



METIS2OpenAIRE

Implementing the CERIF XML Guidelines for CRIS managers

Jan Dvořák

 <https://orcid.org/0000-0001-8985-152X>


Pablo de Castro



 <https://orcid.org/0000-0001-6300-1033>

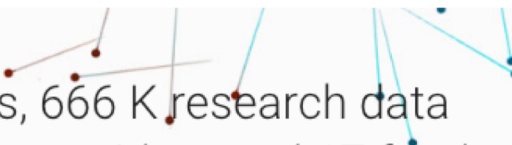
Outline

1. Intro
2. METIS OpenAIRE endpoint
3. Recommendations to implementors
4. Ongoing activities & next steps


OpenAIRE

EXPLOREPROVIDECONNECTMONITORDEVELOP

OpenAIRE | EXPLORESEARCHSHARELINKCONTENT PROVIDERSSIGN IN 




24 mi publications, 666 K research data
from 13 K content providers and 17 funders
linked together for an integrated research
search



Explore

Discover linked open
research


Search for publications,
datasets, software and other
research products. See how
these are linked together and
how they are linked to funding
and organizations.
View statistics on projects and



Share

Deposit in a repository
of your choice

Select an OpenAIRE compatible
repository (2.0 +) so that your
research is linked to your
funding information.
Use [ZENODO](#), a catch-all
repository hosted by CERN to
deposit all your research results



Link

Connect all your
research

If you can't find your research
results in OpenAIRE, don't
worry!
Use our **LINK OUT service**, that
reaches out to many external
sources via APIs, to connect
your research results and [claim](#)
them to your project.

feedback

<https://explore.openaire.eu/>

What does OpenAIRE collect?

- Scientific literature metadata and full-texts from **institutional and thematic repositories**, Open Access **journals** and publishers;
- Dataset metadata from **data repositories** and **data journals**;
- Scientific literature, data and software metadata from **Zenodo**;
- Metadata about data sources, organizations, projects, and funding programs from **entity registries**, i.e. authoritative sources such as CORDA and other funder databases for projects, OpenDOAR for publication repositories, re3data for data repositories, DOAJ for Open Access journals;
- Coming soon: metadata of open source research software from **software repositories** (currently available only on <https://beta.explore.openaire.eu/>)
- Coming soon: metadata about **other types of research products**, like workflow, protocols, methods, research packages (currently available only on <https://beta.explore.openaire.eu/>)
- Coming soon: metadata about scientific literature, datasets, persons, organisations, projects, funding, equipment and services are collected through **CRIS** (Common Research Information Systems)

Standardised metadata aggregation: OpenAIRE Guidelines

OpenAIRE Guidelines

Welcome to the OpenAIRE Guidelines. The intention of this is to provide a public space to share OpenAIREs work on interoperability and to engage with the community. The OpenAIRE Guidelines helps repository managers expose publications, datasets and CRIS metadata via the OAI-PMH protocol in order to integrate with OpenAIRE infrastructure.

OpenAIRE Guidelines have been released for publication repositories, data archives and CRIS systems respectively:

Current Guidelines

- [OpenAIRE Guidelines for Literature Repositories](#)
- [OpenAIRE Guidelines for Data Archives](#)
- [OpenAIRE Guidelines for CRIS Managers based on CERIF-XML](#)



Type

Sort by results number

☐ Journal (12,062)

☐ Institutional Repository (685)

☐ Journal Aggregator/Publisher (102)

☐ Data Repository (93)

☐ Publication Repository (76)

☐ Thematic Repository (57)

☐ Publication Repository Aggreg... (53)

☐ other (28)

☐ Publication Catalogue (11)

☐ Institutional Repository Aggreg... (9)

OpenAIRE data provider validation



DASHBOARD

SOURCES

COMPATIBILITY

CONTENT

METRICS

Validate your datasource

Literature Repository

Run compatibility test against the OpenAIRE literature guidelines.

Data Repository

Run compatibility test against the OpenAIRE Guidelines for Data Archives.

CRIS Systems

Run compatibility test against the OpenAIRE Guidelines for CRIS Managers based on CERIF-XML.

<https://provide.openaire.eu/compatibility/validate>

July 3, 2015

Dataset

Open Access

OpenAIRE Guidelines for CRIS Managers 1.0

Houssos, Nikos; Joerg, Brigitte; Dvořák, Jan

The Guidelines specify the interoperability layer between Current Research Information Systems (CRIS) and the OpenAIRE infrastructure. The information interchange is based on the Common European Research Information Format (CERIF) data model, the CERIF XML exchange format, and the OAI-PMH protocol. The Guidelines are intended mainly for implementers and administrators of CRIS who plan to communicate research information to OpenAIRE. OpenAIRE (openaire.eu) is the European infrastructure enabling researchers to comply with the European Union requirements for Open Access to research results. OpenAIRE collects metadata from a variety of data sources: publication repositories, data archives and CRIS across Europe and beyond. Interoperability guidelines are defined for each type of source. CERIF is a standard data model for research information and a recommendation by the European Union to its Member States. The custody of CERIF has been entrusted by the European Union to euroCRIS (eurocris.org), an international not-for-profit organisation dedicated to the interoperability of CRIS.

The Guidelines consist of the following files: Guidelines document: - OpenAIRE_Guidelines_for_CRIS_Managers_v.1.0.pdf OpenAIRE CERIF Semantics: - OpenAIRE_CERIF_Semantics_v.1.0.xlsx - OpenAIRE_CERIF_Semantics_v.1.0.xml OpenAIRE CERIF XML Schema: - cerif-1.6-2_openaire-1.0.xsd Examples: - openaire_cerif_xml_example_datasets.xml - openaire_cerif_xml_example_funding.xml - openaire_cerif_xml_example_orgunits.xml - openaire_cerif_xml_example_persons.xml - openaire_cerif_xml_example_projects.xml - openaire_cerif_xml_example_publications.xml - openaire_cerif_xml_example_services.xml

[Preview](#)

1,152

views

404

downloads

[See more details...](#)[See more details](#)

Tweeted by 8

6 readers on Mendeley

Indexed in

OpenAIRE

Publication date:

July 3, 2015

DOI:

DOI 10.5281/zenodo.17065

Keyword(s):

OpenAIRE guidelines CERIF CRIS interoperability

June 26, 2018

Project milestone

Open Access

OpenAIRE Guidelines for CRIS Managers 1.1

 Dvořák, Jan;  Bollini, Andrea;  Rémy, Laurent;  Schirrwagen, Jochen

The Guidelines provide orientation for CRIS managers to expose their metadata in a way that is compatible with the OpenAIRE infrastructure. By implementing the Guidelines, CRIS managers support the inclusion and therefore the reuse of metadata in their systems within the OpenAIRE infrastructure. For developers of CRIS platforms, the Guidelines provide guidance to add supportive functionalities for CRIS managers and users. Exchange of information between individual CRIS systems and the OpenAIRE infrastructure is an example of point-to-point data exchange between CRIS systems, since the OpenAIRE infrastructure is itself a CRIS system.

[Preview](#)[Files](#) (548.9 kB)**Name****Size**[OpenAIRE_Guidelines_for_CRIS_Managers-samples_v1.1.0.zip](#)

24.0 kB

 Preview Download

md5:af8e5f43ba8fb38540c060aae7140ae9 ?

[OpenAIRE_Guidelines_for_CRIS_Managers-schemas_v1.1.0.zip](#)

73.5 kB

 Preview Download

md5:23cd4c1e60858205b8bc65f7f4125766 ?

[OpenAIRE_Guidelines_for_CRIS_Managers_v1.1.0.pdf](#)

451.3 kB


 Preview Download

md5:07ea0c5df8480674da406cd3d3ea9345 ?

289

 views

136

 downloads[See more details...](#) Tweeted by 5[See more details](#)

Indexed in

OpenAIRE**Publication date:**

June 26, 2018

DOI:DOI [10.5281/zenodo.1298650](#)**Keyword(s):**

content aggregation


OpenAIRE.eu

OpenAIRE Guidelines 1.1 features

- CERIF Profile
 - Uses the updated CERIF XML format
- Aligned vocabularies
 - COAR Resource Types
 - COAR Access Rights
 - ISSN Media Types
- Scope slightly expanded from 1.0
 - Patents
 - Events
 - a few links

Project (on GitHub)

[https://github.com/openaire/
guidelines-cris-managers](https://github.com/openaire/guidelines-cris-managers)

 Search or jump to... Pull requests Issues Marketplace Explore

openaire / guidelines-cris-managers


Unwatch 19 Star 0 Fork 5

[Code](#) [Issues 9](#) [Pull requests 1](#) [Projects 1](#) [Wiki](#) [Insights](#)

OpenAIRE Guidelines for CRIS Managers based on CERIF-XML

315 commits 6 branches 2 releases 5 contributors

Branch: master New pull request Create new file Upload files Find file Clone or download

 jdvorak001 updated the sample map (person with DAI and ISNI) Latest commit a096a25 11 days ago

conf	added Schematron rules for element OAMandate	5 months ago
docs	updated the sample map (person with DAI and ISNI)	11 days ago
samples	a sample person with a DAI (and ISNI)	11 days ago
schemas	Add person-identifier type: Digital Author Identifier (DAI). (#49)	11 days ago
.gitignore	ignore eclipse stuff	2 years ago
.travis.yml	Travis default git behaviour should be o.k.	6 months ago
README.rst	alt on the build status badge	5 months ago
pom.xml	version 1.1.1-SNAPSHOT	11 days ago

README.rst

OpenAIRE Guidelines for CRIS Managers based on CERIF-XML

This repository contains the project of the OpenAIRE Guidelines for CRIS Managers 1.1,

DOI [10.5281/zenodo.1298650](https://doi.org/10.5281/zenodo.1298650)

Contents


build passing docs passing

XML Schema: [schemas/openaire-cerif-profile.xsd](#)

Examples: map at [docs/_illustrations/OpenAIRE-examples-map.png](#), files at [samples/](#)

Prototype Validator (on GitHub)

<https://github.com/jdvorak001/openaire-cris-validator>

 Search or jump to... Pull requests Issues Marketplace Explore

jdvorak001 / openaire-cris-validator

Unwatch 4 Star 2 Fork 0

<> Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

A validator tool to assess the compliance of an OAI-PMH endpoint with the OpenAIRE Guidelines for CRIS Managers 1.1. Edit

openaire oai-pmh-client validation xml-schema Manage topics

121 commits 2 branches 2 releases 1 contributor Apache-2.0

Branch: master New pull request Create new file Upload files Find file Clone or download

jdvorak001 Release 1.1.0 description Latest commit 7b28208 5 minutes ago

samples	check the examples from the OpenAIRE Guidelines for CRIS Managers	7 months ago
src	added the schema for provenance (fixes #1)	18 days ago
.gitignore	ignore important files	6 months ago
.travis.yml	build instructions rearranged	6 months ago
CHECKS.md	the checks moved in an extra file	6 months ago
LICENSE	Initial commit	8 months ago
README.md	Release 1.1.0 description	5 minutes ago
pom.xml	priority to local cached schemas when copying resources	6 months ago

README.md

OpenAIRE CRIS validator

A tool to assess whether an OAI-PMH endpoint can provide research information complying with the [OpenAIRE Guidelines for CRIS Managers 1.1](#). It covers [these checks](#).

This is a command-line Java tool that is organized as a [JUnit](#) test suite. You can also run it in your IDE.

Please read below how to [build](#) it, [run](#) it, [explore](#) its internals and give [feedback](#). This is Open Source software, available under the terms of the [Apache 2.0 License](#).

Current status

build passing ← checking if the software builds and runs on the [example files from the standard](#).

METIS2OpenAIRE

Project coordinated by euroCRIS

Funded by OpenAIRE

METIS2OpenAIRE Concept

- Main aim: to make a first CRIS System OpenAIRE-compatible (METIS at Radboud University Nijmegen)
- In-house-built system chosen as allows full institutional control on dev
- To widen the project scope, further CRIS solutions included as budget-neutral partners beyond METIS: Omega-PSIR & PURE
- Project WPs include:
 - Setting up an OAI-PMH end-point on METIS
 - Building a minimally sufficient validator for CERIF-XML CRIS feeds
 - Test METIS feed against CERIF-XML Guidelines in several iterations until compliance is achieved
 - Support parallel mapping exercises by budget-neutral project partners

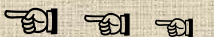
METIS

- In-house built CRIS at Radboud University (Nijmegen, NL)
- Currently in use at two Dutch Universities

Now with an OpenAIRE CRIS Guidelines 1.1 compatible endpoint



<https://oamemtfa.uci.ru.nl/metis-oaipmh-endpoint/OAIHandler>



(RU acceptance testing environment)

METIS
OAI
Endpoint

Identify →

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0">
  <responseDate>2018-11-24T16:58:43Z</responseDate>
  <request verb="Identify">https://oamemtfa.uci.ru.nl/metis-oaipmh-endpoint/OAIHandler</request>
  <Identify>
    <repositoryName>Metis Radboud University</repositoryName>
    <baseURL>https://oamemtfa.uci.ru.nl/metis-oaipmh-endpoint/OAIHandler</baseURL>
    <protocolVersion>2.0</protocolVersion>
    <adminEmail>admin@ru.nl</adminEmail>
    <earliestDatestamp>1990-01-01</earliestDatestamp>
    <deletedRecord>persistent</deletedRecord>
    <granularity>YYYY-MM-DD</granularity>
    <compression>gzip</compression>
    <compression>deflate</compression>
    <description>
      <Service xmlns="https://www.openaire.eu/cerif-profile/1.1/" id="xxx">
        <Compatibility xmlns="https://www.openaire.eu/cerif-
profile/vocab/OpenAIRE_Service_Compatibility">https://www.openaire.eu/cerif-
profile/vocab/OpenAIRE_Service_Compatibility#1.1</Compatibility>
        <Acronym>metis.ru.nl</Acronym>
        <Name xml:lang="en">Metis Radboud University</Name>
        <Description xml:lang="en">Metis, the CRIS of Radboud University that complies with the OpenAIRE
Guidelines for CRIS Managers v.1.1.1.</Description>
        <WebsiteURL>http://www.ru.nl/research-information-services/</WebsiteURL>
        <OAIPMHBaseURL>https://oamemtfa.uci.ru.nl/metis-oaipmh-endpoint/OAIHandler</OAIPMHBaseURL>
        <Owner>
          <OrgUnit id="00000000">
            <Acronym>RU</Acronym>
          </OrgUnit>
        </Owner>
      </Service>
    </description>
    <!-- ... -->
  </Identify>
</OAI-PMH>
```

METIS
OAI
Endpoint

ListMetadata
Formats →

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/">
  <responseDate>2018-11-24T16:58:51Z</responseDate>
  <request verb="ListMetadataFormats">http://oamemtfa.uci.ru.nl:80/metis-oaipmh-endpoint/OAIHandler</request>
  <ListMetadataFormats>
    <metadataFormat>
      <metadataPrefix>didl</metadataPrefix>
      <schema>http://standards.iso.org/ittf/PubliclyAvailableStandards/MPEG-
21_schema_files/did/didl.xsd</schema>
      <metadataNamespace>urn:mpeg:mpeg21:2002:02-DIDL-NS</metadataNamespace>
    </metadataFormat>
    <metadataFormat>
      <metadataPrefix>oai_dc</metadataPrefix>
      <schema>http://www.openarchives.org/OAI/2.0/oai_dc.xsd</schema>
      <metadataNamespace>http://www.openarchives.org/OAI/2.0/oai_dc/</metadataNamespace>
    </metadataFormat>

    <metadataFormat>
      <metadataPrefix>oai_cerif_openaire</metadataPrefix>
      <schema>https://github.com/openaire/guidelines-cris-managers/raw/master/schemas/openaire-cerif-
profile.xsd</schema>
      <metadataNamespace>https://www.openaire.eu/cerif-profile/1.1/</metadataNamespace>
    </metadataFormat>

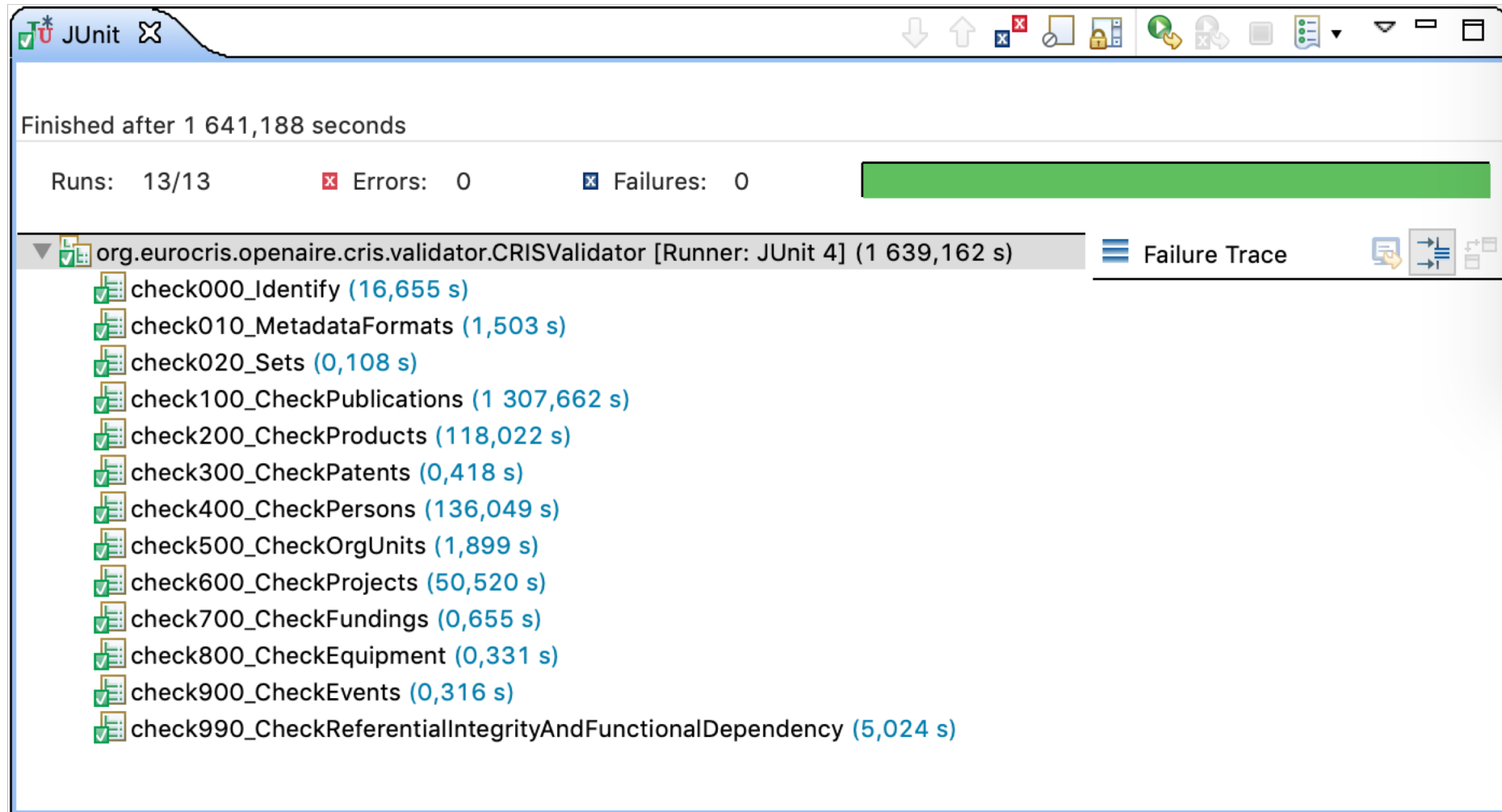
    <metadataFormat>
      <metadataPrefix>pk_dare_all</metadataPrefix>
      <schema>http://purl.utwente.nl/ns/metis.xsd</schema>
      <metadataNamespace>http://poas.uci.kun.nl/metis</metadataNamespace>
    </metadataFormat>
  </ListMetadataFormats>
</OAI-PMH>
```

METIS
OAI
Endpoint

ListSets →

```
<?xml version="1.0" encoding="UTF-8" ?>
<OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:dc="http://purl.org/dc/elements/1.1/">
  <responseDate>2018-11-24T16:58:58Z</responseDate>
  <request verb="ListSets">http://oamemtfa.uci.ru.nl:80/metis-oaipmh-endpoint/OAIHandler</request>
  <ListSets>
    <set>
      <setSpec>openaire_cris_products</setSpec>
      <setName>OpenAIRE_CRIS_products</setName>
      <setDescription>
        <dc:description>The list of CERIF XML records for datasets and other research products.</dc:description>
      </setDescription>
    </set>
    <set>
      <setSpec>openaire_cris_projects</setSpec>
      <setName>OpenAIRE_CRIS_projects</setName>
      <setDescription>
        <dc:description>The list of CERIF XML records for projects.</dc:description>
      </setDescription>
    </set>
    <set>
      <setSpec>openaire_cris_persons</setSpec>
      <setName>OpenAIRE_CRIS_persons</setName>
      <setDescription>
        <dc:description>The list of CERIF XML records for persons.</dc:description>
      </setDescription>
    </set>
    <set>
      <setSpec>openaire_cris_orgunits</setSpec>
      <setName>OpenAIRE_CRIS_orgunits</setName>
      <setDescription>
        <dc:description>The list of CERIF XML records for organisations and organisation units.</dc:description>
      </setDescription>
    </set>
    <!-- ... -->
  </ListSets>
</OAI-PMH>
```

METIS Check with the Prototype Validator



The image shows a JUnit test runner window. At the top, it says "Finished after 1 641,188 seconds". Below this, it shows "Runs: 13/13", "Errors: 0", and "Failures: 0". A green progress bar is visible. The main section lists the test results for the package `org.eurocris.openaire.cris.validator.CRISValidator`. Each test is marked with a green checkmark and its duration in seconds.

Test Name	Duration (s)
<code>check000_Identify</code>	16,655
<code>check010_MetadataFormats</code>	1,503
<code>check020_Sets</code>	0,108
<code>check100_CheckPublications</code>	1 307,662
<code>check200_CheckProducts</code>	118,022
<code>check300_CheckPatents</code>	0,418
<code>check400_CheckPersons</code>	136,049
<code>check500_CheckOrgUnits</code>	1,899
<code>check600_CheckProjects</code>	50,520
<code>check700_CheckFundings</code>	0,655
<code>check800_CheckEquipment</code>	0,331
<code>check900_CheckEvents</code>	0,316
<code>check990_CheckReferentialIntegrityAndFunctionalDependency</code>	5,024

METIS Contents Summary

OpenAIRE_CRIS_publications:	~ 132,500 items (incl. ~14,200 journals)
OpenAIRE_CRIS_products:	~ 17,000 items
OpenAIRE_CRIS_patents:	—
OpenAIRE_CRIS_persons:	~ 25,000 items
OpenAIRE_CRIS_orgunits:	~ 550 items
OpenAIRE_CRIS_projects:	~ 5,100 items
OpenAIRE_CRIS_funding:	—
OpenAIRE_CRIS_events:	—
OpenAIRE_CRIS equipments:	—

295 MB of total XML
268 MB of CERIF XML
~7 mio. lines
takes ~30 min

Recommendations to Implementers

The do's and don'ts
of setting up
an OpenAIRE CRIS Guidelines 1.1 compatible endpoint
in your institutional CRIS

Recommendations for the devel process

Start with a reasonably sized sample of your CRIS data

E.g. 2 larger departments × projects, events & outputs from 3 years

Real data!

Test & validate as much as possible

Preferably as a part of your Continuous Integration setup

=> Invest some time up front! <=
(... and avoid getting grey hair from waiting ...)

Recommendations for the QA process

Correct problems in your data as they surface

Unless you've always had strict validations on inputs and perfect data migrations.

Typos in idanteflyer sintex, strænge cHaråktørß, h-y-p—h_e—ns' hell, `quo»tes“,

n

e

w

l

i

n

e

s, ...

E.g.:

In Publication[@id="448992"]: cvc-pattern-valid:

Value '**10.1049/ip-ifs: 20055019**' is not facet-valid with respect to pattern '10\.\d{4,}(\.\d+)*/[^\s]+' for type '**DOI__SimpleType**'.

=> Be nice to your librarian! <=
(You will need her/his assistance.)

Suggested feature implementation sequence

0. An empty, but compliant OAI endpoint (Service)
1. Your institution and its organisational structure (OrgUnits)
2. Your academic staff with affiliations to your departments (Persons, OrgUnits)
3. The funders you have met (OrgUnits)
4. Funding programmes, incl. the links to their funders (Funding, OrgUnits)
5. Equipment
6. Projects
7. Journals your staff has published in (Publications)
8. Your staff's journal articles (Publications)
9. Proceedings, conferences (Publications, Events)
10. Your staff's other publication outputs (Publications)
11. Datasets, software, websites, videos and other outputs (Products)
12. Patents

=> Have the foundations in place before starting with the roof! <=

Make sure you only expose public information

Do not show:

- Project proposals
- Contractual research projects
- Metadata of internal documents and non-disclosable outputs
- Persons who exercised their right to be forgotten
 - or some other opt-out

=> Do not show what you wouldn't put on your institution's public website! <=

Make your data understandable from outside

Only show what is relevant for RTD:

- No support staff
- No support units

Start with a year you're having

- consistent &
- reasonably complete information for

=> You will need to describe your contents! <=



*

The Road Ahead

1. More CRISs with OpenAIRE CERIF XML endpoint
2. The OpenAIRE Pipeline: Validator → Aggregator → Portal
3. Updates of the Guidelines
4. The DRIS as the registry of CRISs

* Landscape near Paracas, Peru. Photo CC BY 4.0 Jan Dvořák.

More CRISs with OpenAIRE CERIF-XML endpoint

Budget-neutral partners in the METIS2OpenAIRE project:

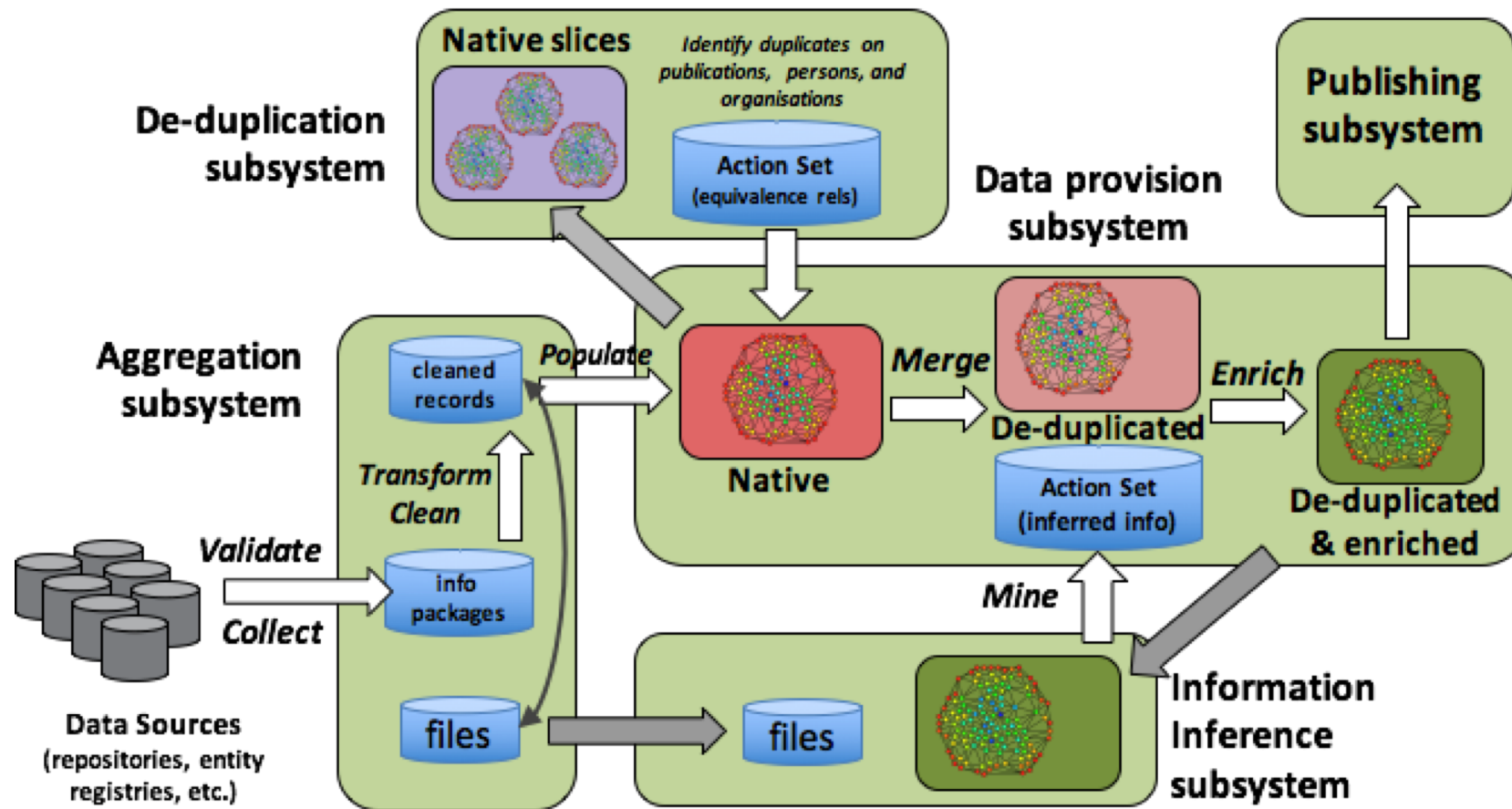
- Omega-PSIR (→ see the Omega-PSIR track tomorrow)
- PURE (→ see the talk tomorrow by Anna Clements)

And:

- VIRTAs (→ see the talk tomorrow by Joonas Nikkanen and Dragan Ivanović)
- NARCIS (in development)
- DSpace-CRIS (planned)

... to be tested with the Prototype Validator

OpenAIRE Aggregation Pipeline



OpenAIRE Final Validator

- To be realized by the OpenAIRE team
- Based on the Prototype Validator
- To make the “handshake” happen
between producers and the consumer of the research information

Update to the Guidelines

Small **extensions** based on feedback:

1. Allow multiple parents for OrgUnits
 - To represent interdisciplinary centres faithfully
2. Add Digital Author Identifiers
 - From NARCIS [NL]

→ v.1.1.1

=> Please tell us about your experience! <=

Support in the DRIS

- DRIS = Directory of Research Information Systems
- Created and maintained by euroCRIS
- Currently undergoing a deep change
- Will track specifically OpenAIRE CERIF XML Guidelines endpoint
 - URL
 - compatibility level

(→ separate talk by PdC in the next session)

=> Your input is welcome! <=

Conclusions

1. METIS harvested
2. Recommendations for implementations
3. Coming:
 - More CRISs
 - OpenAIRE official validator
 - Minor extensions to the standard
 - Support by the DRIS