https://vsr.informatik.tu-chemnitz.de/research/projects/PIROL/**RDM-LD/**

SEMANTiCS Karlsruhe 2019

Analysis of current RDM applications for the interdisciplinary publication of research data

André Langer, Ellen Bilz and Martin Gaedke

*VSR.Informatik.*TU-Chemnitz.*de*

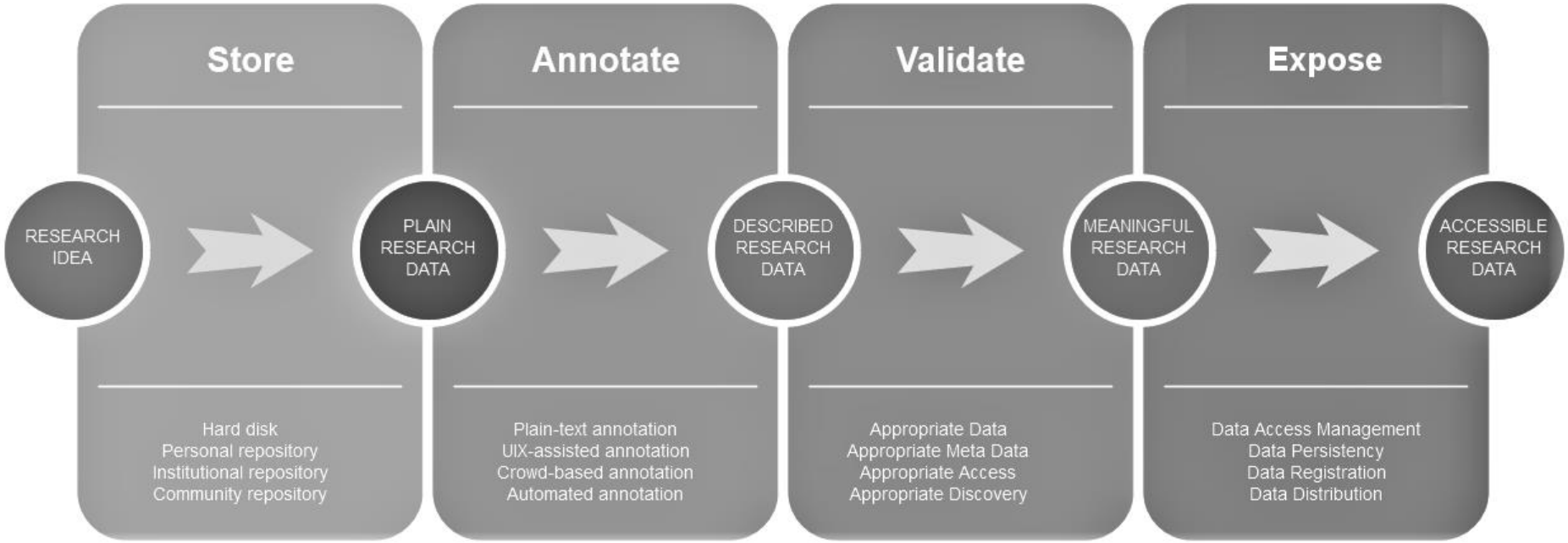
SEMANTiCS Conference Karlsruhe 2019
1st International Workshop on Approaches for Making Data Interoperable (AMAR 2019)

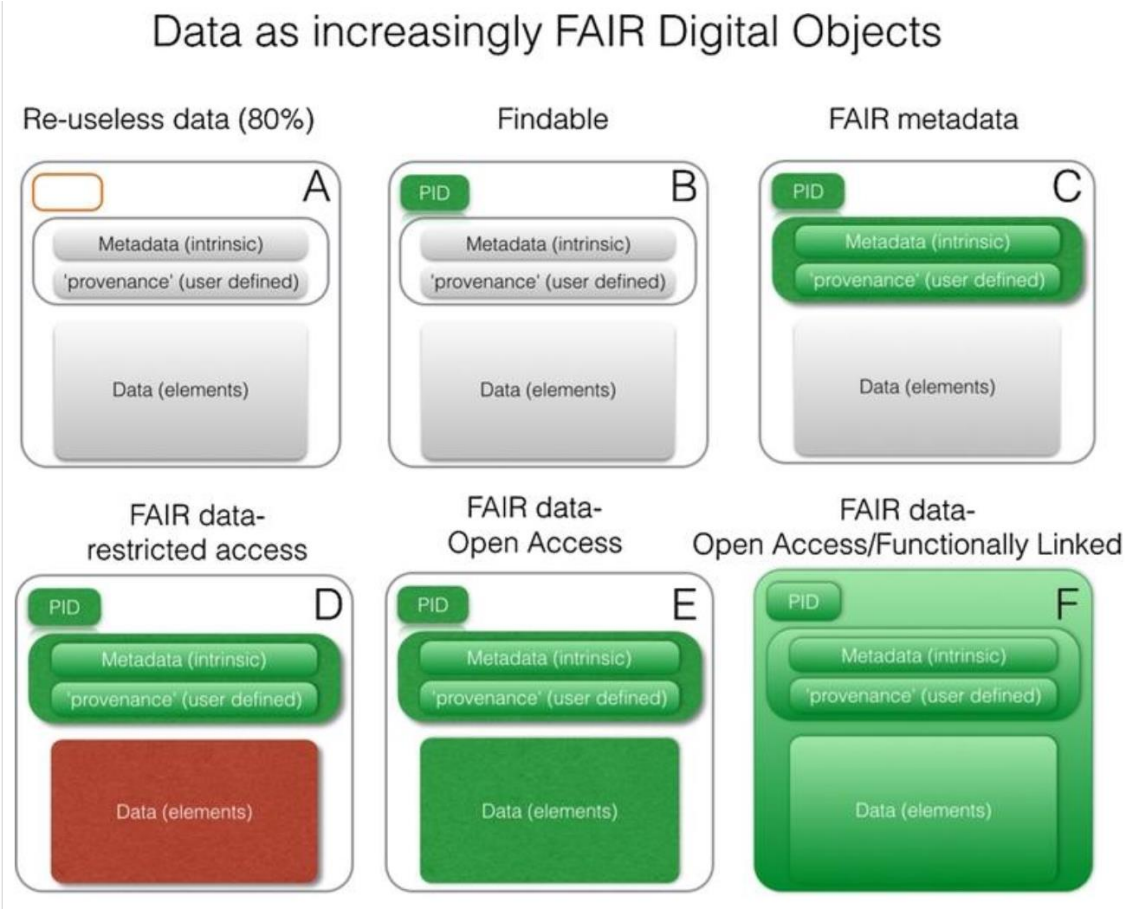
1 Research Data

Research data

„any kind of digital artifact that is associated with scientific research“

[Sousa et. al, 2014]


Store
Annotate
Validate
Expose
RESEARCH IDEA
PLAIN RESEARCH DATA
DESCRIBED RESEARCH DATA
MEANINGFUL RESEARCH DATA
ACCESSIBLE RESEARCH DATA
Hard disk
Personal repository
Institutional repository
Community repository
Plain-text annotation
UIX-assisted annotation
Crowd-based annotation
Automated annotation
Appropriate Data
Appropriate Meta Data
Appropriate Access
Appropriate Discovery
Data Access Management
Data Persistency
Data Registration
Data Distribution



[Mons et. al, 2017]

Research Publications

Research Data

Research Information


2
Problem
Description

Institutions (optional)

Enter the name of the institution

Categories for this data

Enter the name of the category

Description of this data

3000 characters left

Steps to reproduce (optional)

3000 characters left

Subjects optional

Specify subjects from a taxonomy or controlled vocabulary. Each term must be uniquely identified (e.g. a URL). For free form text, use the keywords field in basic information section.

Subjects Term Identifier

Term Identifier

+ Add another subject

Delete Save Publish

3 Research Question

Which relevant platforms for research data publishing are currently used?

To which extent do they support Linked Data based meta data annotations for interdisciplinary reuse?

Assumptions

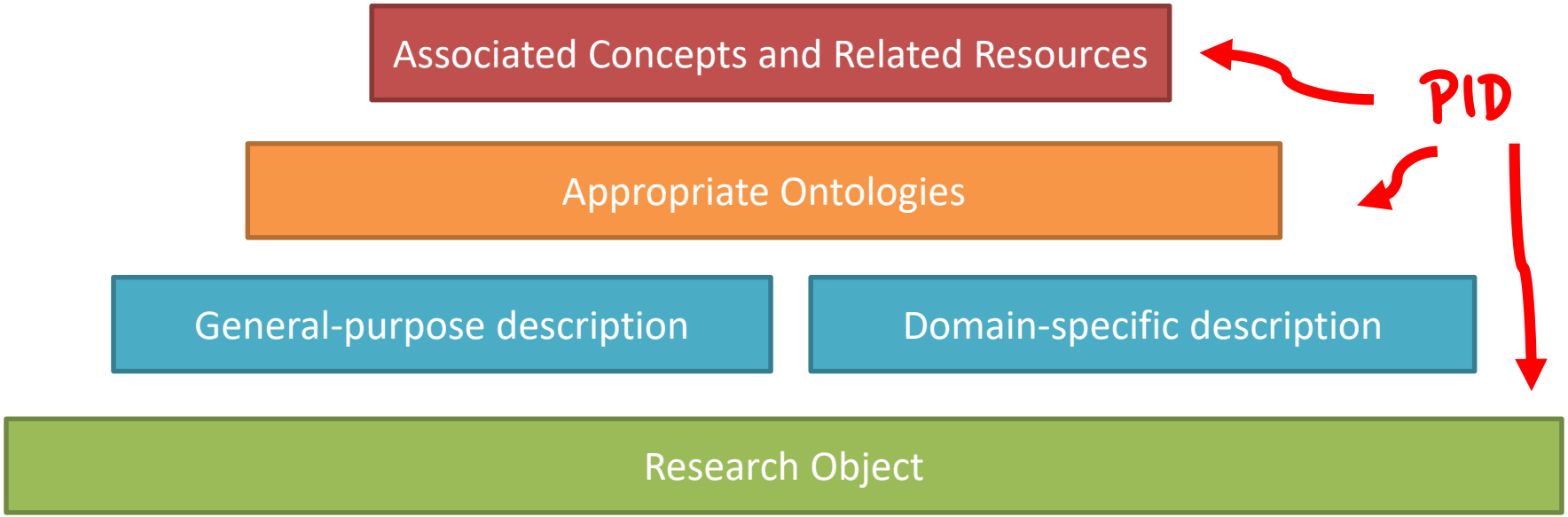
There is an interest in publishing existing research data

Appropriate domain-specific ontologies already exist

Benefits from a structured, unambiguous data description


4
Approach

Interdisciplinary Access to Research Data



In order to identify relevant systems for research data publishing, we ran a systematic mapping on scientific publications of the last ten years

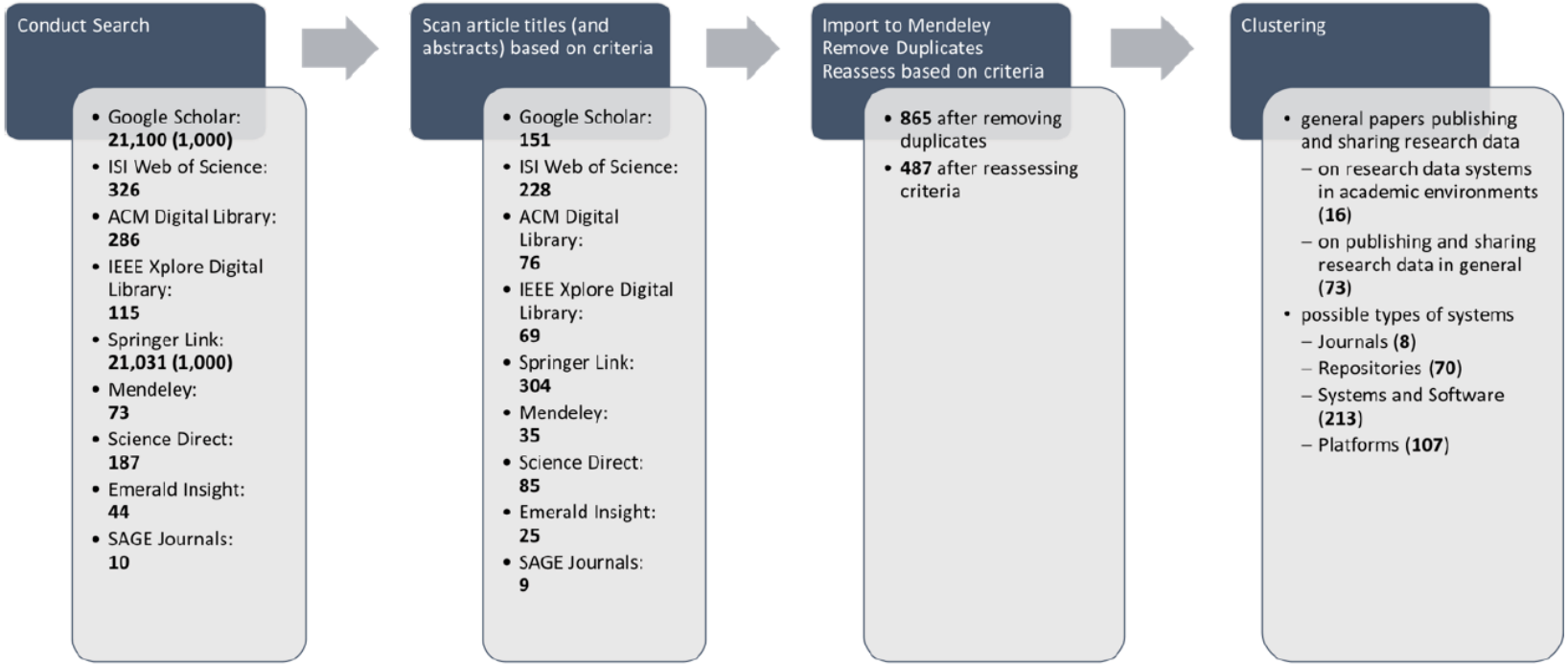
Inclusion criteria

+ dealing with research data management and data publishing or data sharing
+ papers published between 2008 and 2018
+ in English

Exclusion criteria

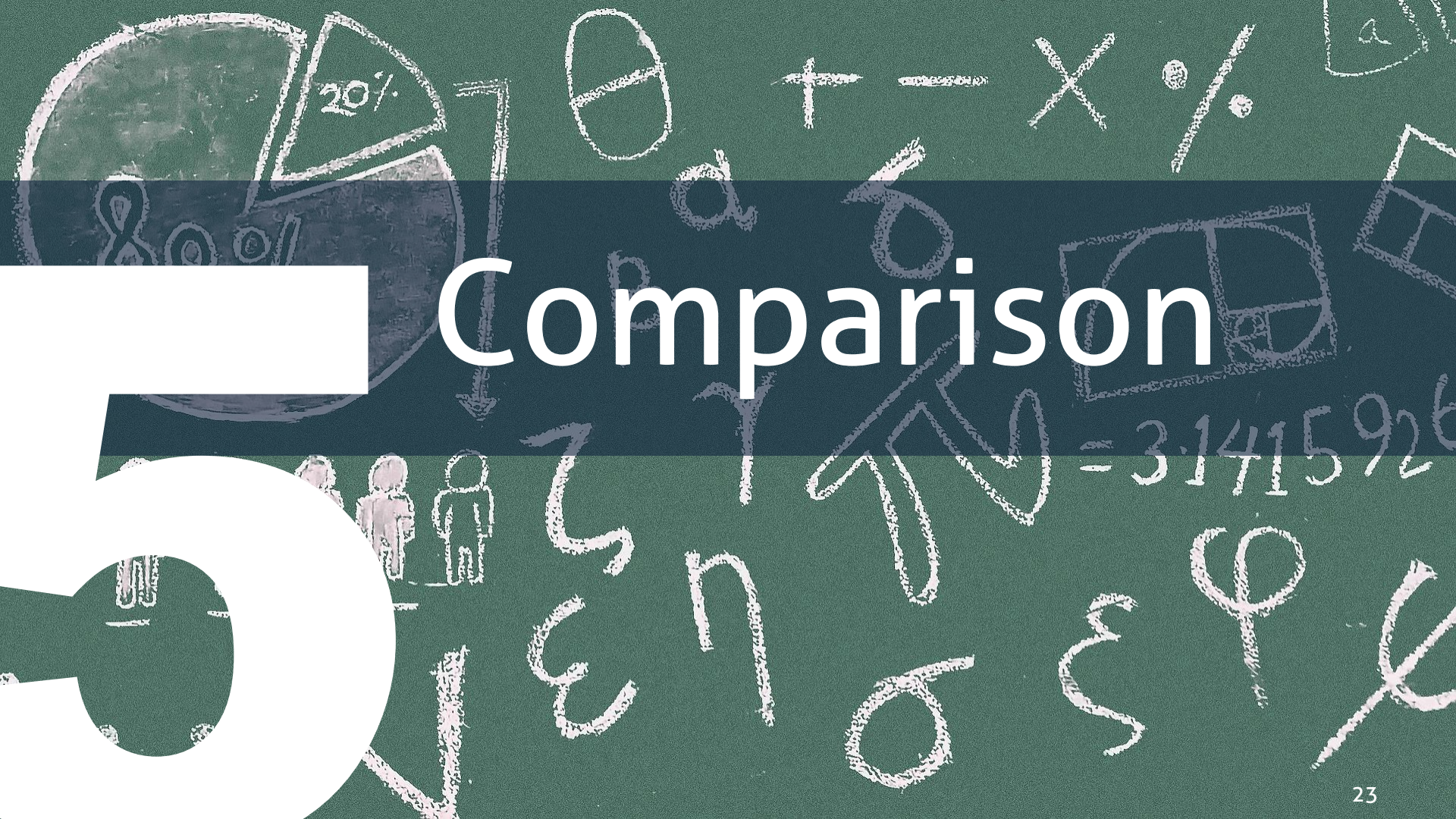
- domain-specific solutions
- different topic focus
- deprecated solutions

Conduction


Conduct Search
• Google Scholar: **21,100 (1,000)**
• ISI Web of Science: **326**
• ACM Digital Library: **286**
• IEEE Xplore Digital Library: **115**
• Springer Link: **21,031 (1,000)**
• Mendeley: **73**
• Science Direct: **187**
• Emerald Insight: **44**
• SAGE Journals: **10**
Scan article titles (and abstracts) based on criteria
• Google Scholar: **151**
• ISI Web of Science: **228**
• ACM Digital Library: **76**
• IEEE Xplore Digital Library: **69**
• Springer Link: **304**
• Mendeley: **35**
• Science Direct: **85**
• Emerald Insight: **25**
• SAGE Journals: **9**
Import to Mendeley
Remove Duplicates
Reassess based on criteria
• **865** after removing duplicates
• **487** after reassessing criteria
Clustering
• general papers publishing and sharing research data
– on research data systems in academic environments **(16)**
– on publishing and sharing research data in general **(73)**
• possible types of systems
– Journals **(8)**
– Repositories **(70)**
– Systems and Software **(213)**
– Platforms **(107)**

Search Results

System	Count	URL
GitHub	101	https://github.com/
Dryad	67	https://datadryad.org/
e!DAL	59	http://edal.ipk-gatersleben.de/
DSpace	51	https://duraspace.org/dspace/
Figshare	44	https://figshare.com/
Fedora	37	https://duraspace.org/fedora/
Eprints	35	https://www.eprints.org/
Dropbox	32	https://www.dropbox.com/
CKAN	27	https://ckan.org/
Dataverse	27	https://dataverse.org/
Zenodo	25	https://www.zenodo.org/
myExperiment	23	https://www.myexperiment.org
Globus	22	https://www.globus.org/
Virtuoso	22	https://virtuoso.openlinksw.com/
B2SHARE / EUDAT B2 Services	21	https://b2share.eudat.eu/
Drupal	20	https://www.drupal.org/
XSEDE	14	https://www.xsede.org/
Dendro	13	http://dendro.fe.up.pt/
D2R	11	http://d2rq.org/d2r-server
HUBzero	10	https://hubzero.org/
Google Drive	9	https://drive.google.com/



Comparison

In order to compare the identified research data publishing systems, we ran a comparative study based on 15 criteria derived from the FAIR principles for data sharing

Findable (LD)

C1 Is a particular research data set in a current version accessible via a unique PID?
C2 Is the research data information through that platform indexed in data catalogs, registries and search engines?
C3 Is a search interface available with filter possibilities for structured Linked Data?

Accessible (LD

C4 Can new resea
C5 Is the user in and identifi
C6 Can the resea
C7 Do authentica

Interoperable (

C8 Is the meta da
C9 Can particula way such as
C10 Can domain-s
C11 Can each conc URI?

Reusable (LD)

C12 Can a data lic
C13 Can the data
C14 Are data sets community,
C15 Is the provide

Type	Name	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	Total
RDM Systems	CKAN	+	+	o	+	-	+	-	(-)	-	-	-	o	+	o	-	(-)
	e!DAL	+	+	-	+	o	+	-	+	+	?	o	+	+	?	+	o
	ePrints	+	+	o	+	-	+	o	-	-	-	-	+	+	o	-	-
	Dataverse	+	+	o	+	+	+	o	+	+	-	o	+	+	o	-	+
	DSpace	+	+	o	o	(o)	+	+	(-)	+	(-)	(-)	o	+	o	-	(o)
	HUBzero	o	o	o	o	-	+	-	-	-	-	-	-	%	o	-	-
	Invenio	+	+	o	+	-	+	-	-	o	-	o	-	o	-	+	-
RDM Applications	Dryad	+	+	o	+	-	+	-	-	+	-	-	-	-	-	-	-
	EUDAT B2	+	+	+	+	+	+	-	-	+	-	o	+	+	+	-	+
	Figshare	+	+	o	+	o	+	o	-	+	-	o	o	+	+	-	o
	Globus	o	-	o	o	-	o	+	-	+	-	-	-	o	-	-	-
	myExperiment	+	o	o	+	+	+	+	+	+	-	+	+	+	+	-	+
	XSEDE	o	-	o	o	-	+	-	-	-	-	-	-	-	-	-	-
	Zenodo	+	+	o	+	o	+	+	+	+	o	+	+	+	+	o	+
RDM Tools	D2R	+	%	+	%	-	o	%	%	-	-	o	%	%	%	%	o
	Dendro	+	%	+	o	+	+	o	+	+	o	+	o	+	+	+	+
	Fedora	+	%	%	+	%	+	+	+	%	%	+	%	+	%	%	+
	Virtuoso	+	%	+	+	-	+	+	+	+	+	+	%	%	+	-	+

\+ was assigned, if the criterion was entirely fulfilled
o was assigned, if the criterion was partially fulfilled
\- was assigned, if the criterion was not fulfilled
% was assigned, if the criterion was not applicable
? was assigned, if it was not possible to assess the mentioned criterion
() was assigned, if the feature is limited in the native version but might be there with plugins

6 Conclusion

Conclusion

- Comparative study with 18 identified web-based general-purpose research data publishing solutions
- Support for Linked Data input and exposure differs among all examined platforms
- Input capabilities for LD meta data often limited to basic discovery meta information

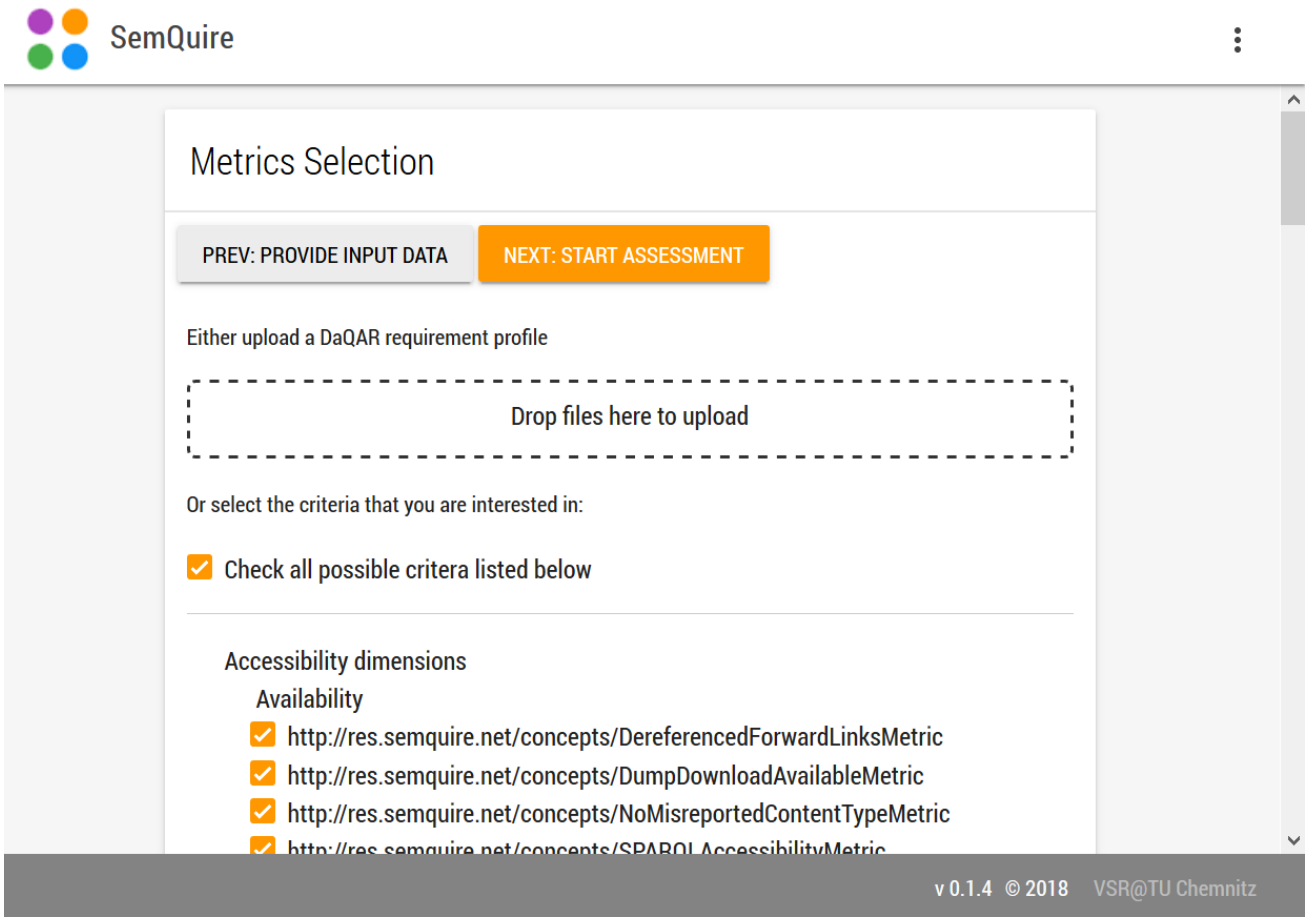
VSR

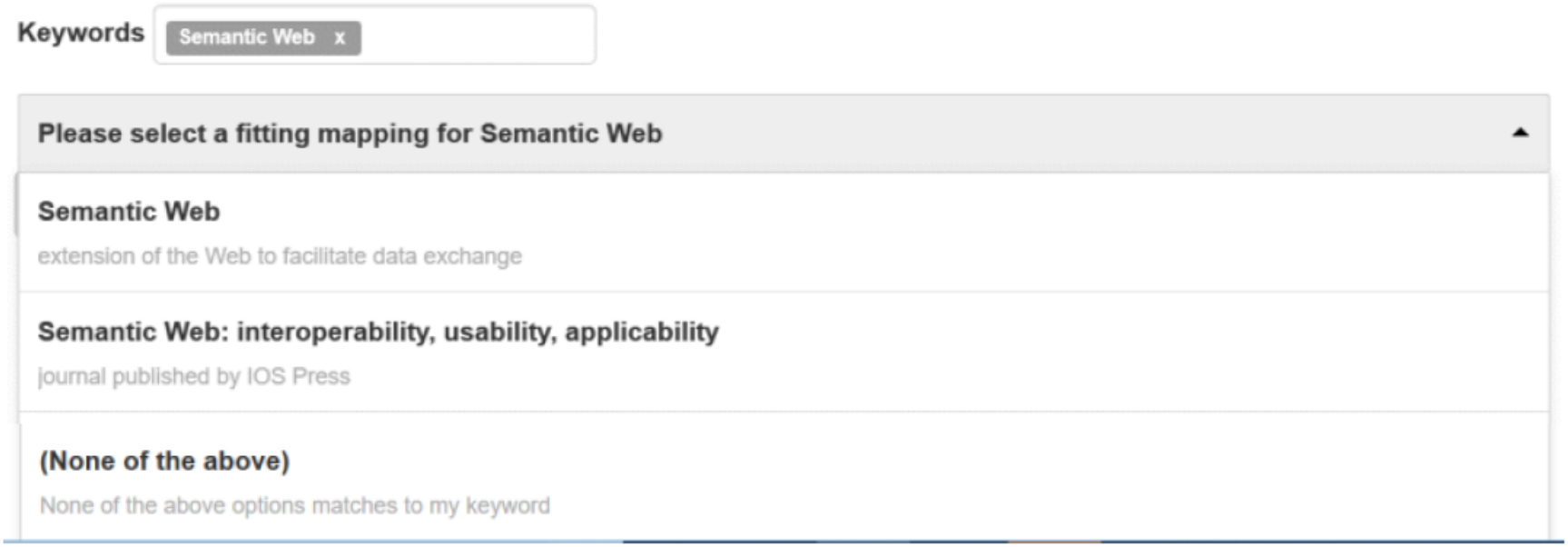
Inspired and Interested?

Andre.Langer@Informatik.TU-Chemnitz.de

VSR.*Informatik*.TU-Chemnitz.*de*

@myVSR /myVSR


SemQuire
Metrics Selection
PREV: PROVIDE INPUT DATA
NEXT: START ASSESSMENT
Either upload a DaQAR requirement profile
Drop files here to upload
Or select the criteria that you are interested in:
Check all possible critera listed below
Accessibility dimensions
Availability
http://res.semquire.net/concepts/DereferencedForwardLinksMetric
http://res.semquire.net/concepts/DumpDownloadAvailableMetric
http://res.semquire.net/concepts/NoMisreportedContentTypeMetric
v 0.1.4 © 2018 VSR@TU Chemnitz


Keywords
Semantic Web x
Please select a fitting mapping for Semantic Web
Semantic Web
extension of the Web to facilitate data exchange
Semantic Web: interoperability, usability, applicability
journal published by IOS Press
(None of the above)
None of the above options matches to my keyword

Langer, A. & Gaedke, M. (2016). Fame.Q - A formal approach to master quality in enterprise linked data. In Isaias, P. (2018). *Proceedings of the 15th International Conference WWW/Internet (ICWI2016)*. Mannheim, Germany. October 28 - 30, 2016.; IADIS; pp. 51-58; ISBN/ISSN 978-1-5108-3297-8

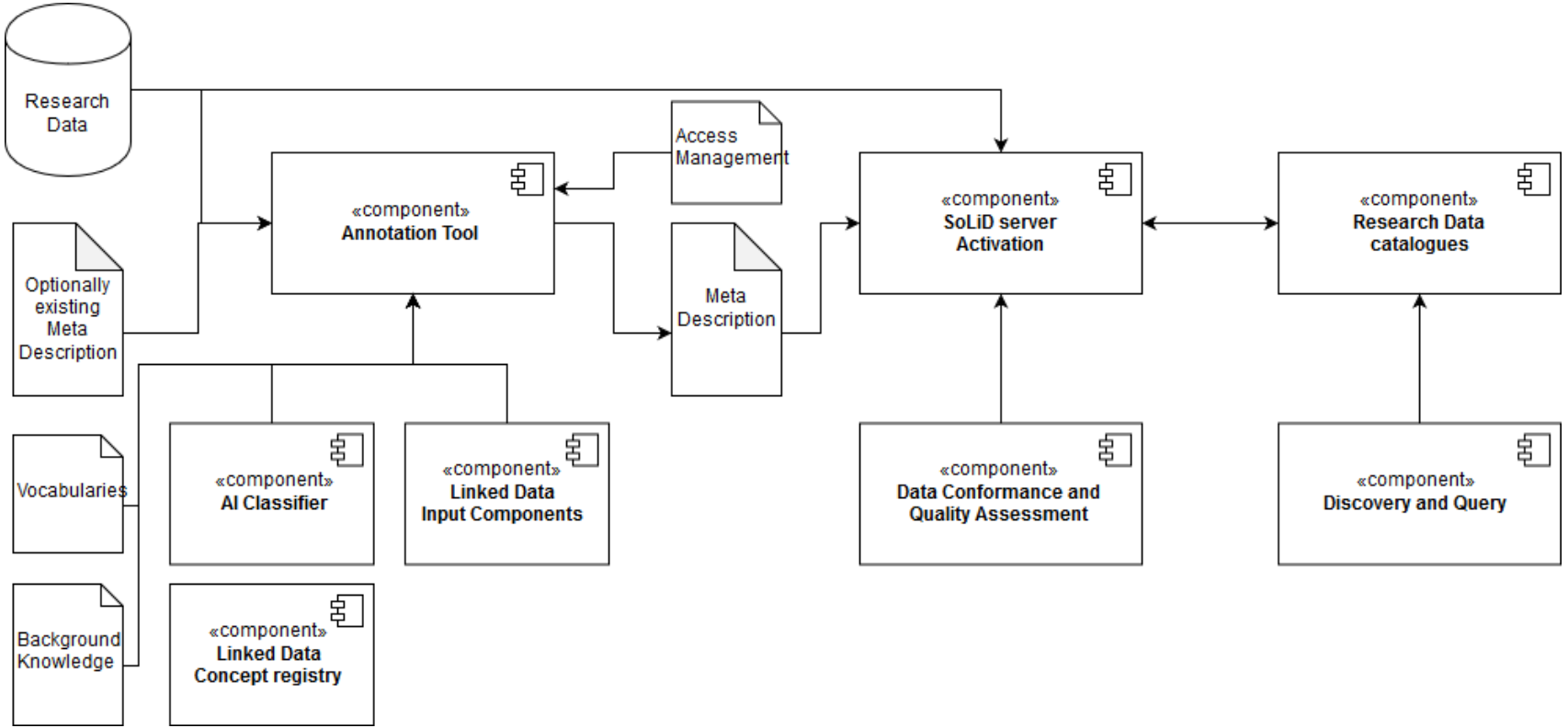
Langer, A., Krug, M., Moreno, L. & Gaedke, M. (2017). Utilizing Linked Data Structures for Social-aware Search Applications. In Eibl, M. & Gaedke, M. (2017). *Informatik 2017: Bände I bis III, 47. Jahrestagung der Gesellschaft für Informatik e.V. (GI)*. Chemnitz, Germany. September 25 - 29, 2017; Bonn: Gesellschaft für Informatik e.V. (GI); pp. 1903-1914; ISBN/ISSN: 978-3-88579-669-5; doi: https://doi.org/10.18420/in2017_190

Langer, A. & Gaedke, M. (2018). DaQAR - An ontology for the uniform exchange of comparable LD quality assessment requirements. In Mikkonen, T., Klamma, R. & Hernández, J. (2018). *Web engineering: 18th International Conference, ICWE 2018, Cáceres, Spain, June 5-8, 2018, Proceedings.*), Lecture Notes in Computer Science. vol. 10845 LNCS. Cham, Switzerland: Springer; pp. 234–242. ISBN/ISSN: 978-3-319-91661-3; doi: https://doi.org/10.1007/978-3-319-91662-0

Langer, A., Siegert, V., Göpfert, C. & Gaedke, M. (2018). SemQuire - Assessing the Data Quality of Linked Open Data Sources. In Pautasso, C., Figueroa, F., Systä, K. & Murillo, J. (2018). *Current trends in web engineering : ICWE 2018 International Workshops, MATWEP, EnWot, KD-WEB, WEOD, TourismKG, Cáceres, Spain, June 5, 2018, Revised Selected Papers.* Cham: Springer; pp. 163–175; ISBN/ISSN: 978-3-030-03055-1; doi: https://doi.org/10.1007/978-3-030-03056-8

Langer, A., Göpfert, C. & Gaedke, M. (2018): F.I.E.L.D.S. - Analyzing Form Input interfaces for Explicit Linked Data handling in document Submission systems. In Isaías, P. & Weghorn, H. (2018). *Proceedings of 17th International Conference WWW/Internet (ICWI2018*); Budapest, Hungary. October 21 – 23, 2018; IADIS; pp. 3-10; ISBN/ISSN: 978-989-8533-82-1

Langer, A., Göpfert, C. & Gaedke, M. (2018) URI-aware user input interfaces for the unobtrusive reference to Linked Data. In *IADIS International Journal on Computer Science and Information Systems*, Vol. 13, No. 2, pp. 62-75; ISSN: 1646-3692


Research Data
Access Management
«component»
Annotation Tool
«component»
SoLiD server Activation
«component»
Research Data catalogues
Optionally existing Meta Description
Meta Description
Vocabularies
«component»
AI Classifier
«component»
Linked Data Input Components
«component»
Data Conformance and Quality Assessment
«component»
Discovery and Query
Background Knowledge
«component»
Linked Data Concept registry