

COAR Interest Group

“Controlled Vocabularies for Repository Assets”

Introductory Webcast

14 January 2014

Jochen Schirrwagen | jochen.schirrwagen@uni-bielefeld.de



Agenda

- Introduction
- Organization & Objectives
- The “info:eu-repo” vocabulary
- Challenges
- Relation to other initiatives
- Communication Channels / Activities
- Discussion / AOB



Purpose

- This group will provide a forum about **controlled vocabularies** to describe Open Access **scientific results** in repositories and eJournal platforms.
- It has a narrower focus on the “**info:eu-repo**” **vocabulary** aiming to create an **authority group** and to set up a workflow for its moderation and maintenance.
- It has also a broader focus by inviting and liaising with other initiatives that aim for **standardization and communication** of good practices in **describing scientific results and research information**.



Organization of the Interest Group

- Part of the COAR Working Group 2 „Repository Interoperability“ (Chair: Eloy Rodrigues)
- Interest Group Chair: Jochen Schirrwagen
- Authority Group for “info:eu-repo”:
 - By interested members of this group
 - E.g. members of initiatives using the vocabulary
- Interested members and experts willing to exchange knowledge and good practices about metadata, vocabularies, information management



The *info:eu-repo* vocabulary

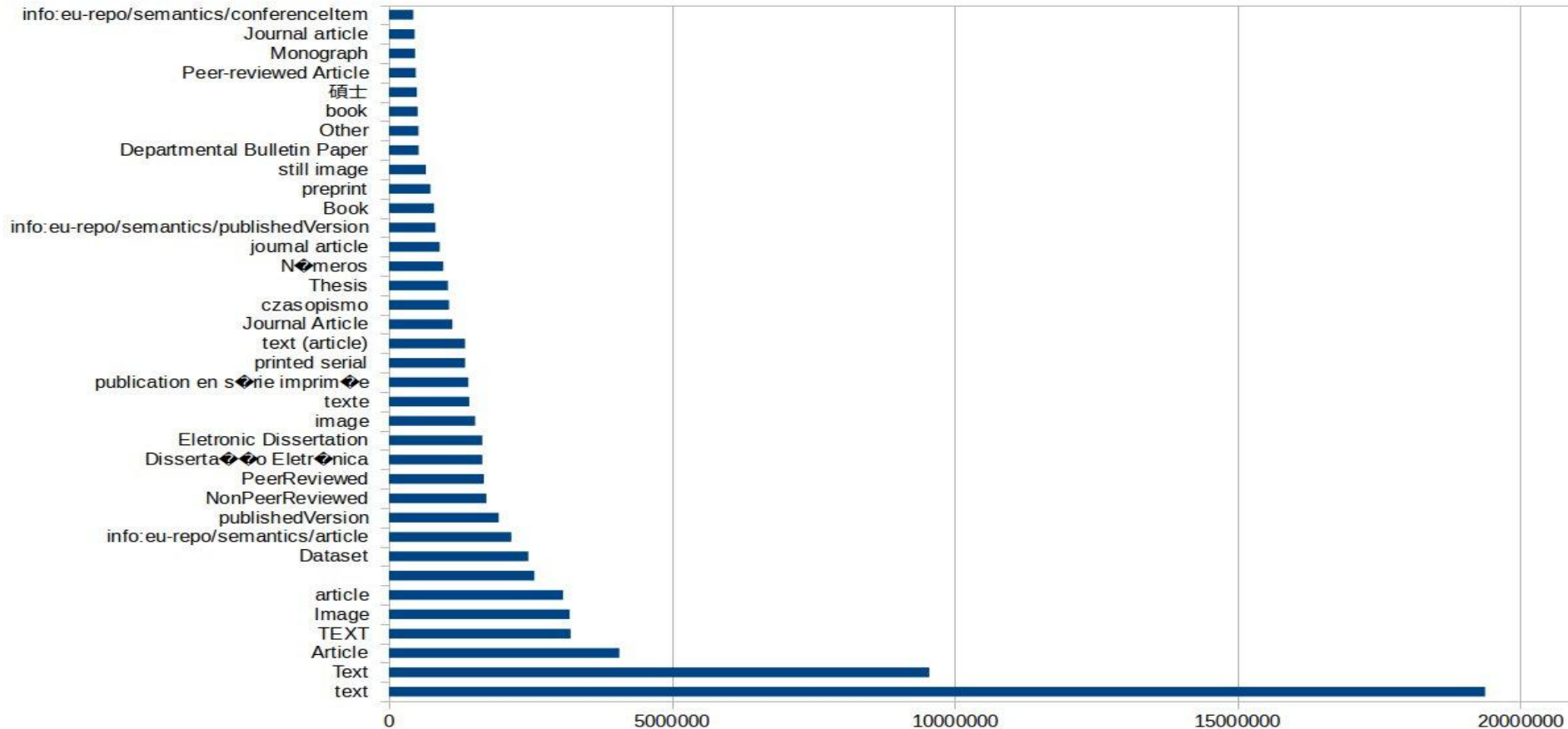
- **Construction**
 - a registered namespace “info:eu-repo”
 - Serving as a placeholder for semantic , controlled terms and identifiers
- **Functionality**
 - URI’s – persistent identifiers for non-informational resources
 - XML Schema – used for validation
- **Coverage**
 - Publication Type Vocabulary (xml-schema)
 - Version Type Vocabulary (xml-schema)
 - Access Mode
 - Identifier
 - Classification Schemes
 - Date Types
 - Versions
 - Object Types



Some Statistics (based on top36 publication types)



rate of publication types



(Statistics based on BASE harvesting (~3000 datasources, ~58 mio records kindly provided by Friedrich Summann, base-search.net)



Statistics of info:eu-repo terms (publication type top25)

value	rate	datasources
info:eu-repo/semantics/article	2163633	466
info:eu-repo/semantics/publishedVersion	822329	392
info:eu-repo/semantics/conferenceltem	430319	14
info:eu-repo/semantics/bookPart	333725	106
info:eu-repo/semantics/doctoralThesis	225874	142
info:eu-repo/semantics/bachelorThesis	197812	80
info:eu-repo/semantics/report	173628	92
info:eu-repo/semantics/book	169356	121
info:eu-repo/semantics/conferenceObject	102988	87
info:eu-repo/semantics/masterThesis	93019	90
info:eu-repo/semantics/other	70967	97
info:eu-repo/semantics/contributionToPeriodical	45937	30
info:eu-repo/semantics/paper	43329	10
info:eu-repo/semantics/conferencePaper	39454	9
info:eu-repo/semantics/lecture	34211	45
info:eu-repo/semantics/review	28261	58
info:eu-repo/semantics/workingPaper	24858	42
info:eu-repo/semantics/conferenceContribution	20343	2
info:eu-repo/semantics/conferencePoster	11258	5
info:eu-repo/semantics/acceptedVersion	8604	22
info:eu-repo/semantics/submittedVersion	8363	19
info:eu-repo/semantics/Article	4512	2
info:eu-repo/semantics/patent	4258	25
info:eu-repo/semantics/image	4214	3



Implementation by Guidelines (sample)

- DRIVER-Guidelines v2 (will be superseded by the OpenAIRE Guidelines)
 - http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf
- OpenAIRE Guidelines series
 - https://guidelines.openaire.eu/wiki/Main_Page
- NEEO Application Profile
 - <http://homepages.ulb.ac.be/~bpauwels/NEEO/WP5/WP5%20Technical%20guidelines.pdf>
- Directrices Sistema Nacional de Repositorios Digitales (Argentina)
 - http://repositorios.mincyt.gob.ar/pdfs/Directrices_SNRD_2013.pdf
- “Gemeinsames Vokabular für Publikations- und Dokumenttypen” (Common Vocabulary for publication and document types) (DINI, Germany)
 - <http://edoc.hu-berlin.de/series/dini-schriften/12/PDF/12.pdf>



Some Challenges

- “individual extensions” by repository managers which are not officially registered and supported
 - E.g. “info:eu-repo/UGOV/Doctoral Thesis”
- “info:eu-repo” had its initial focus on European repositories, but we want to promote its global use by considering regional requirements
- The “info” URI registry (<http://info-uri.info>) discontinues for new namespaces
 - Used for URI identification
 - Supersede by new principles, e.g. Linked Data
 - **Need to migrate the vocabulary terms and concepts towards new web practices**
- SURF-NL has been the initial and reliable host and moderator
 - But COAR represents the OA repository community in a global way
 - COAR could improve the vocabulary and its dissemination
 - COAR has experts and links to related initiatives



Activities

- Hosting and maintaining the “info:eu-repo” vocabulary; COAR provides the organizational infrastructure and a web platform
- Set up of an COAR authority group of the “info:eu-repo” vocabulary
- Set up of a workflow for members of the authority group to moderate the vocabulary (e. g. decision making about adding / changing terms, identifiers, descriptions)
- Maintaining a knowledge space (wiki) about metadata guidelines and vocabularies related to the Open Access community
- Review and monitor new and ongoing developments regarding controlled vocabularies and Open Access indicators and provide feedback to the COAR WG “Repository Interoperability” and the COAR Board



Vocabulary Authority Group and Workflow

- Pre-phase:
 - Invitation of interested members
 - What is the preferred workflow to add / change identifiers, terms, description ?
 - Decision making
- Realisation-phase:
 - Provision of guidance for initiatives using the vocabulary
 - Maintenance of the vocabulary in the wiki
 - Ensure compliancy with Linked Data principles
 - ...

Are you
interested?



Related Initiatives

- AIMS-FAO (Agricultural Information Management Standards)
 - Communities of practice in information management
 - AGROVOC thesaurus
 - VEST – catalog og controlled vocabularies
 - LOD-DB – Linked Open Data enabled Bibliographic Data
- CASRAI
(Consortia Advancing Standards in Research Administration Information)
 - Forum and standards development about data in the research life-cycle domain (e.g. researchers, institutions, funders)
- V4OA (Vocabularies for Open Access, supported by JISC)
 - Establishment of an on-going process about agreements of use of vocabularies in particular circumstances in the OA context
- RIOXX (UKOLN, JISC)
 - Guidelines and application profile to comply with RCUK policy on Open Access
- NISO Open Access Metadata and Indicators



- Draft on recommended practices for **Open Access Metadata and Indicators**
 - Includes various stakeholder views (publishers, funders, legal experts, scientific communities)
 - Indication of “free to read” content
 - Indication of use / re-use of content by link (URI) to license terms
 - Consideration of embargo periods; may affect change of license terms starting at a specified date
 - Focus on fulltext documents (not supplementary material, components)
- **Questions**
 - Is this a publisher gold OA issue only ? Does it have any impact on green OA in repositories ?
 - NISO WG recommends the extension of existing metadata schemes – are any vocabularies or metadata schemes known that would already cover those fields: <free_to_read>, <license_ref> ?
 - Other questions? Or Answers ;-) ?



Communication Channels

- Web-Site
 - <https://www.coar-repositories.org/activities/repository-interoperability/ig-controlled-vocabularies-for-repository-assets/>
- Wiki
 - Will be integrated
- Email-List
 - <https://listserv.gwdg.de/mailman/listinfo/coar-ig-oa-vocabularies>



Discussion

- Placeholder for any questions, feedback, ideas



It's a wrap!

Thank you for your interest, feedback and participation!

Feel free to join the group, keep up to date:

<https://listserv.gwdg.de/mailman/listinfo/coar-ig-oa-vocabularies>

More information:

<https://www.coar-repositories.org/activities/repository-interoperability/ig-controlled-vocabularies-for-repository-assets/>

Next Meeting: tbc. / before the next
Annual COAR conference May 2014

