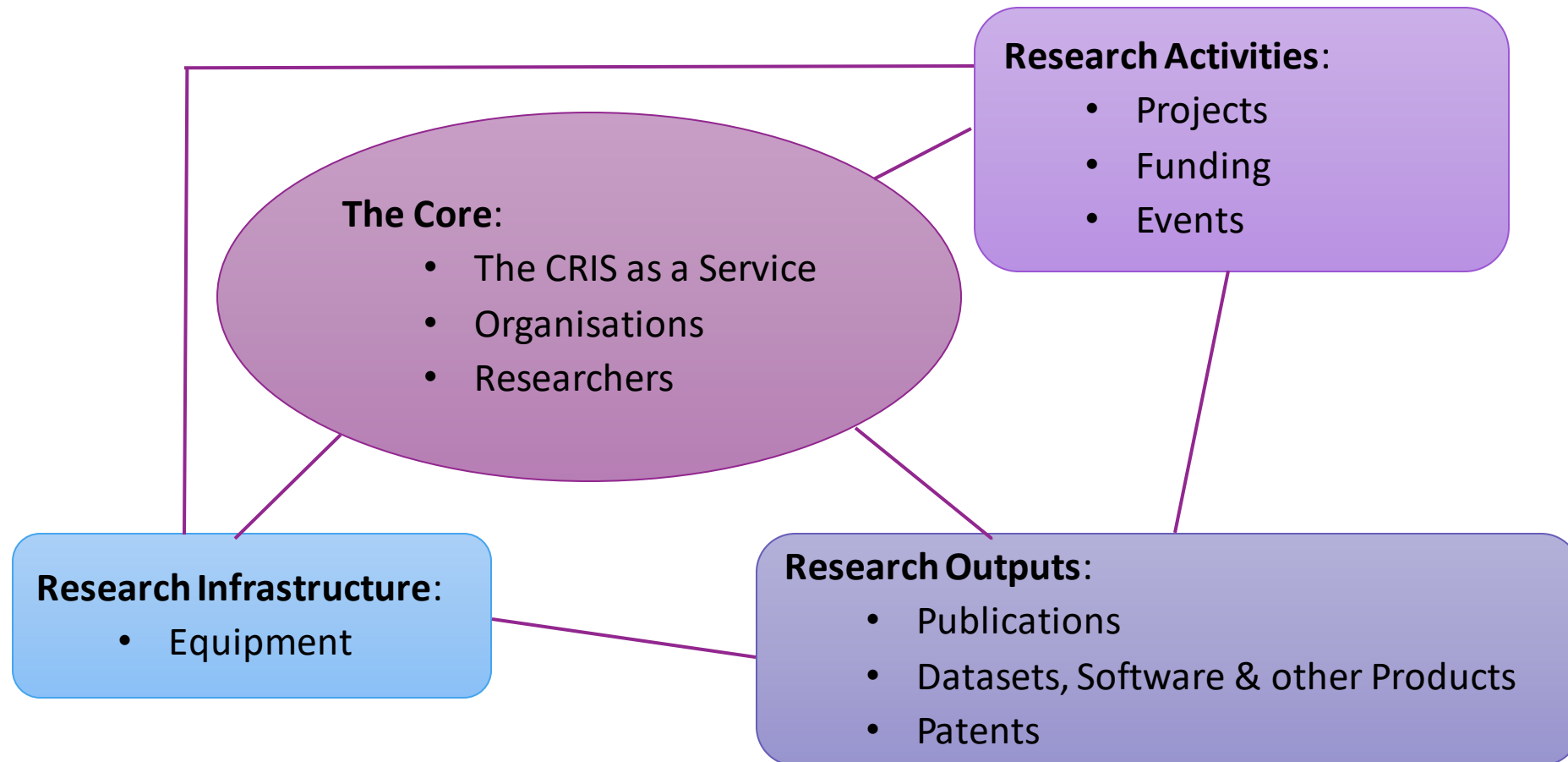
OpenAIRE Guidelines for CRIS Managers

An overview

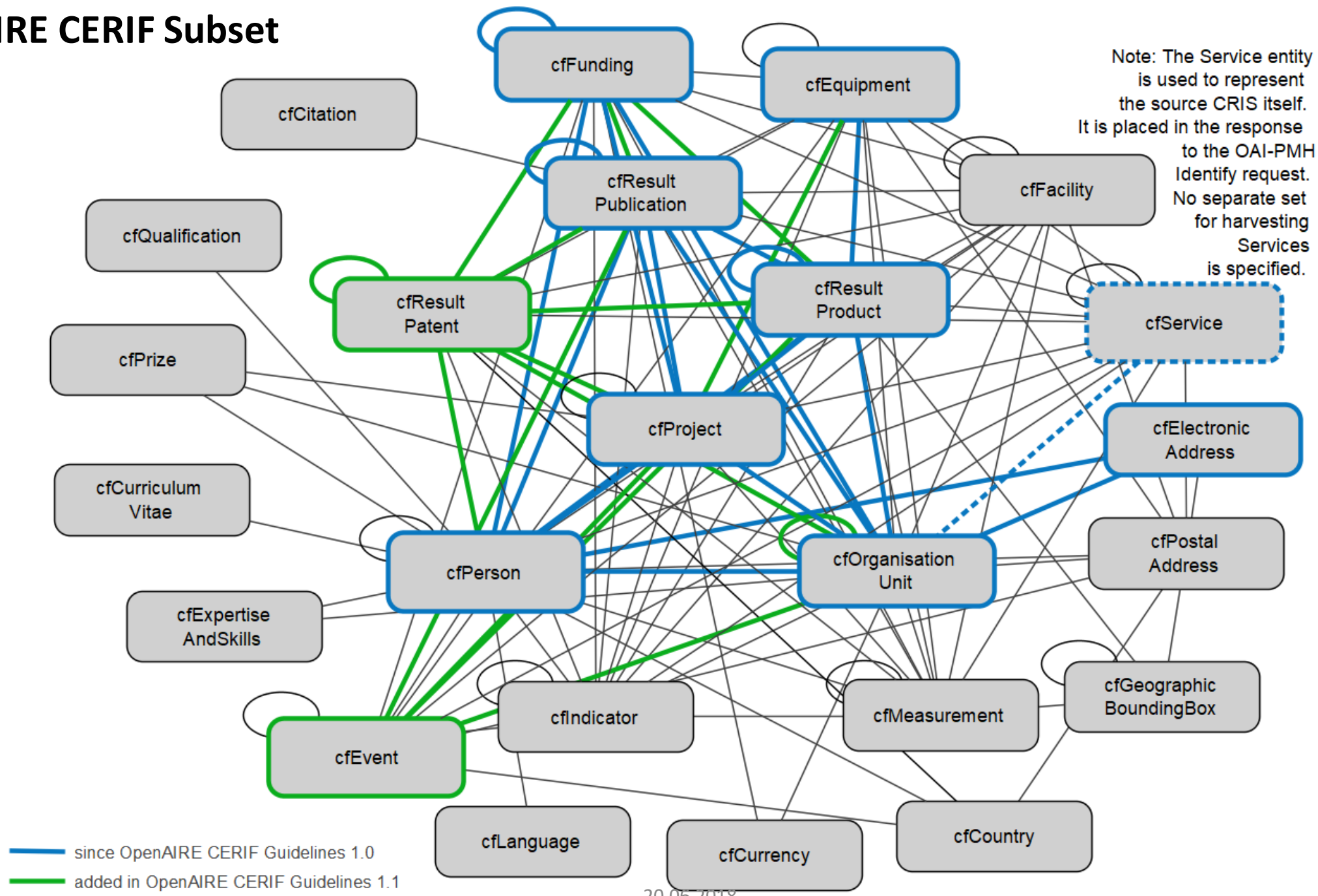
Evolution of the Standard

- **v.1.1.1** (December 2018), <https://doi.org/10.5281/zenodo.2316420>
 - Minor extensions
 - E.g. added DAI as person identifier
 - Documentation improvements
- **v.1.1.0** (June 2018), <https://doi.org/10.5281/zenodo.1298650>
 - Updated CERIF-XML
 - Vocabularies aligned with the other OpenAIRE Guidelines
- v.1.0 (June 2015), <https://doi.org/10.5281/zenodo.17065>
 - Old-style CERIF-XML, a 1:1 mapping from the CERIF data model

Information Scope of the OpenAIRE Guidelines for CRIS Managers 1.1

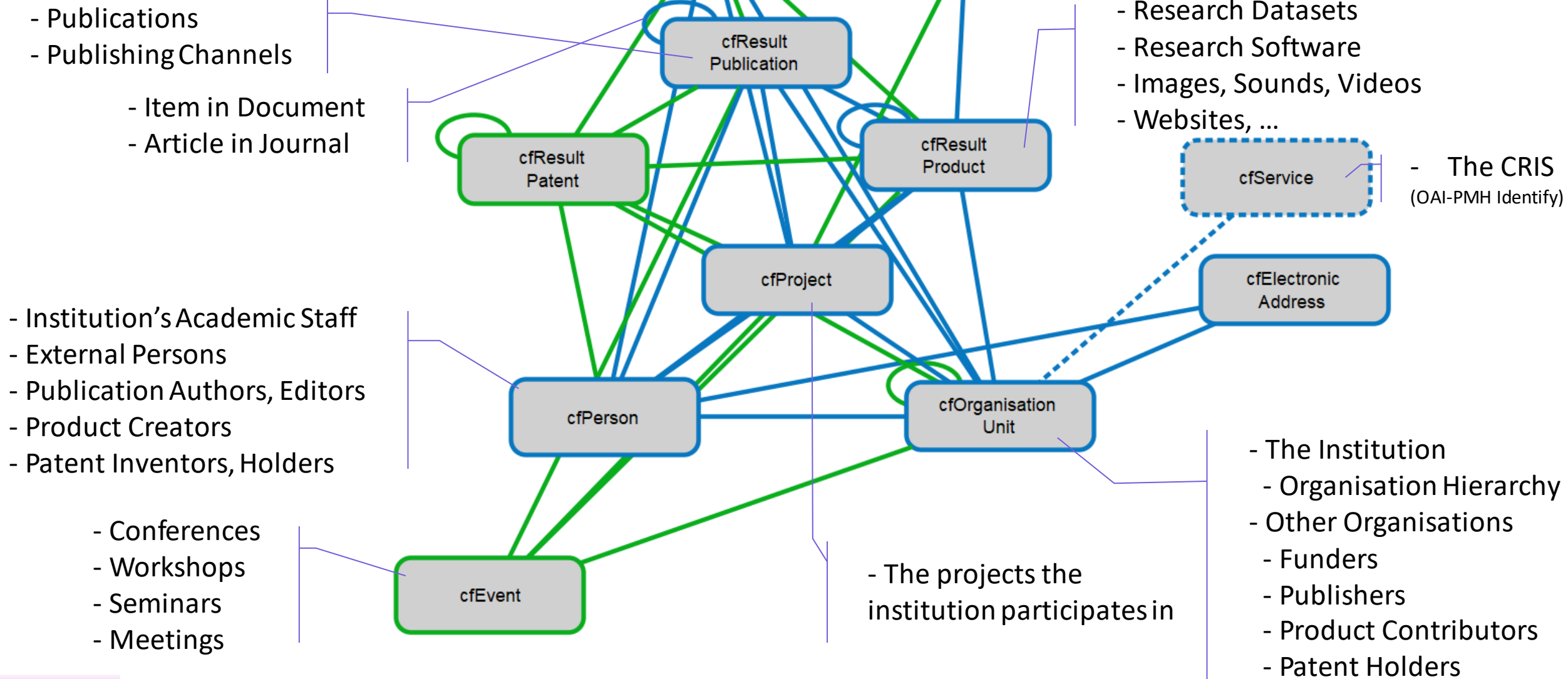


OpenAIRE CERIF Subset



Scope: Institutional CRIS

JD



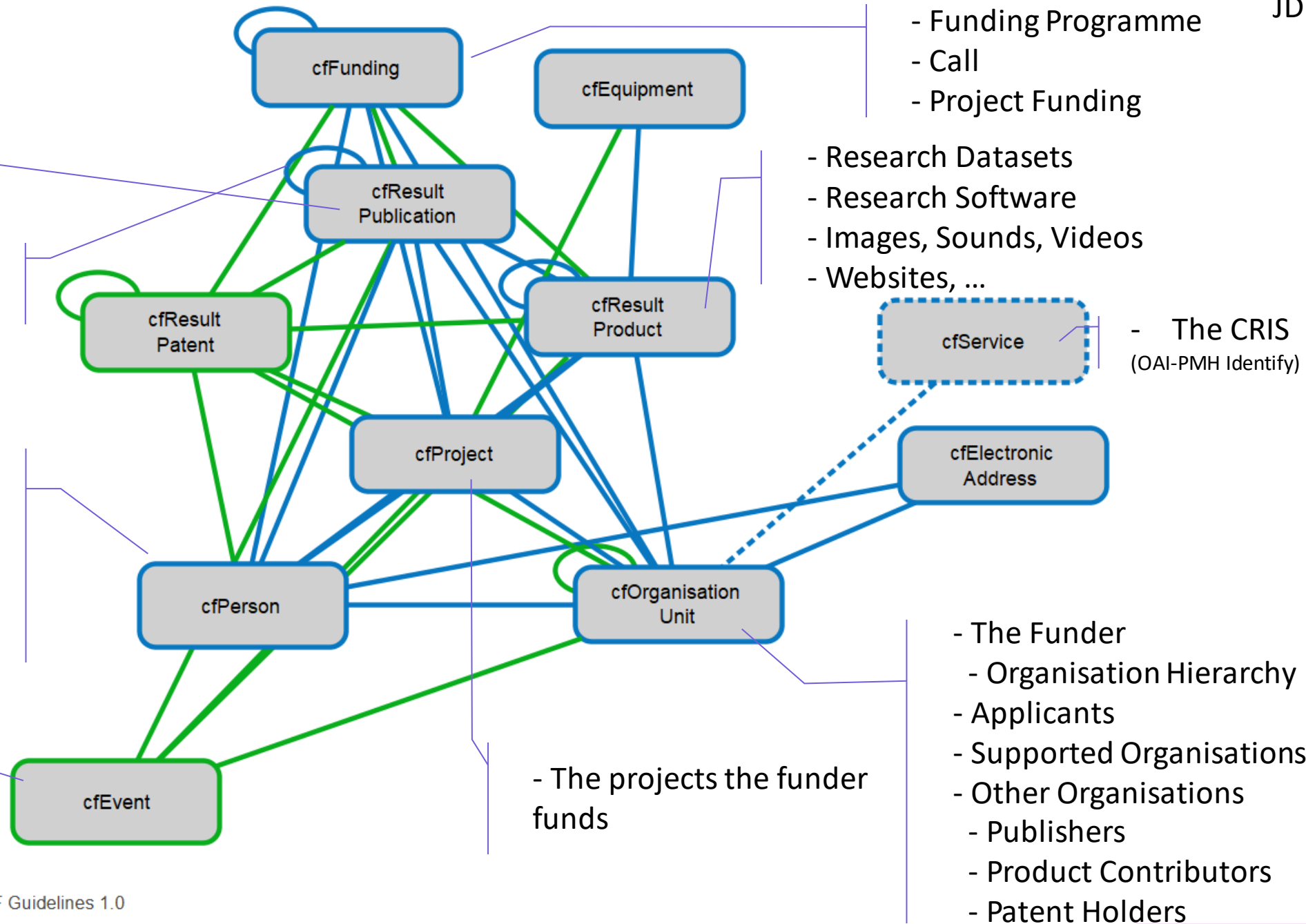
— since OpenAIRE CERIF Guidelines 1.0

— added in OpenAIRE CERIF Guidelines 1.1

Scope: Funder CRIS

JD

- Publications
- Publishing Channels
- Item in Document
- Article in Journal
- Applicants
- Principal Investigators
- Organisation Contacts
- Publication Authors, Editors
- Product Creators
- Patent Inventors, Holders
- Conferences
- Workshops
- Seminars
- Meetings



— since OpenAIRE CERIF Guidelines 1.0

— added in OpenAIRE CERIF Guidelines 1.1

Tour of the OpenAIRE CERIF Entities

Publications in an Institutional CRIS

- Publications authored by our institution's staff/students
- Publications edited by our institution's staff/students
- Publications our institution published
- Publishing channels our staff direct or sit on the editorial board
- Publishing channels our staff/students published in
- Publishing channels with indicator values
(Journal Citation Report, Scimago Journal Report, Leiden Ranking, ...)

Publications in a Funder CRIS

- Publications from projects we funded
- Publications referenced in grant applications
- Publishing channels with indicator values
- Administrative documents (kept separately):
 - Call documentation
 - Project proposals (= grant applications)
 - Project proposal reviews
 - Panel decisions
 - Funding agreements
 - Reports from the funded projects: interim & final reports
 - Evaluation reports by committees/panels
 - Funding scheme proposals, justifications, approvals, evaluations, ...

Publication Types

Origin:

- Where does your classification come from?
- How long have you been using it?

Mapping:

- Is the COAR Resource Types somewhere too fine so you couldn't map unambiguously?
- Do you have counterparts for all COAR Resource Types?

https://openaire-guidelines-for-cris-managers.readthedocs.io/en/latest/cerif_xml_publication_entity.html#type

Publications our staff/students authored

Bibliographic record origin:

- Entered manually?
- Imported from
 - Web of Science?
 - Scopus?
 - Crossref?
 - ORCID?
 - other?

Affiliations:

- Do you check your institution is listed as an affiliation of "your" authors?
- How deep in the organisational structure do you go when recording affiliations?
- Do you also record other affiliations of your staff/students?

Other Publication Context

- PublishedIn Publication?
 - Controlled list of journals?
- PartOf Publication?
- Editors? (Person or OrgUnit)
 - with your institution's affiliation?
 - with any affiliation?
- Publisher? (OrgUnit or Person)
 - Controlled list of publishers?
- OriginatesFrom?
 - Projects?
 - Fundings?
- References?
 - Datasets?
 - Patents?
 - Other Publications?
- PresentedAt Event?
- Coverage of Event?
- OutputFrom Event?

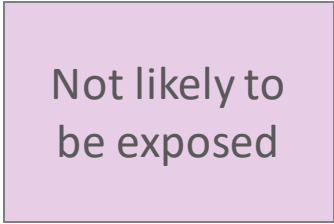
Projects

In an institutional CRIS

- Projects our institution participates in
 - With external funding?
 - Contractual research?
 - Internal projects?
- Projects our institution applied for funding for
 - Unsuccessful grant proposals
 - CRIS supports the internal approval process?

In a funder CRIS

- Projects we fund
- Project proposals



Not likely to be exposed

Project Context

The Consortium

- Coordinator? (OrgUnit or Person)
- Partners? (OrgUnit or Person)
- Contractors? (OrgUnit or Person)
- In-kind contributors? (OrgUnit or Person)
- Member? (OrgUnit or Person)

The Team

- Principal Investigator? (Person)
- Contact? (Person)
- Members? (Person)

with Affiliations?

- How deep in the org structure?

Funding: Funding Programmes & Calls & Tenders

In Institutional CRISs

- Funding opportunities
 - Prospective
 - Past

In Funder CRISs

- Funding schemes we have
- The calls we despatch

- Tracking “competition”?

Funding: Per-project Funding

In Institutional CRISs

- The funding we receive
- The funding we distribute
(as a Coordinator)
- The funding we apply for?

In Funder CRISs

- The funding we distribute
- Amounts typically public
 - E.g. CORDIS Open Data
 - Authoritative info

Amounts not likely to be exposed

Events (conferences, workshops, ...)

In Institutional CRISs

- The events where papers by our staff/students have been accepted
- The events where papers by our staff/students are being submitted
- The events our staff/students attend
- The events we organize

In Funder CRISs

- The events where paper with our funding have been accepted
- The events that were/will be attended with our funding
- The events that were/will be organized with our funding
- The events we organize

Products: Datasets, Software & other stuff

- Do you have the records?
- Are they properly classified?
- Were there difficulties mapping to the COAR Resource Types scheme?

Equipment

- Equipment we used for producing a dataset

OrgUnits

- How deep org structure do you keep
 - Of your own organization?
 - Of other organizations?
- How deep do you expose?
- Multilingual names?

OpenAIRE: validation and production phase

An introduction

Agenda

- Prerequisites for validation and experiences with validating (using the [prototype] validator) and/or the XML schema
- OpenAIRE beta infrastructure
 - Validate
 - Registrate
 - Explore
- Harvesting: Collecting, Validating and Transforming
- Production phase

Prerequisites for validation

- Use of one Current Research Information System (CRIS)
- OAI-PMH endpoint that provide the CRIS-CERIF standard
- Consider that the CRIS supports the OpenAIRE CRIS Guidelines, latest version



Software	
IRiS	70
Pure	51
In-house-built	43
Omega-PSiR	20
DSpace-CRIS	9
Converis	7
Symplectic Elements	4
UX00-INV	3
Worktribe	3
Metis	2
next >	

source: <https://dspacecris.eurocris.org/cris/explore/drisc>

Prerequisites for validation

- Login at (β.)PROVIDE dashboard
- OpenAIRE -- Prototype Validator
 - <https://github.com/jdvorak001/openaire-cris-validator>
 - OpenAIRE Guidelines for CRIS managers (XML)
 - <https://github.com/openaire/guidelines-cris-managers>
 - <https://www.openaire.eu/schema/cris/current/>

Experiences with validating

- Prototype validation of a CRIS repository

```
JUnit version 4.12
.Fetching and validating https://....api/cerif?verb=Identify
.Fetching and validating https://....api/cerif?verb=ListMetadataFormats
E.Fetching and validating https://....api/cerif?verb=ListSets
...
.Fetching and validating
https://....api/cerif?verb=ListRecords&metadataPrefix=oai_cerif_openaire&set=openaire_cris_persons
In Person[@id="0000-0002-7693-178X"]: cvc-pattern-valid: Wert '0000-0002-7693-178X' ist nicht Facet-gültig in
Bezug auf Muster 'https://orcid.org/0000-0001-5912-0041[0-9]{3}-[0-9]{3}[0-9X]' für Typ
'#AnonType_ORCID__Type'.
In Person[@id="0000-0003-1476-3087"]: cvc-pattern-valid: Wert '0000-0003-1476-3087' ist nicht Facet-gültig in
Bezug auf Muster 'https://orcid.org/0000-0001-5912-0041[0-9]{3}-[0-9]{3}[0-9X]' für Typ
'#AnonType_ORCID__Type'.
...
```

Experiences with validating

- Prototype validation of a CRIS repository

Time: 207,213

There was 1 failure:

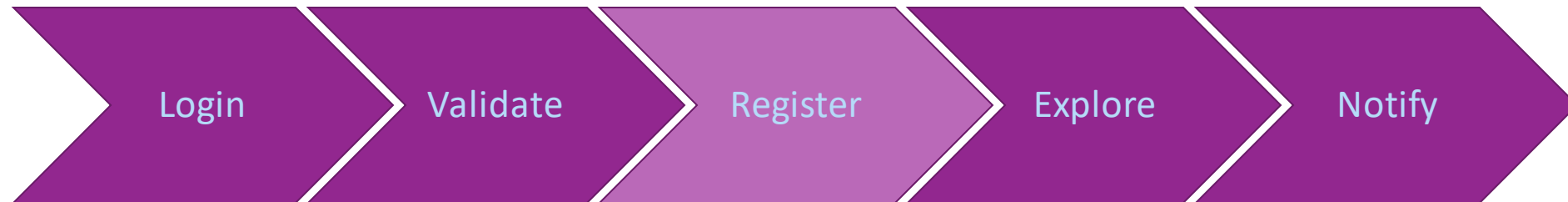
```
1) check010_MetadataFormats(org.eurocris.openaire.cris.validator.CRISValidator)
java.lang.AssertionError: Please reference the official XML Schema at https://www.openaire.eu/schema/cris/ (2)
    at org.junit.Assert.fail(Assert.java:88)
    at org.junit.Assert.assertTrue(Assert.java:41)
    at org.eurocris.openaire.cris.validator.CRISValidator$3.test(CRISValidator.java:337)
    at org.eurocris.openaire.cris.validator.CRISValidator$3.test(CRISValidator.java:324)
    at org.eurocris.openaire.cris.validator.util.CheckingIterable$MatchCountingIterator.next(CheckingIterable.java:153)
    at org.eurocris.openaire.cris.validator.util.CheckingIterable.run(CheckingIterable.java:30)
    at org.eurocris.openaire.cris.validator.CRISValidator.check010_MetadataFormats(CRISValidator.java:320)
    at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
    at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:62)
    at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
    at java.lang.reflect.Method.invoke(Method.java:498)
    at org.junit.runners.model.FrameworkMethod$1.runReflectiveCall(FrameworkMethod.java:50)
    at org.junit.internal.runners.model.ReflectiveCallable.run(ReflectiveCallable.java:12)
    at org.junit.runners.model.FrameworkMethod.invokeExplosively(FrameworkMethod.java:47)
    at org.junit.internal.runners.statements.InvokeMethod.evaluate(InvokeMethod.java:17)
    ...
    at org.junit.runner.JUnitCore.run(JUnitCore.java:137)
    at org.junit.runner.JUnitCore.run(JUnitCore.java:115)
    at org.junit.runner.JUnitCore.runMain(JUnitCore.java:77)
    at org.junit.runner.JUnitCore.main(JUnitCore.java:36)
    at org.eurocris.openaire.cris.validator.CRISValidator.main(CRISValidator.java:153)
```

FAILURES!!!

Tests run: 13, Failures: 1

(1/13 fails; ~7% but only 2 records)

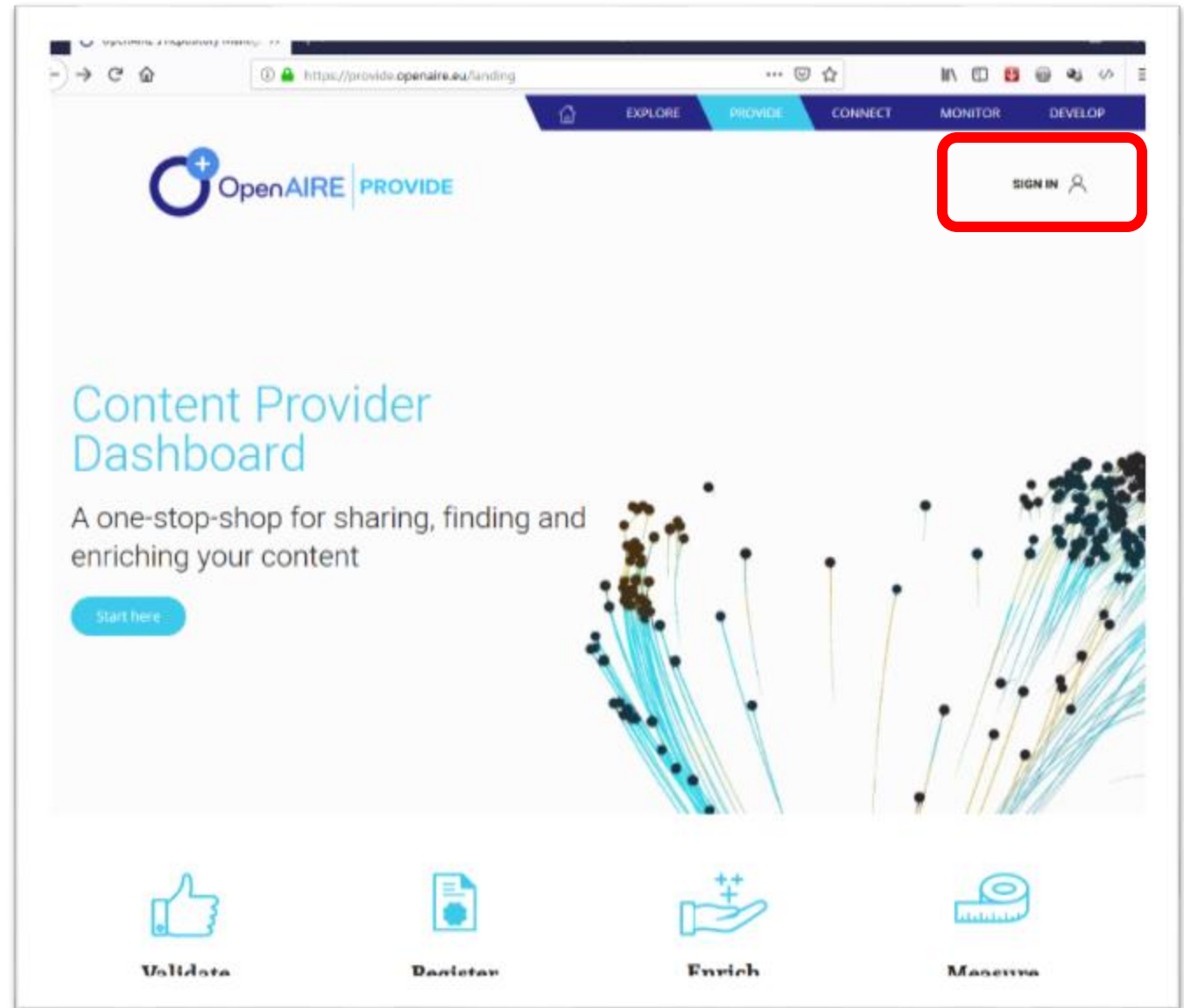
(β -)Infrastructure: Step's to proceed



Step's to proceed

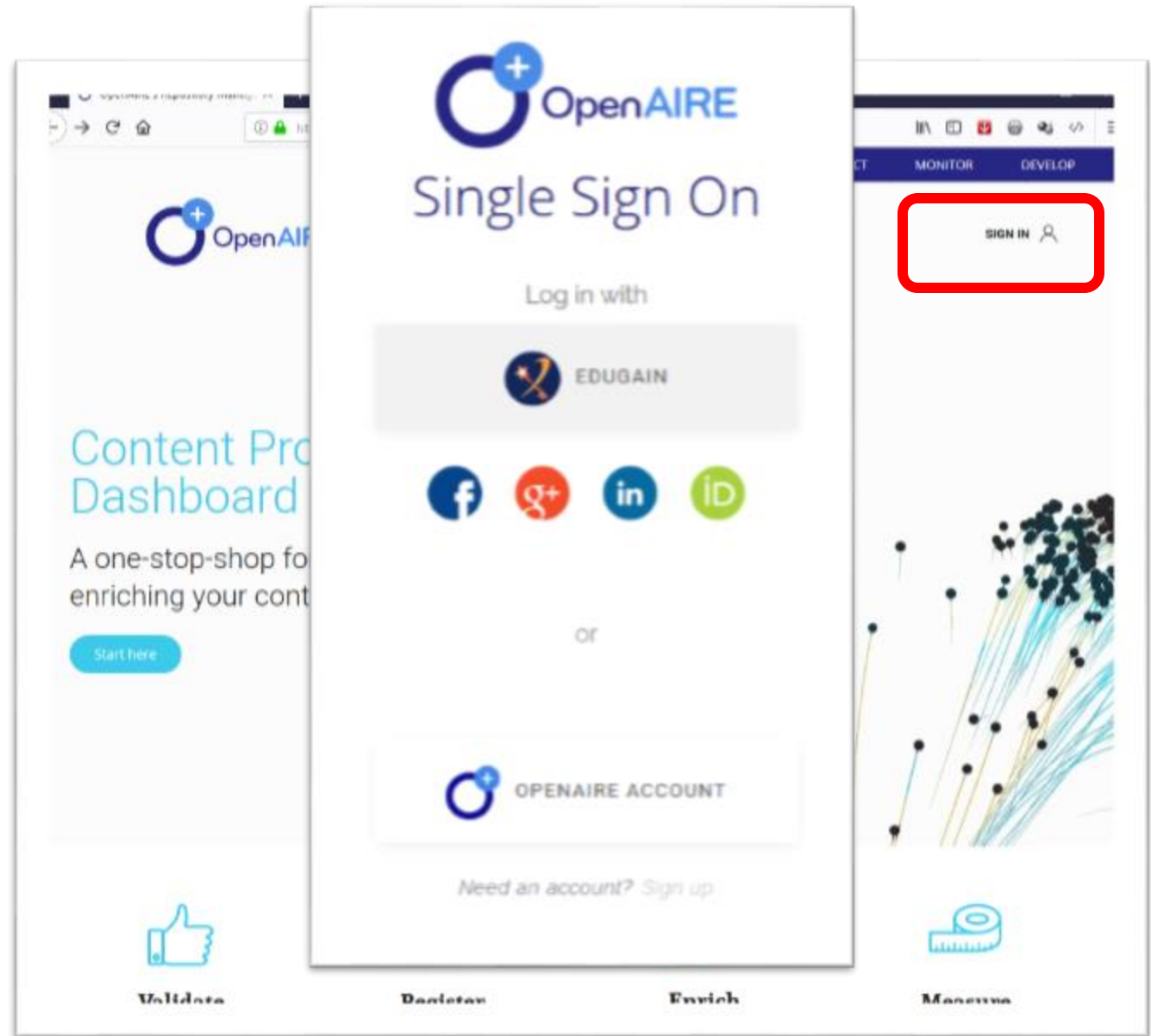


Login



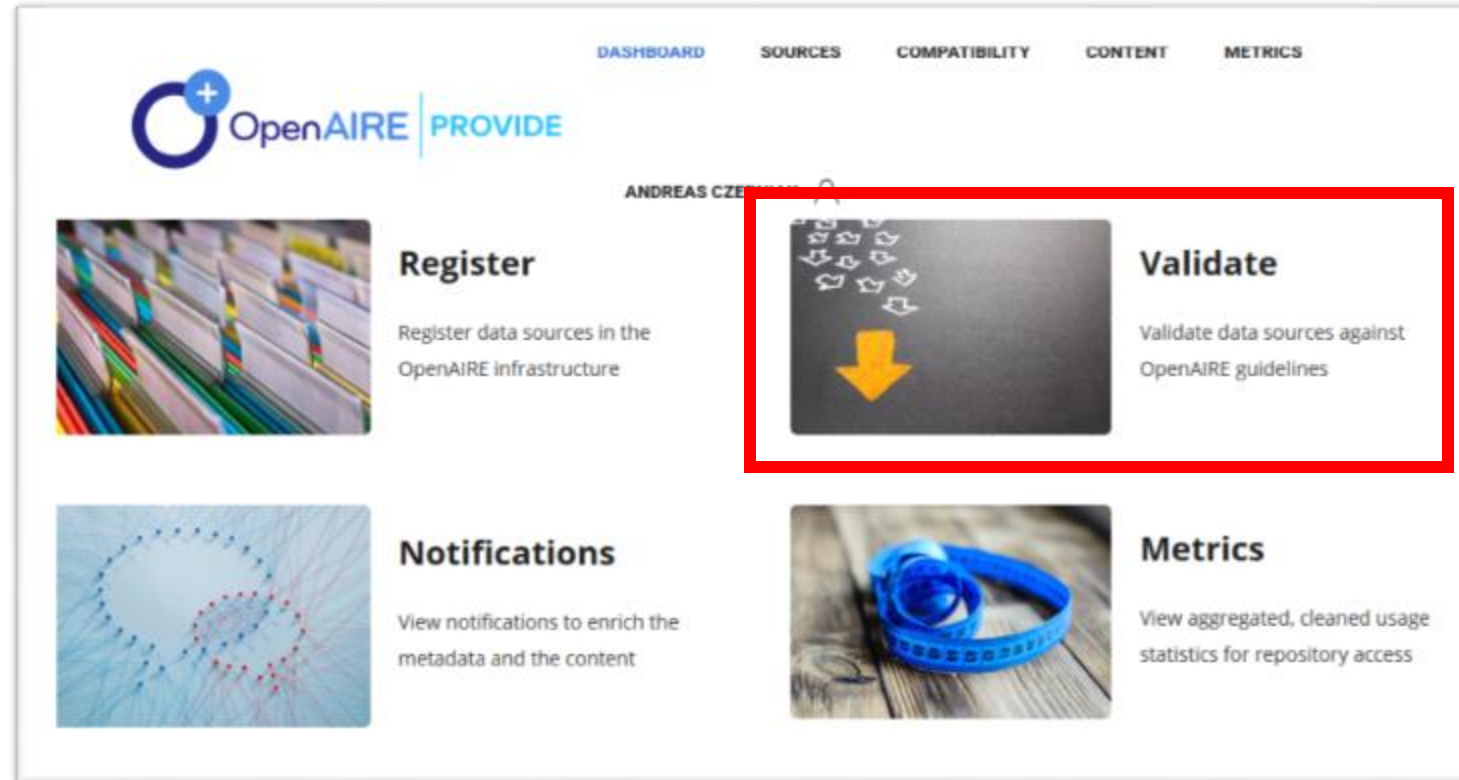
Step's to proceed

Login

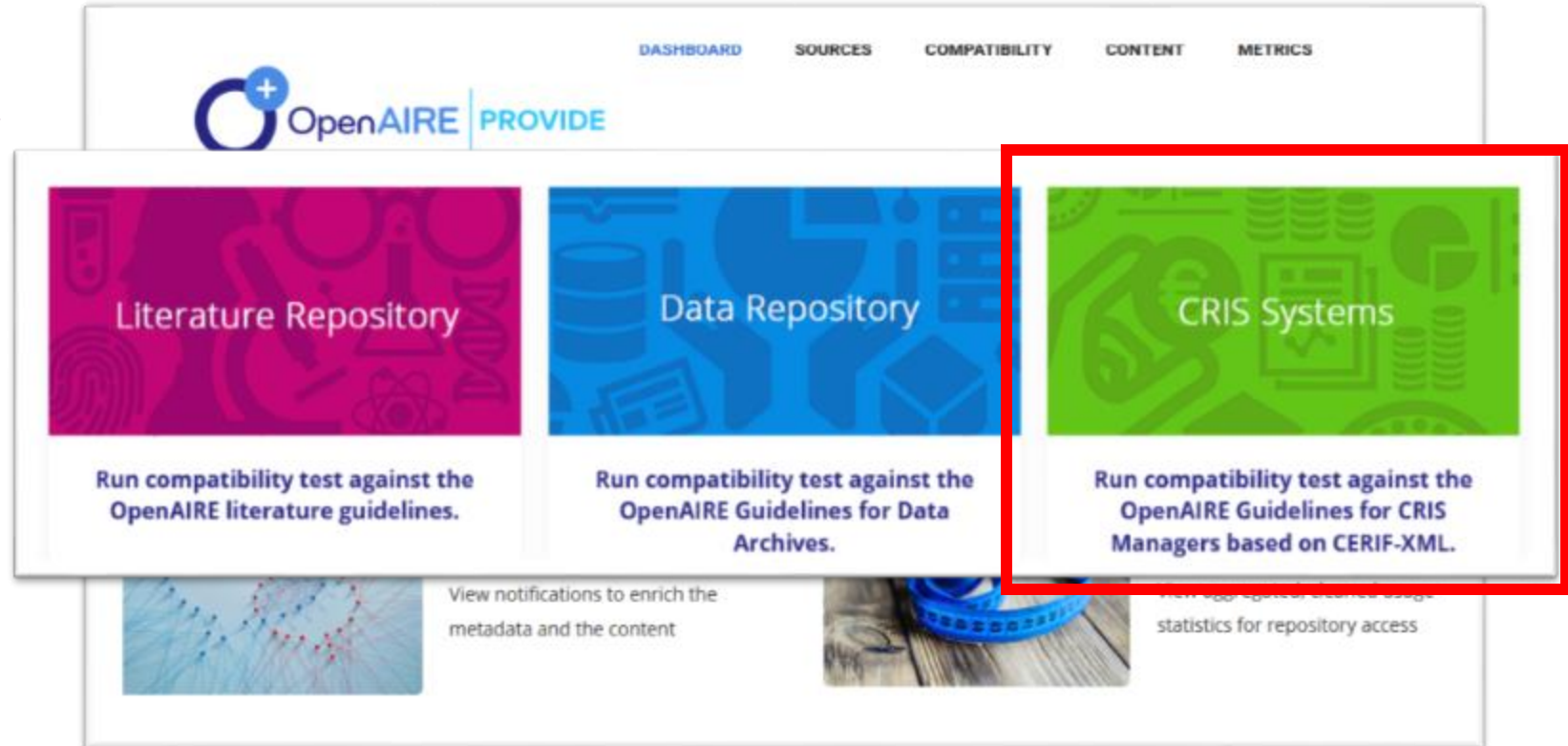


Step's to proceed

Validate



Step's to proceed



Step's to proceed



Run compatibility test

Select datasource Select guidelines Select parameters Finish

☐ SELECT BASE URL FROM ONE OF YOUR REGISTERED REPOSITORIES

-- none selected --

☒ OR ENTER NEW

<https://dwitjutife1.csc.fi/api/cerif>

Next >>

Step's to proceed

Validate



Step's to proceed



Run compatibility test

Select datasource Select guidelines Select parameters Finish

Cris Entities

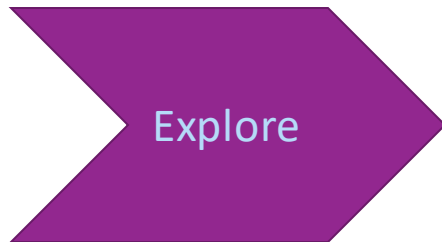
- ☒ Select / Deselect All Entities
- ☒ Funding
- ☒ Organisation
- ☒ Person
- ☒ Project
- ☒ Publication
- ☒ Product
- ☒ Service

Check Referential Integrity

☒ YES ☐ NO

« Back Finish »

Step's to proceed



After a new Index – explore your content on (beta.)explore.openaire.eu :

Compatibility Level: ✗ OpenAIRE CRIS v1.1

Results per page: 10 ⬆ ⬇ ⬇ ⬆

📍 ↓ (CSV) 2 content providers, page 1 of 1

Type (1)

☐ CRIS System(2)

Compatibility Level (1)

☒ OpenAIRE CRIS v... (2)

Metis Radboud University

CRIS SYSTEM OPENAIRE CRIS V1.1

Organization: [Radboud University Nijmegen](#)

Country: Netherlands

Website URL: <https://metistst.ru.nl/>

OAI-PMH URL: <https://oamemtf.uci.ru.nl/metis-oaipmh-endpoint/OAIHandler>

VIRTA

CRIS SYSTEM OPENAIRE CRIS V1.1

Organization: [CSC](#)

Country: Finland

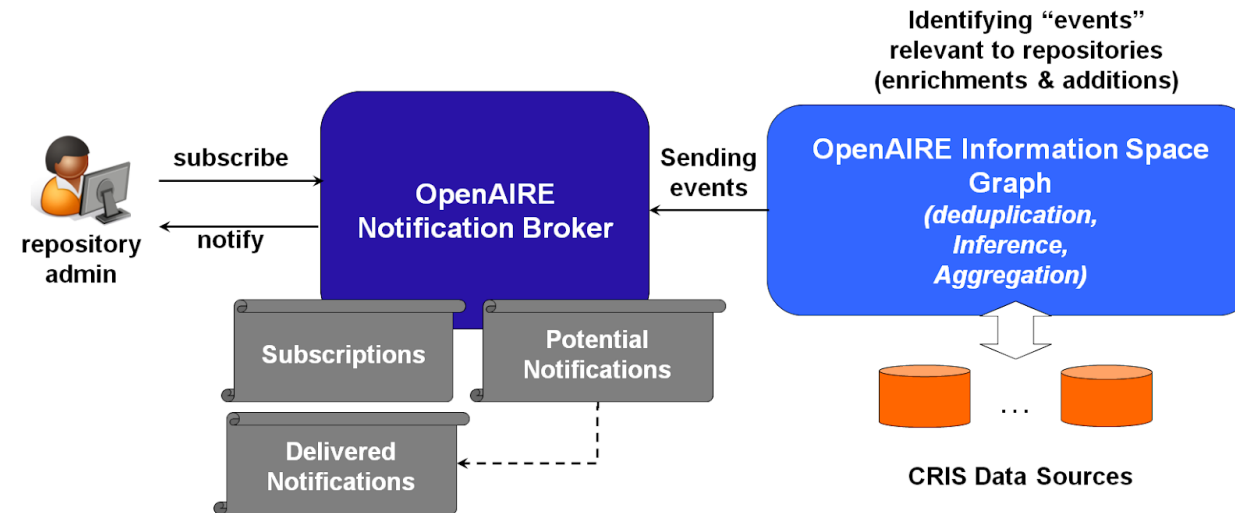
Website URL: <https://www.csc.fi>

OAI-PMH URL: <https://dwitjutife1.csc.fi/api/cerif>

Step's to proceed

Notify

Repository manager can use the Broker service to get information about the content.



Harvesting process



Harvesting: Collecting

- Collection interval: ~14 days
- Collecting every set separately
 - openaire_cris_publications
 - openaire_cris_patents
 - openaire_cris_products
 - openaire_cris equipments
 - ...

Harvesting: Transforming

native metadata is homogenised, enriched, filtered by transformation

`<DisplayName> Victoria L.
Cammann</DisplayName>`

tidy authors (e.g. invert)
=> Cammann, Victoria L.

`<DOI>10.1007/9783319446721_3
</DOI>`

dig out hidden identifiers
=> ISBN 9783319446721

`<URL>....pdf</URL>`

deduce implicit access levels
=> open access

missing landing page URL

construct landing page URLs
(in case of regular pattern)

Production phase

CRIS's in β -phase would be:

- automatic transfer into Production
 - New CRIS systems are welcome @ β -phase

In discussion

- Using the Directory of CRIS systems at euroCRIS in the same way as OpenDOAR for OA repositories.



any questions?
Thank you very much for your
attention

Contributors to this Presentation

Acro	Name	Affiliation(s)	Email
JD	Jan Dvořák orcid.org/0000-0001-8985-152X	euroCRIS / Czech Technical University / Charles University (Prague)	Jan.Dvorak@ff.cuni.cz
AC	Andreas Czerniak orcid.org/0000-0003-3883-4169	OpenAIRE / Bielefeld University	andreas.czerniak@uni-bielefeld.de
AL	Aenne Loehden	OpenAIRE / Bielefeld University	aenne.loehden@uni-bielefeld.de