CRIS in OpenAIRE

we take you on-board

Aenne Loehden
Jochen Schirrwagen
Andreas Czerniak
Bielefeld University

euroCRIS member meeting | Helsinki / FI | 27-29 MAY 2019

OpenAIRE

European Commission

CC BY

@openaire_eu

Agenda

- Introduction
- Integration of CRIS's
- Content curation
- Benefits
- OpenAIRE legal entity and euroCRIS

Introduction

CC BY

What is OpenAIRE (in short)

- OpenAIRE => **Open A**ccess **I**nfrastructure for **R**esearch in **E**urope.

- European Commission funded since 2009 and has ~55 partner institutions.
- Promote and provide support for Open Access and Open Science across the European countries and has:
  - 34 National open access desks (NOADs) in every European country to support the national open access infrastructure.
  - Services for researchers, funders, projects, ...
- Collaborate with euroCRIS, COAR, EOSC-hub, RDA Europe, and many other organisations/initiatives.

OpenAIRE Guidelines for CRIS Managers

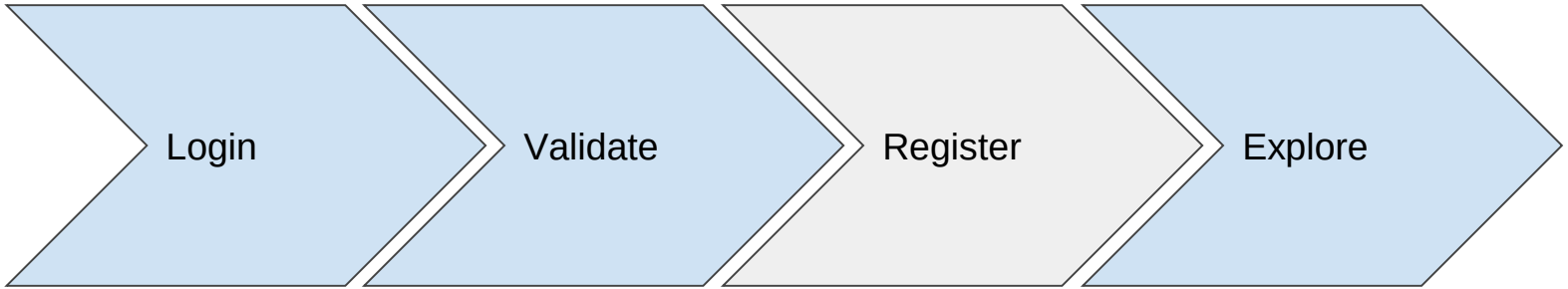
- Evolution of the standard:
  - v.1.0 (June 2015), https://doi.org/10.5281/zenodo.17065
    - Old-style CERIF-XML, a 1:1 mapping from the CERIF data model
  - **v.1.1.0** (June 2018), https://doi.org/10.5281/zenodo.1298650
    - Updated CERIF-XML
    - Vocabularies aligned with the other OpenAIRE Guidelines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

Integration of CRIS systems

we take you on-board

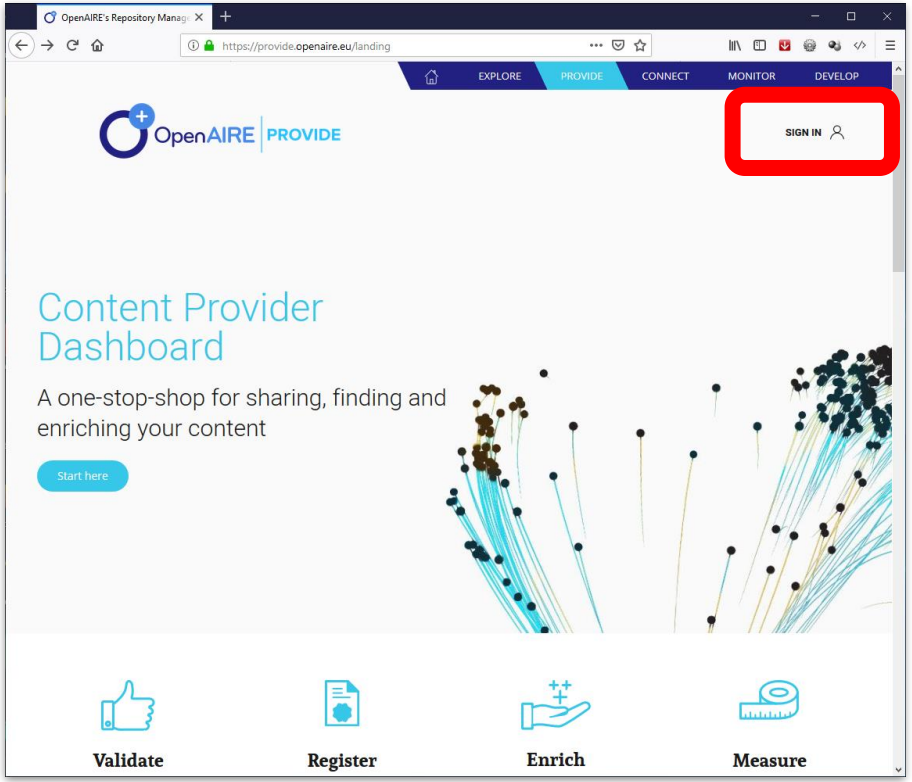
CC BY

Integration of CRIS's - procedures



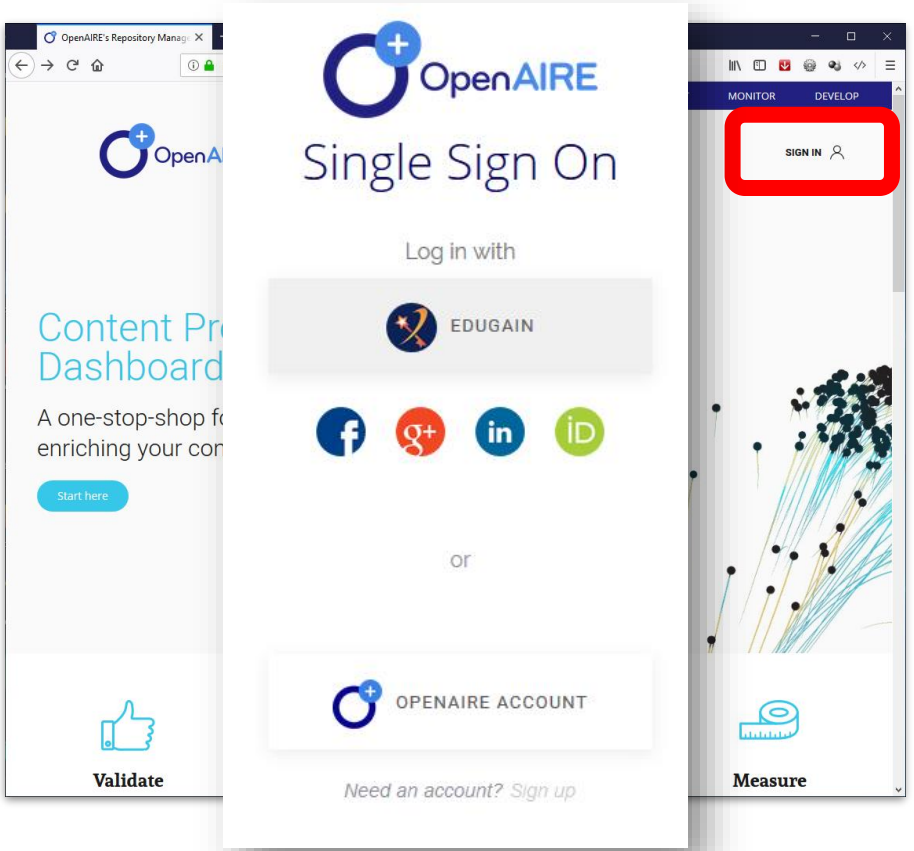
Integration of CRIS's - login procedure

- Login



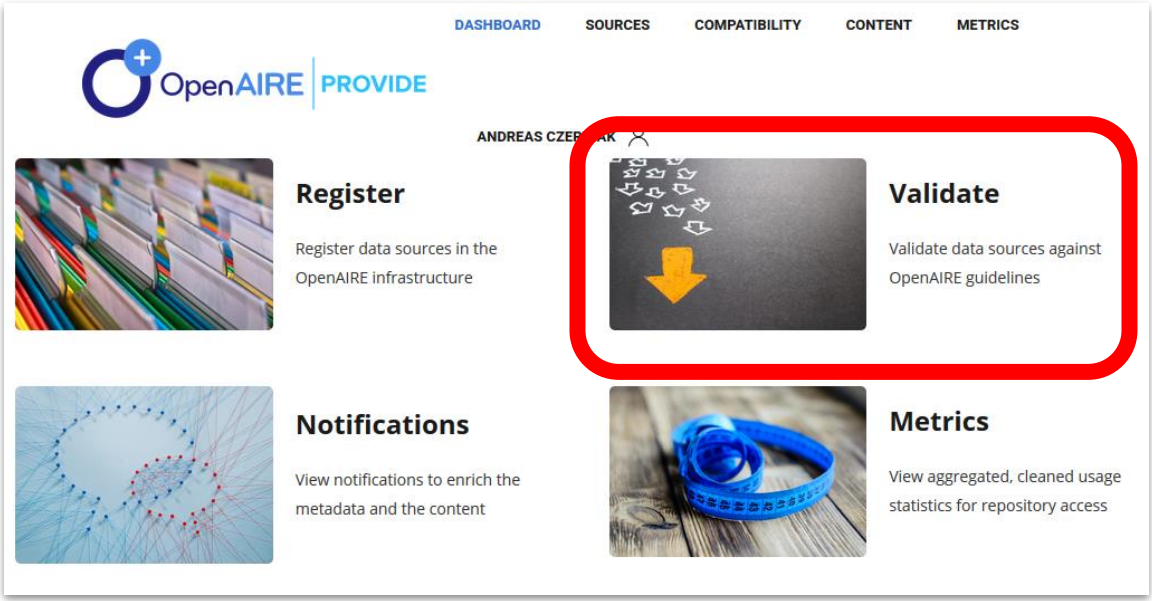
Integration of CRIS's - login procedure

- Login
  - single sign on



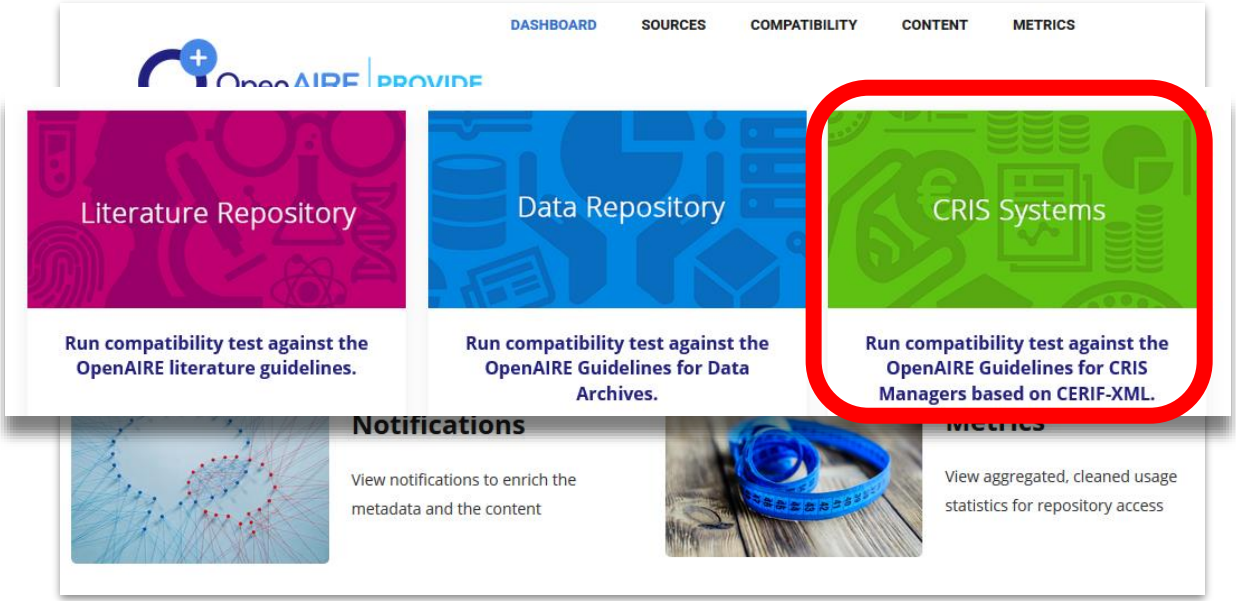
Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"



Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"



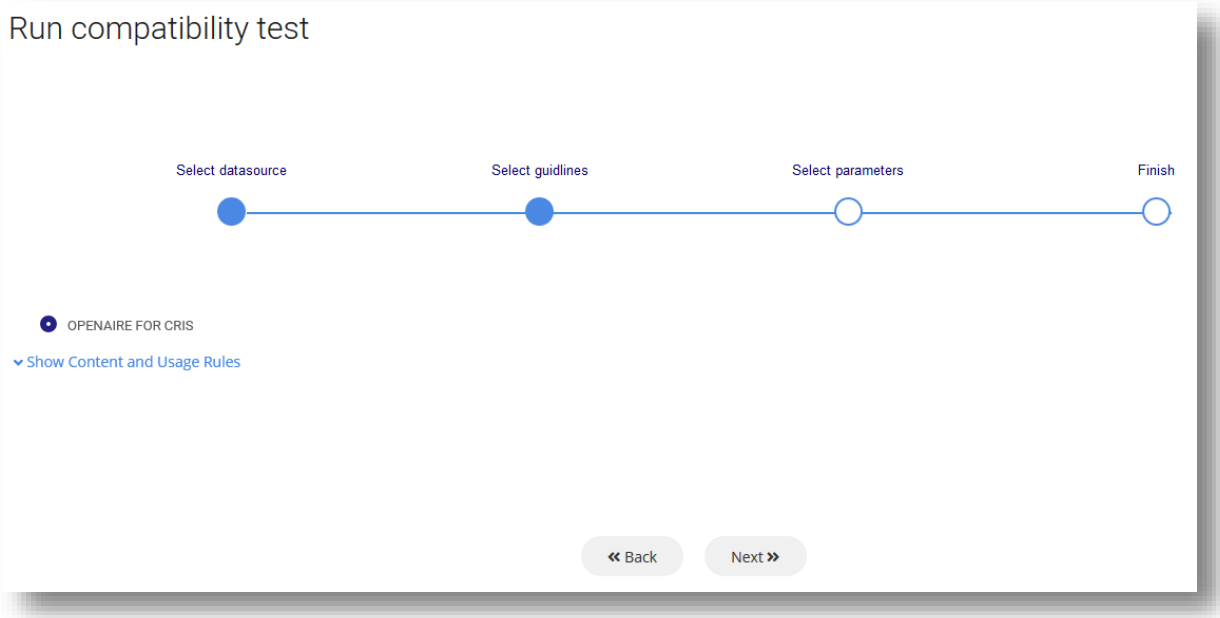
Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"
- enter your OAI-PMH endpoint



Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"
- enter your OAI-PMH endpoint



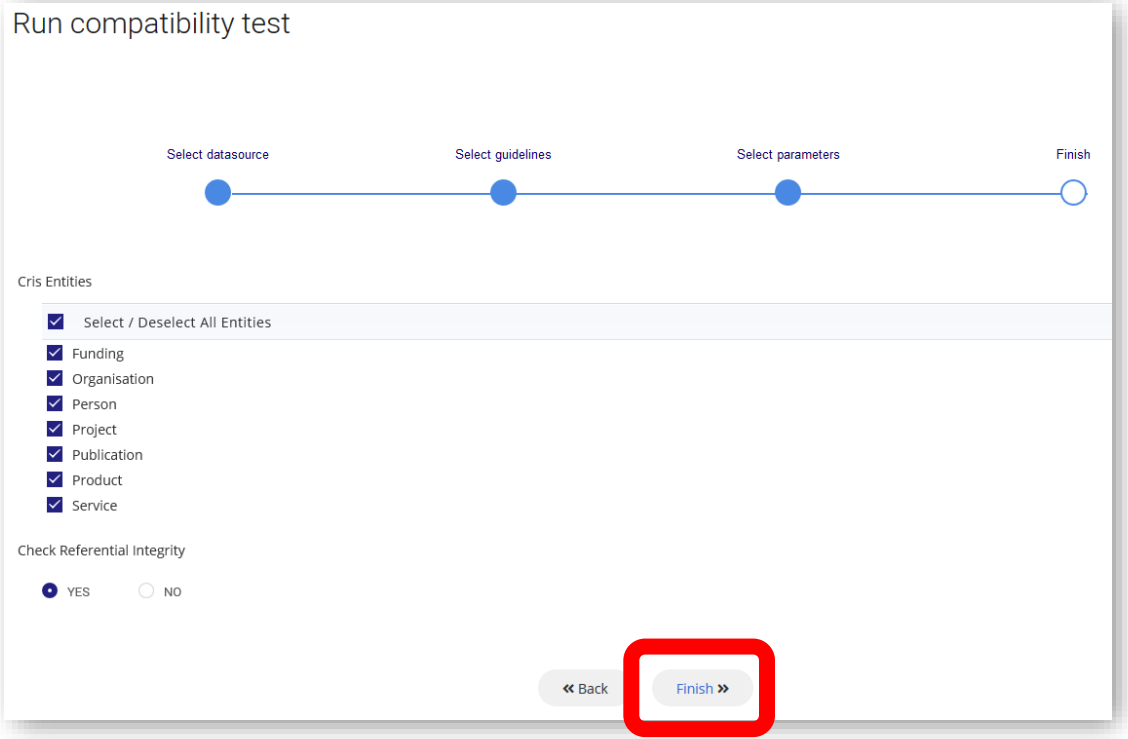
Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"
- enter your OAI-PMH endpoint
- choose the "Information elements"



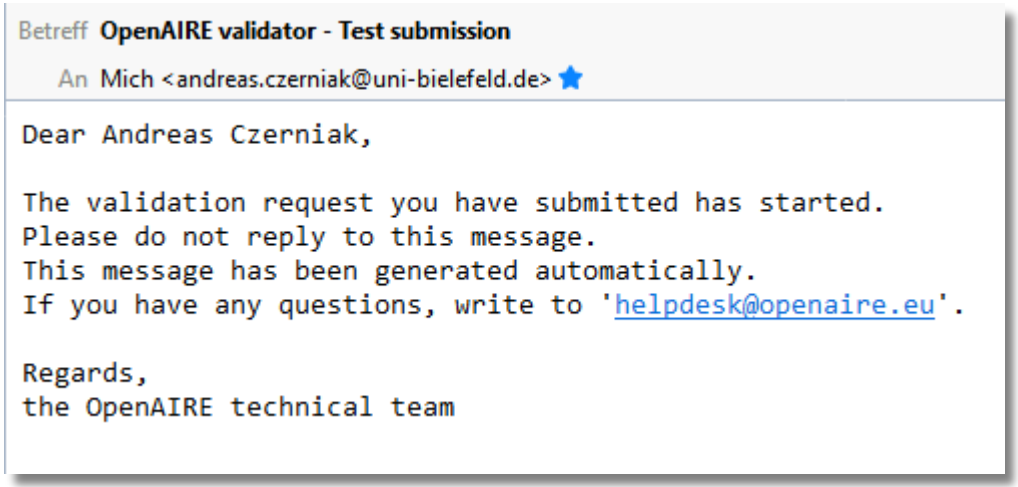
Integration of CRIS's - validation procedure

- login
  - single sign on
- choose "Validate"
- choose "CRIS Systems"
- enter your OAI-PMH endpoint
- choose the "Information elements"



Integration of CRIS's - validation procedure

- Validation email



Betreff **OpenAIRE validator - Test submission**

An Mich <andreas.czerniak@uni-bielefeld.de>

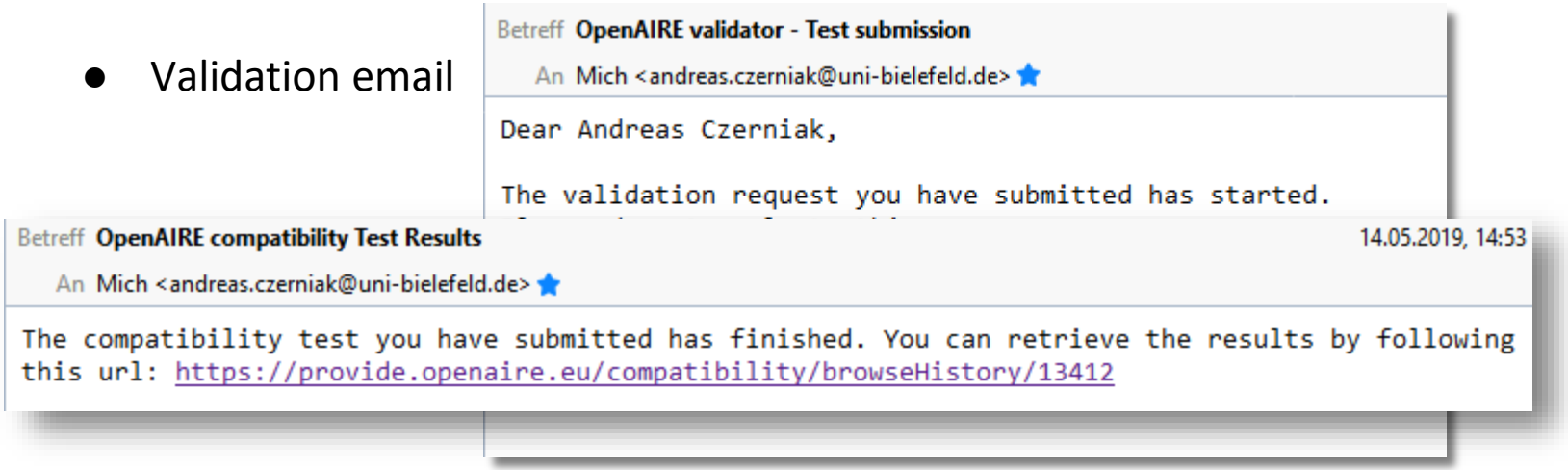
Dear Andreas Czerniak,

The validation request you have submitted has started.
Please do not reply to this message.
This message has been generated automatically.
If you have any questions, write to 'helpdesk@openaire.eu'.

Regards,
the OpenAIRE technical team

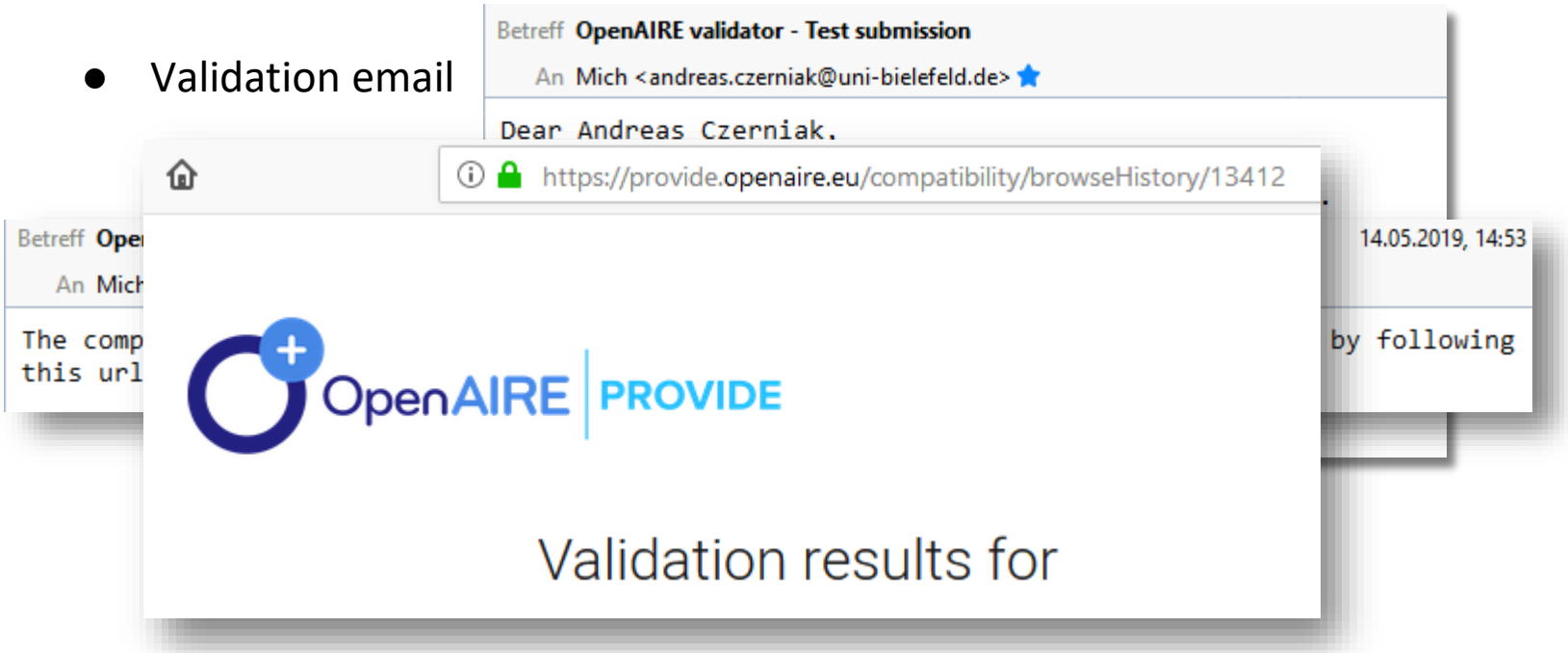
Integration of CRIS's - validation procedure

- Validation email



Integration of CRIS's - validation procedure

- Validation email



Integration of CRIS's - registration procedure

- Registration

  will be first available on our beta infrastructure soon

Integration of CRIS's - registration procedure

- Registration

  will be first available on our beta infrastructure soon



Integration of CRIS's - registration procedure

- Registration

  will be first available on our beta infrastructure soon

  contact us https://www.openaire.eu/support/helpdesk if you are interested in becoming a part of it.



Metadata curation + examples

CC BY

Purposes of research item metadata

Metadata serves to discover and fathom research items:

- What is the described research item about?
- Where does it stem from?
- What is the context of the item?
- How may it be obtained and used?

Metadata quality aspects

- Timeliness: Metadata is up-to-date, i.e. records reflect the current state incl. recent changes.
- Completeness: All proper, suitable, relevant information is leveraged.
- Accuracy: Information is veritable, correct, non-contradictory. It is clear to which items statements refer (e.g. access rights or PIDs may refer to the metadata record itself, to the described item, to a cited item). Statements are sound, i.e. PIDs are valid.
- Legibility: Descriptions are easily comprehensible; due to language (e.g. titles), structure (of e.g. thesauri/vocabulary), homogeneity (e.g. order of author names).
- Consolidation: Records and statements are non-redundant, i.e. without duplicates. Records are contextualised, i.e. linked with other records (citations/references, versions/variants).
- Wastelessness: No test records, no records out of scope, no predatory sources, no spam, etc.
- Format-conformance: Compliance with format standards, utilisation of vocabularies/thesauri.

Diagnostic findings of CRIS metadata

- Abstracts, subjects often not given, subjects not based on thesauri.
- Language attribution missing, undeclared, or wrong for titles, abstracts, subjects.
- Resource types frequently too unspecific (other or text).
- Reviews lacking explicit specification of reviewed item (as prose in title).

What is the described item about?

- Author names quite differently formatted and sometimes lacking IDs.
- Container publication (journal, book) without IDs (ISSNs, ISBNs).

Where does it stem from?

Diagnostic findings of CRIS metadata

- Project funding hints generally without funder names, let alone IDs.
- Citations generally missing.
- Dates sometimes missing or incomplete, version hints often missing or unintelligible.

What is the item's context?

- Access rights mostly not stated, licenses seldom indicated (or as prose in abstract).
- Landing page links repeatedly missing, incomplete, defective.
- Persistent identifiers (DOIs, Handles, ...) sometimes not resolvable.

How may it be accessed and used?

Potential causes of metadata weaknesses

Background of metadata issues relate partly to utilized software ...

- Faulty mapping from CRIS database to exposition layer (e.g. OAI-PMH interface).
- Import from other data sources with flawed metadata or flawed mapping.
- Imperfect automatic extraction from full texts (OCR).
- Platforms creating PIDs without registering them.

Potential causes of metadata weaknesses

… partly more to intellectual metadata curation ...

- Metadata input by different (non-librarian) parties.
- Non-availability or unclarity of information (e.g. project numbers, funding streams).
- Disregarded default values.
- Misuse of metadata fields.

… or to just general issues

- Information getting outdated (e.g. URLs).

Measures to improve metadata quality

Metadata quality can be fostered by platform administrators ...

- Check mappings (in and out).
- Obtain platform add-ons (e.g. for data provision, plausibility checks, PID and URL tests).
- Consult clean external sources.
- Adhere to standards.
- Apply controlled vocabularies and thesauri.

Measures to improve metadata quality

.... as well as by metadata curators ...

- Be precise and specific.
- Beware of defaults.
- Free information from implicit prose to explicit specification.
- State (persistent) identifiers wherever available.
- Contextualize.

➔ Provide metadata trainings and guidelines for authors.

Measures to improve metadata quality

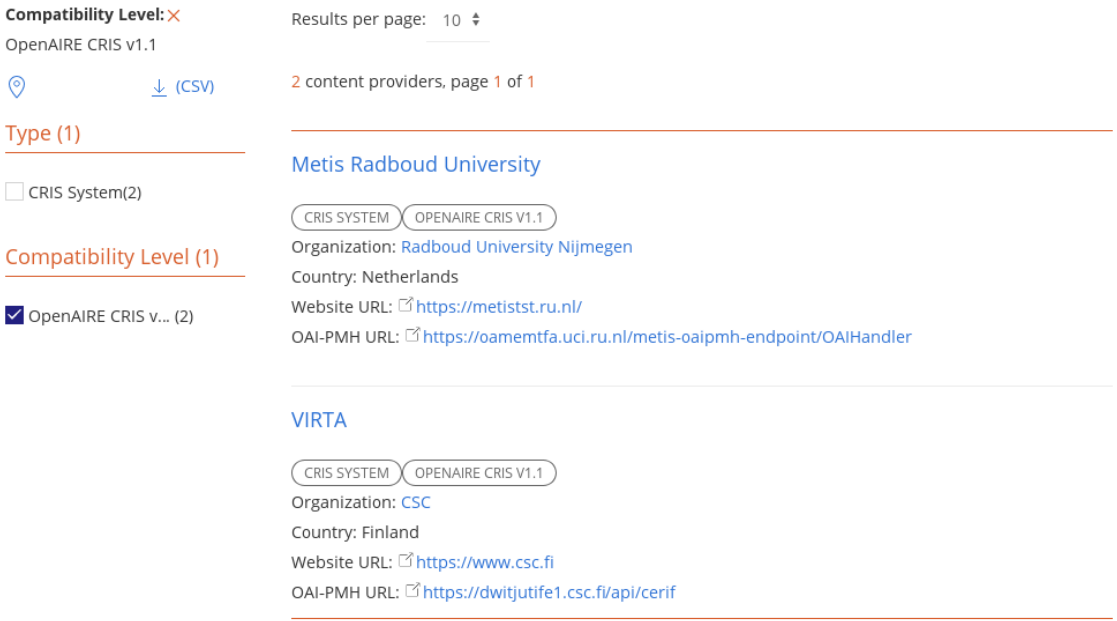
… and by OpenAIRE (homogenise, enrich, filter native metadata)

<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)

Data Sources kindly contributing to OpenAIRE's CRIS integration

- CRIS UNS (Current Research Inform. Syst. Univ. of Novi Sad) (Serbia, in-house softw.)
- DANS (Data Archiving and Networked Services) (Netherlands, in-house software)
- Metis Radboud University (Netherlands, in-house software (Metis))
- Pure-Elsevier (Netherlands, Pure)
- VIRTA (Finland, in-house software)
- WUT Base of Knowledge (Warsaw University of Technology) (Poland, OMEGA-PSIR)

CRIS data sources on OpenAIRE's portal



… more to appear after next indexings!

https://beta.explore.openaire.eu/search/find/dataproviders?datasourcecompatibilityname="OpenAIRE CRIS v1.1"

Benefits

De-duplication

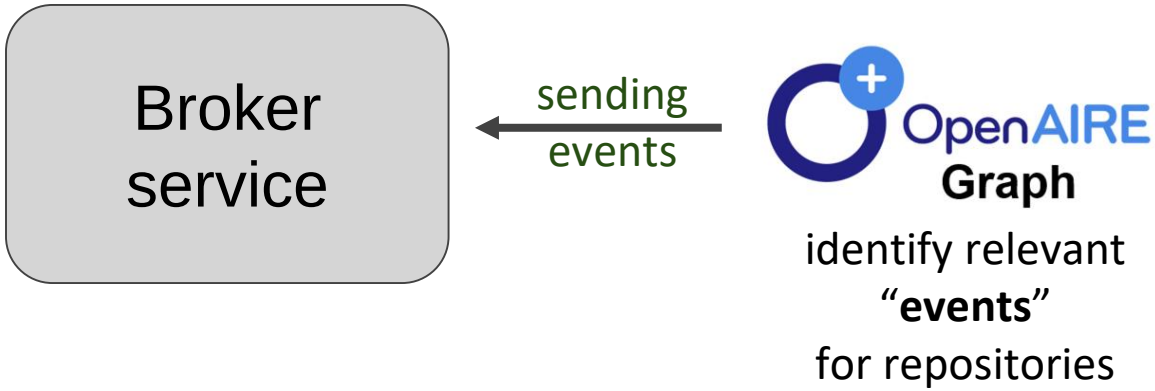
- Publications
  - Products with “equivalent” PIDs, titles, authors, dates are grouped
- Dataset
  - Products with “equivalent” PIDs are grouped
- Software
  - Products with “equivalent” PIDs and original URLs are grouped
- Other products
  - Products with “equivalent” PIDs, titles, authors, dates are grouped

Broker sketch - Information channel

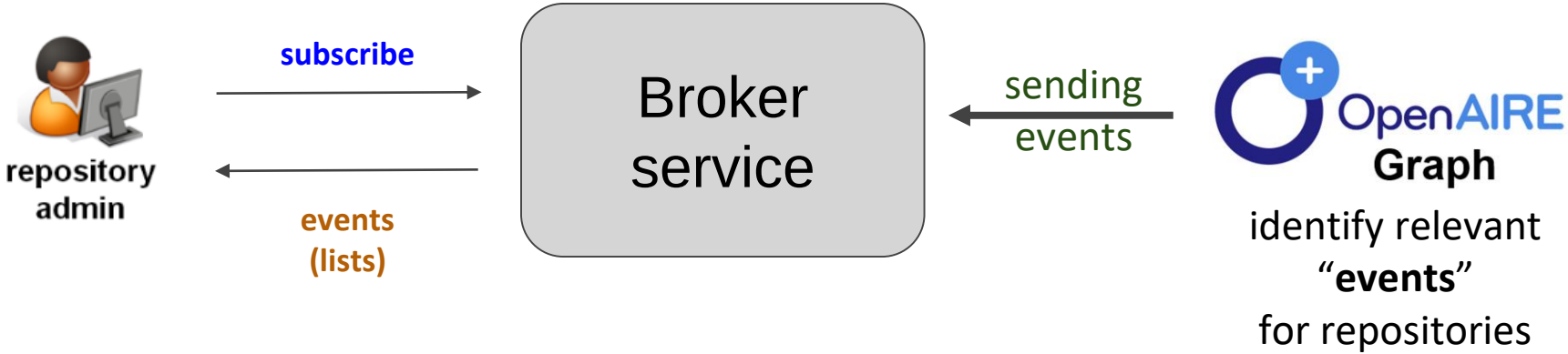


identify relevant
“**events**”
for repositories

Broker sketch - Information channel



Broker sketch - Information channel



Broker - Information channel for CRIS managers

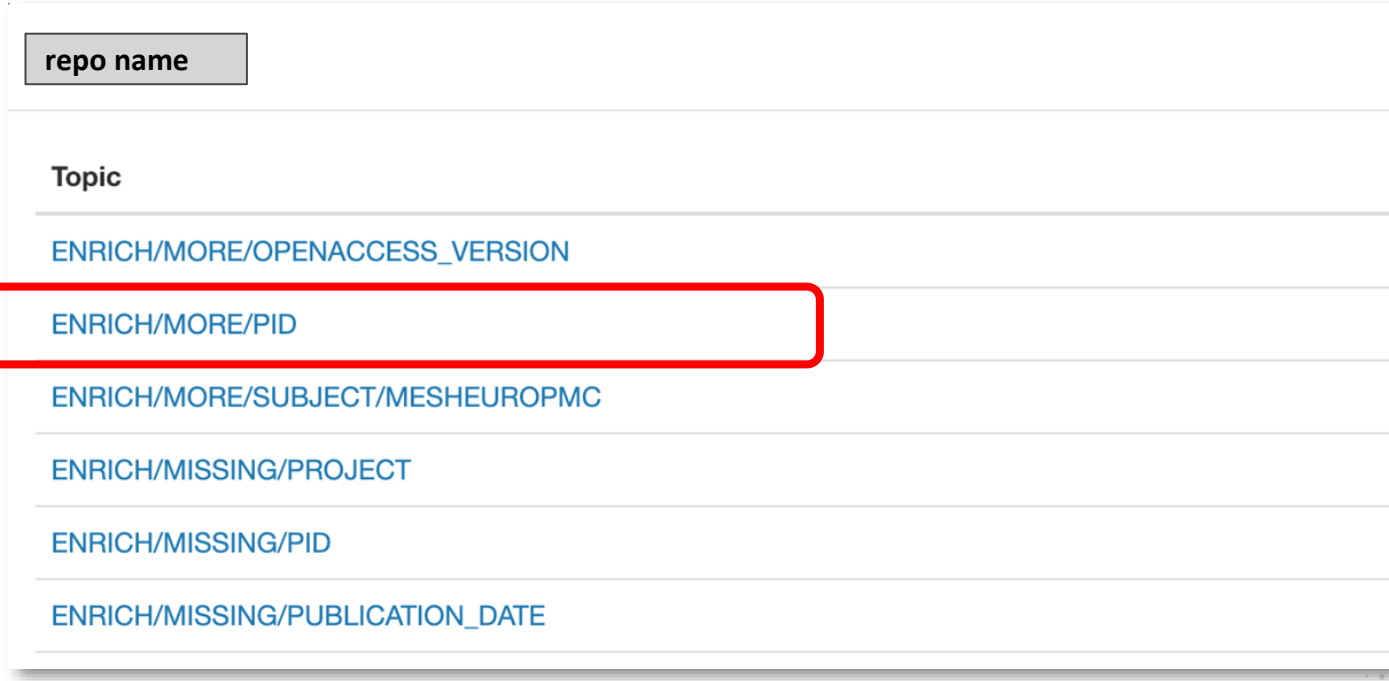
- subscribe to different "**events**"

repo name

Topic	#
ENRICH/MORE/OPENACCESS_VERSION	2323
ENRICH/MORE/PID	600
ENRICH/MORE/SUBJECT/MESHEUROPMC	213
ENRICH/MISSING/PROJECT	205
ENRICH/MISSING/PID	165
ENRICH/MISSING/PUBLICATION_DATE	165
ENRICH/MISSING/SUBJECT/MESHEUROPMC	105
ENRICH/MORE/SUBJECT/JEL	55
ENRICH/MORE/SUBJECT/ARXIV	40
ENRICH/MISSING/OPENACCESS_VERSION	35
ENRICH/MORE/SUBJECT/ACM	35
ENRICH/MORE/SUBJECT/DDC	25
ENRICH/MISSING/SUBJECT/ACM	17
ENRICH/MISSING/SUBJECT/DDC	12
ENRICH/MISSING/SUBJECT/ARXIV	10
ENRICH/MISSING/ABSTRACT	4
ENRICH/MISSING/SUBJECT/JEL	3
Total	**4012**

Broker - Information channel for CRIS managers

- subscribe to different "**events**"



Broker - Information channel for CRIS managers



Trust: 1

ID	oai:drops-oai.dagstuhl.de:6427
Title(s)	Families of DFAs as Acceptors of omega-Regular Languages
Author(s)	Angluin, Dana | Boker, Udi | Fisman, Dana
PID(s)	10.23638/LMCS-14(1:15)2018 **(doi)**
Abstract(s)	Families of DFAs (FDFAs) provide an alternative formalism for recognizing omega-regular languages. The motivation for introducing them was a desired correlation between the automaton states and right congruence relations, in a manner similar to the Myhill-Nerode theorem for regular languages. This correlation is beneficial for learning algorithms, and indeed it was recently shown that omega-regular languages can be learned from membership and equivalence queries, using FDFAs as the acceptors. In this paper, we look into the question of how suitable FDFAs are for defining omega-regular languages. Specifically, we look into the complexity of performing Boolean operations, such as complementation and intersection, on FDFAs, the complexity of solving decision problems, such as emptiness and language containment, and the succinctness of FDFAs compared to standard deterministic and nondeterministic omega-automata. We show that FDFAs enjoy the benefits of deterministic automata with respect to Boolean operations and decision problems. Namely, they can all be performed in nondeterministic logarithmic space. We provide polynomial translations of deterministic Buchi and coBuchi automata to FDFAs and of FDFAs

beyond OpenAIRE-Advance

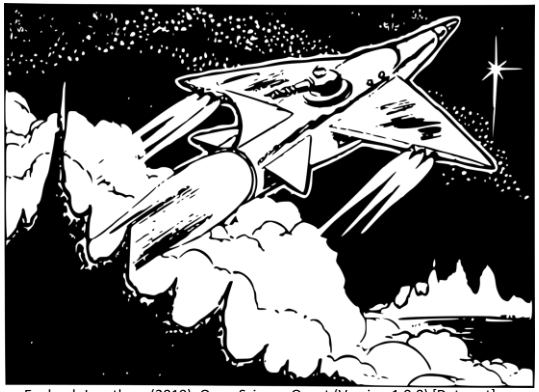
CC BY

beyond OpenAIRE-Advance

- OpenAIRE collaborates with EOSC-hub
  - key infrastructures of the EOSC
    - service integration
    - communication, engagement, support and training
    - governance and strategy
- OpenAIRE A.M.K.E. established as non-profit legal entity in 09/2018
  - making the Open Science human network and services infrastructure sustainable
  - four founding members, NOADs are joining, open for partners
- Memorandum of Understanding between euroCRIS and OpenAIRE A.M.K.E

together

- we
  - take you onboard



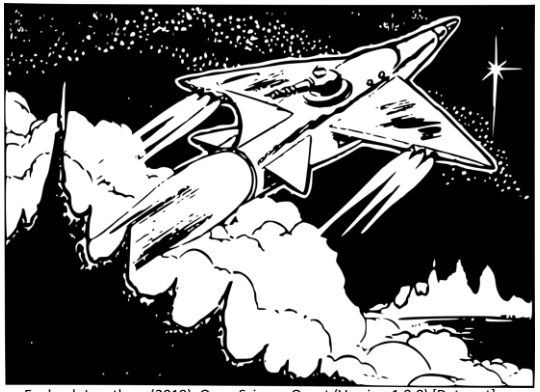
England, Jonathan. (2019). Open Science Quest (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2646121

together

- we
  - take you onboard

    and

  - lift up your treasure



England, Jonathan. (2019). Open Science Quest (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.2646121



CC0: https://pixabay.com/de/illustrations/schatz-m%C3%BCnzen-golden-3176785/

Thank you!

Aenne Loehden
aenne.loehden@uni-bielefeld.de

Andreas Czerniak
andreas.czerniak@uni-bielefeld.de orcid.org/0000-0003-3883-4169

https://www.openaire.eu

CC BY