

Wie offen sind Forschungsdaten in Repositorien?

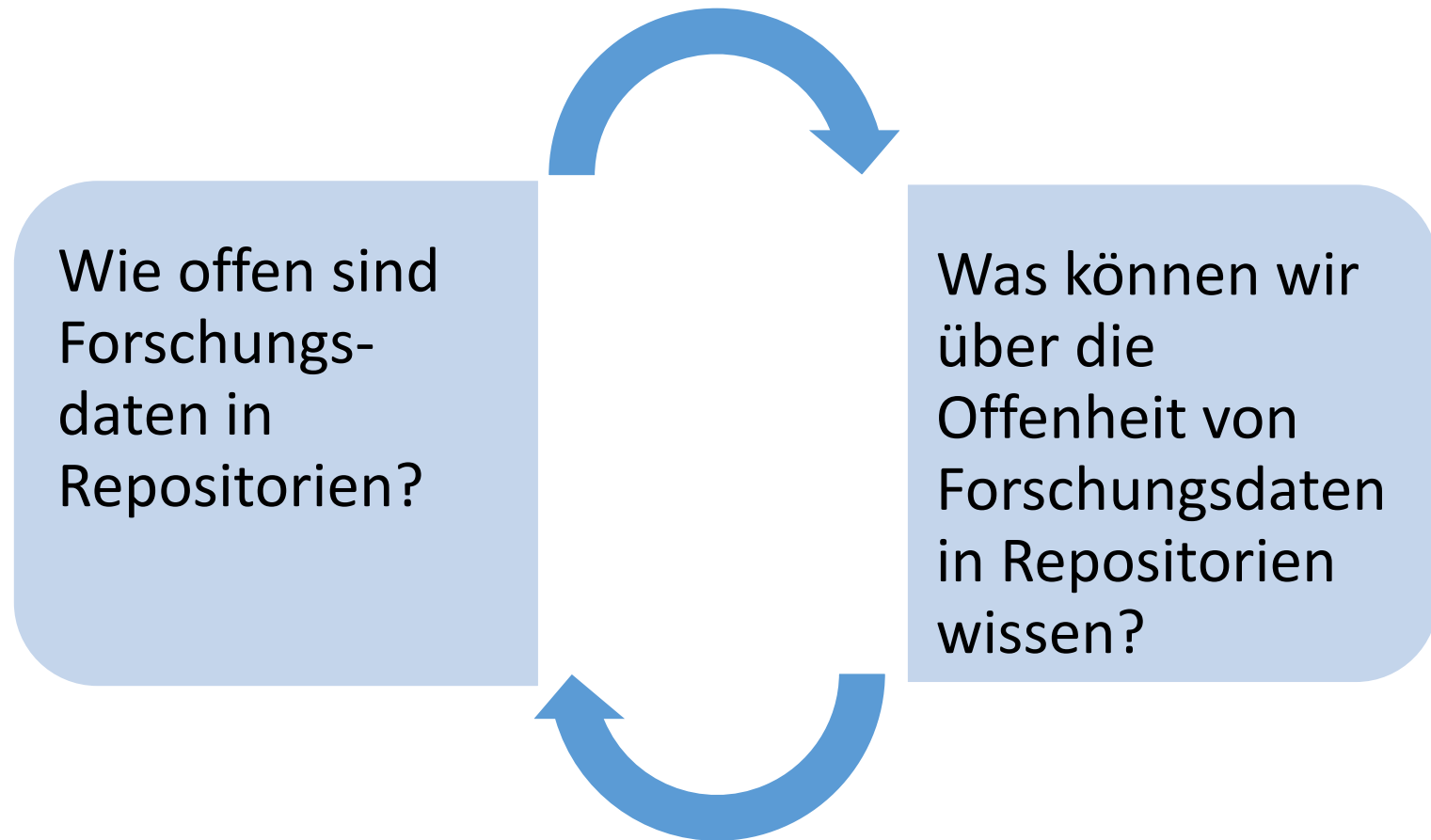
Ergebnisse einer Metadatenanalyse

Maxi Kindling

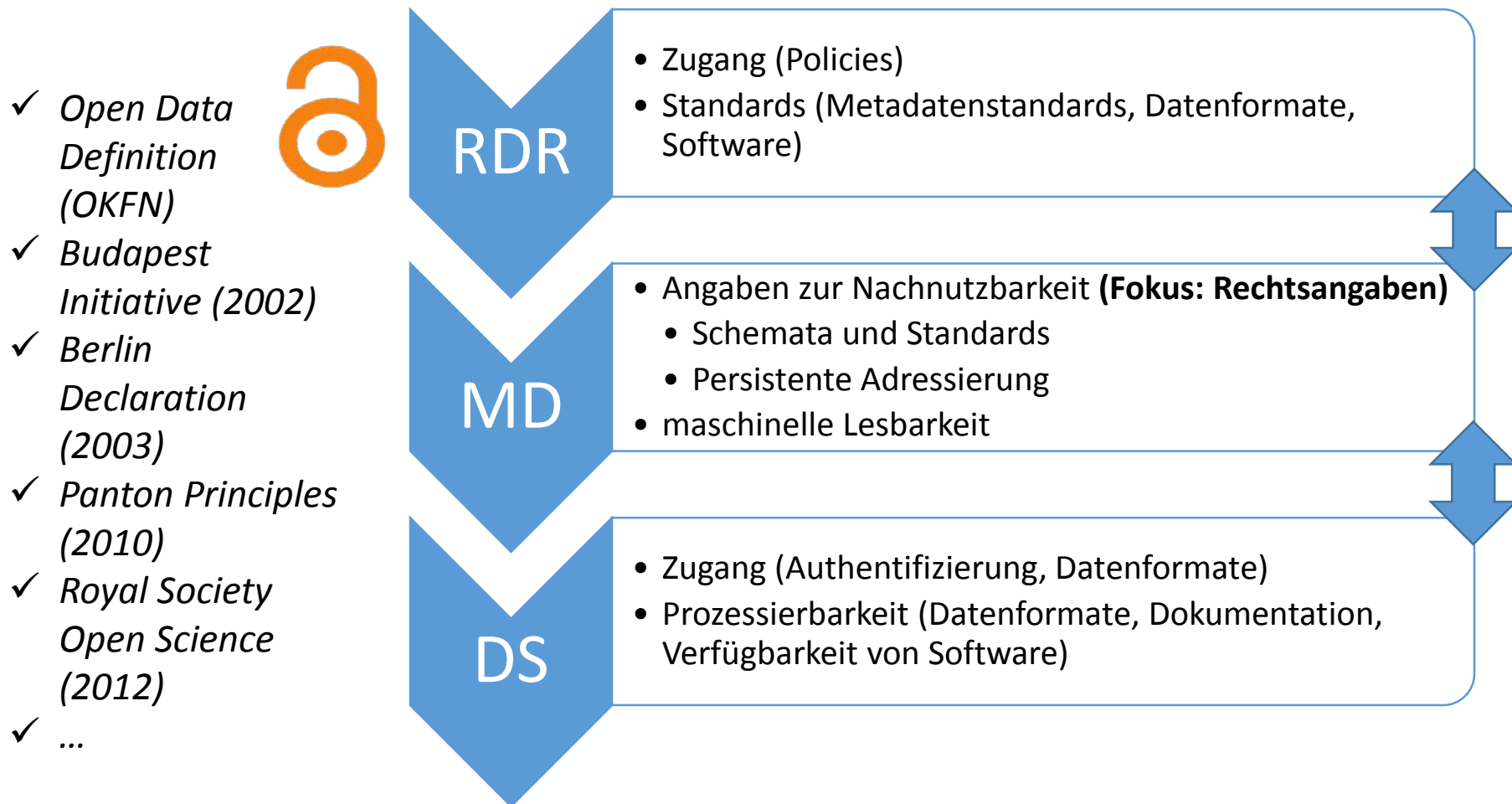
Open-Access-Tage 2015

Universität Zürich

8. September 2015



Offenheit und Nachnutzbarkeit von Forschungsdaten in Repositorien



RDR = Research Data Repository; MD = Metadata; DS = Data Set

Offenheit und Nutzbarkeit von Forschungsdaten in Repositorien



Pampel, H. et al. (2013). Making Research Data Repositories Visible: The re3data.org Registry. PLOS ONE, 8(11), e78080. doi:10.1371/journal.pone.0078080

Vierkant, P., et al. (2014). Schema for the Description of Research Data Repositories. Version 2.2. doi:10.2312/re3.006



The research data repository provides additional information on its services.



The research data repository is either certified or supports a repository standard.



The repository provides access to the data.



The repository provides access to the data.



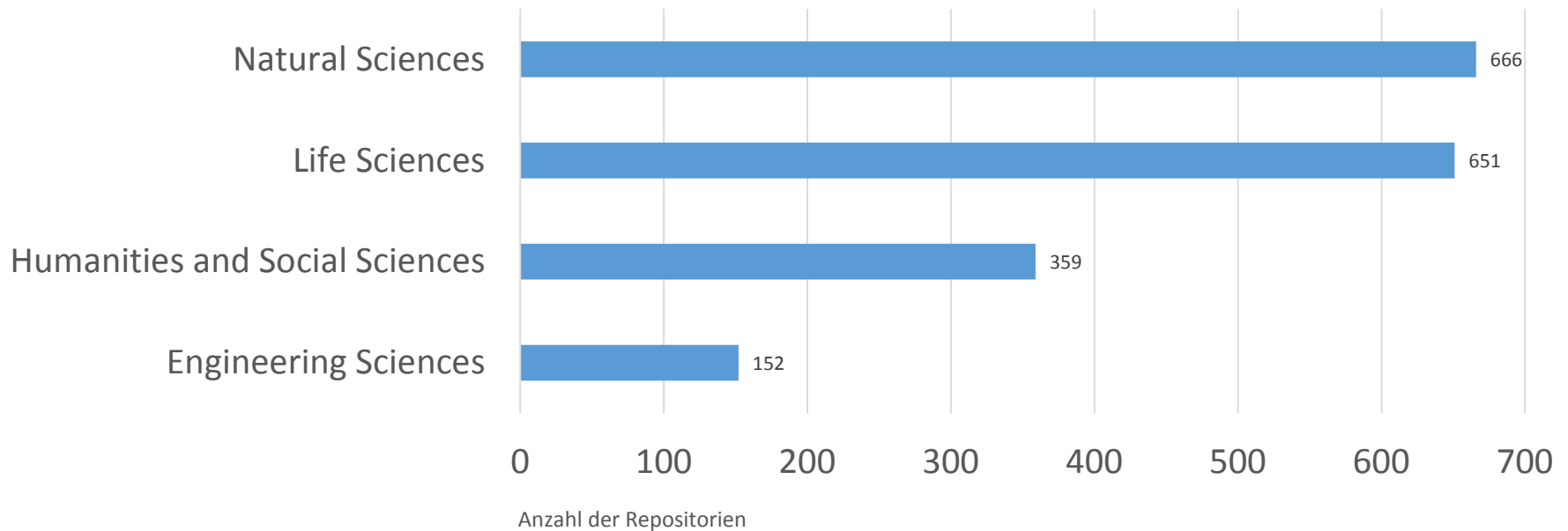
The research data repository uses a persistent identifier system to make its provided data persistent, unique and citable.



The terms of use and licenses of the data are provided by the research data repository.

n = 1312 Repositorien in re3data.org (Stand: 14. August 2015)

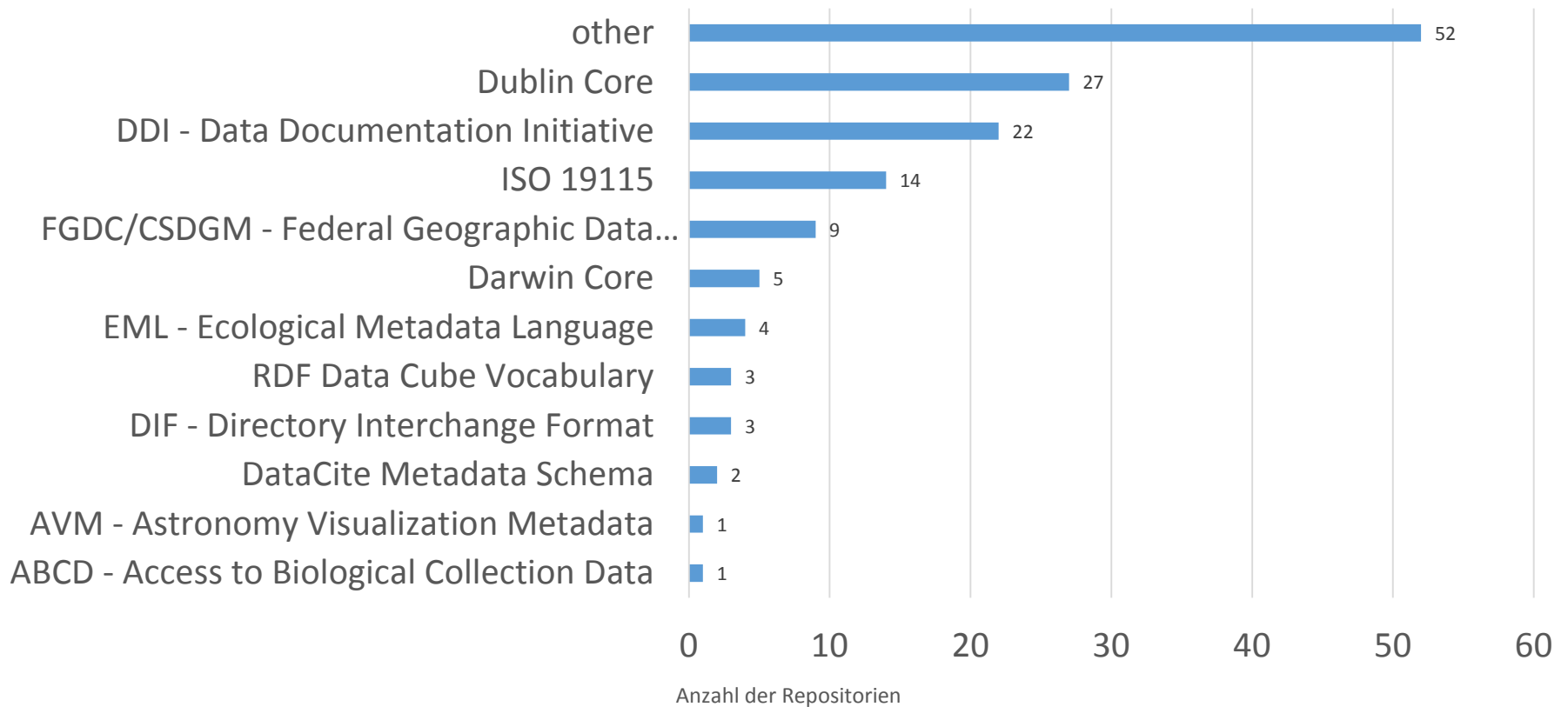
DFG subjects (oberste Kategorien, Mehrfachnennung möglich)



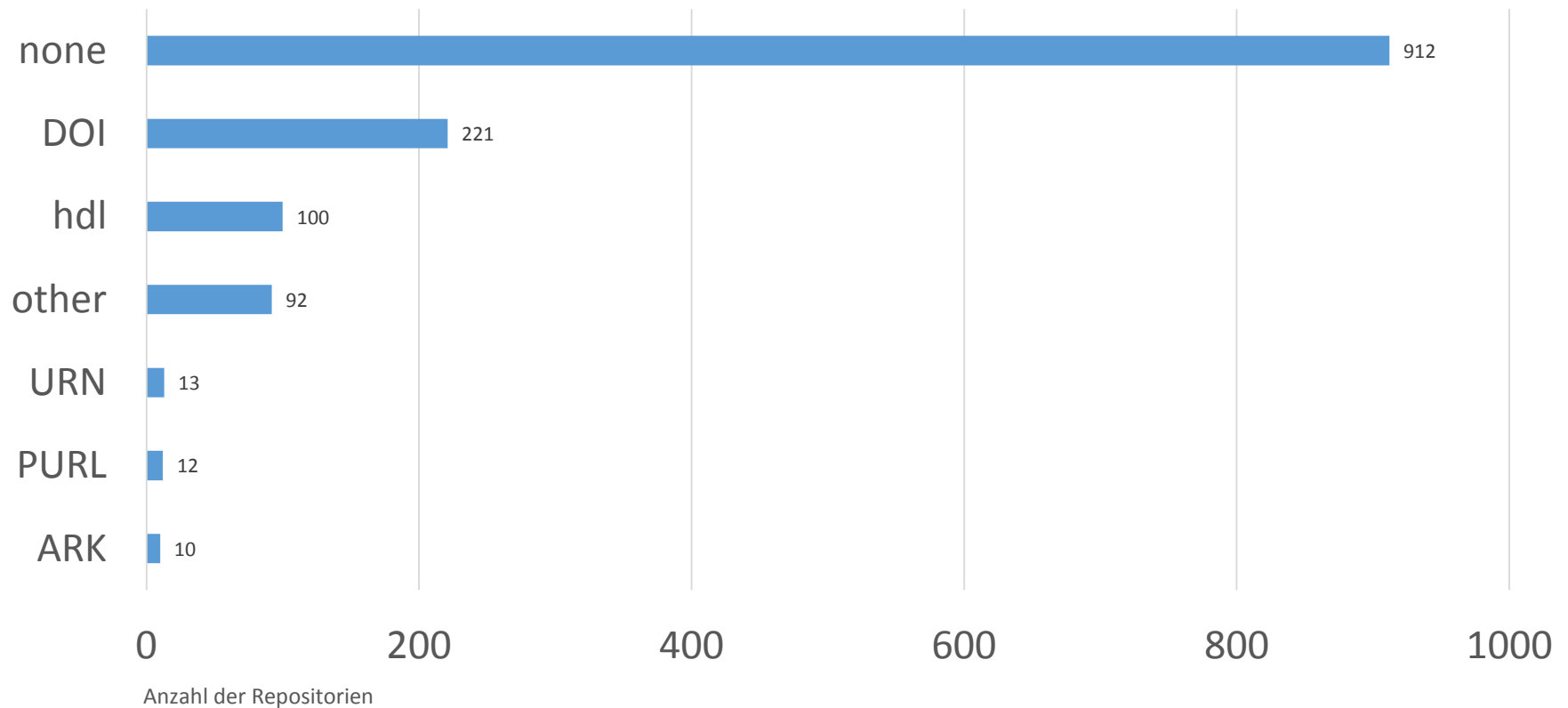
**Database Access type, Data Access type & Data Upload type
für jeweils n = 1312 (Mehrfachnennungen möglich)**

	Open	Teilweise Open
Database Access	1250	0
Data Access	659	468
Data Upload	49	1

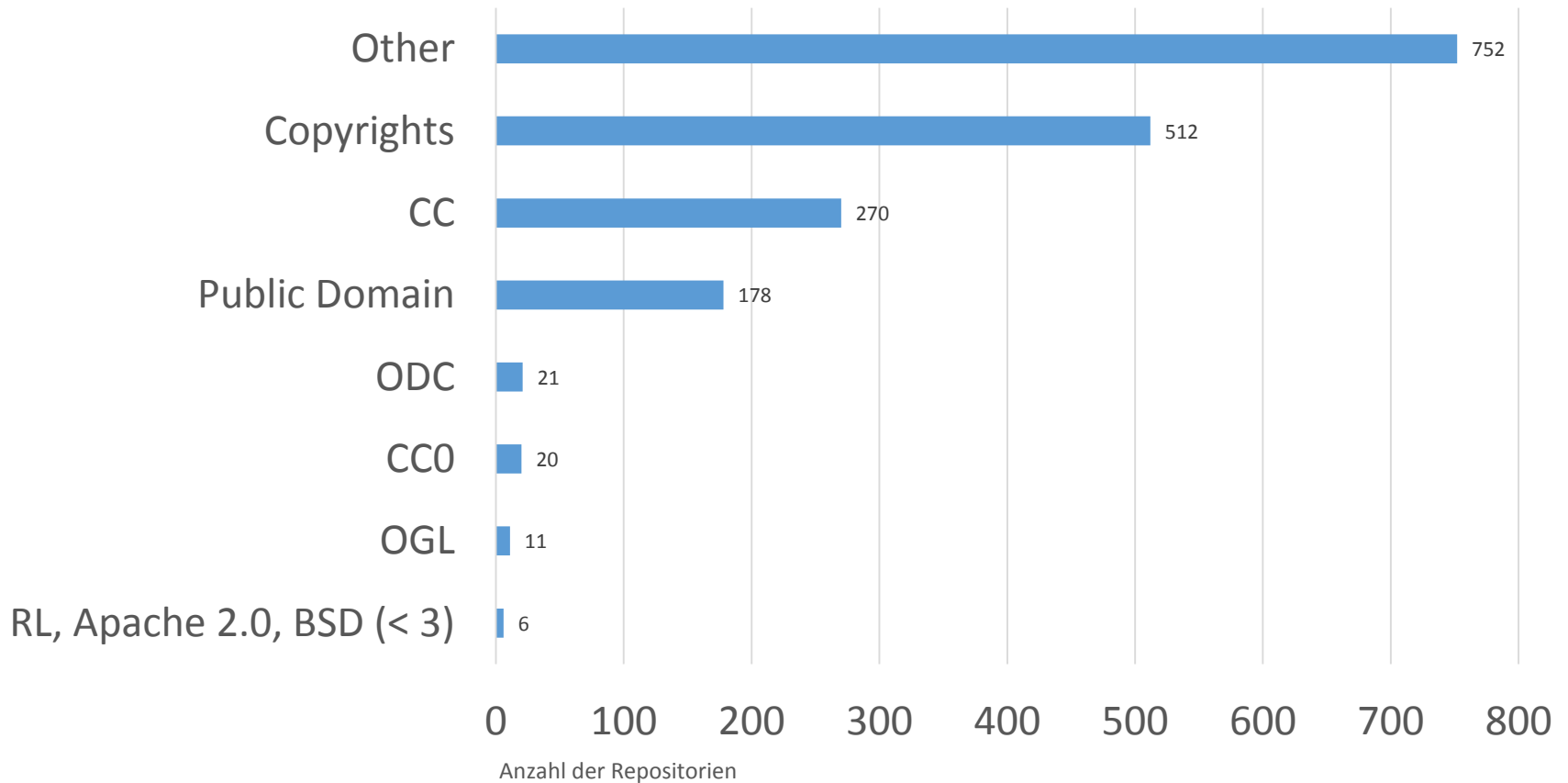
**Metadata standards für n=1312
(Mehrfachnennung möglich, gesamt = 143)**



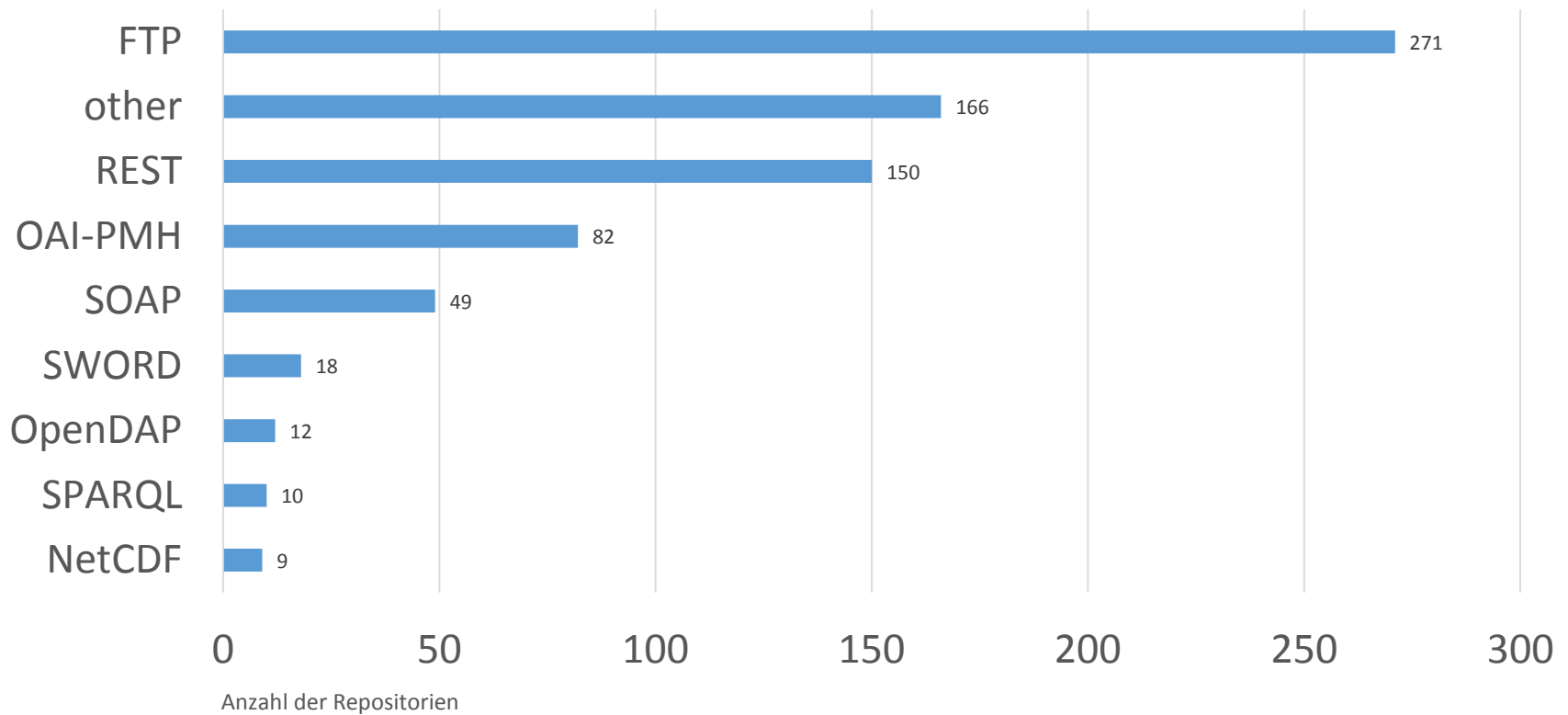
Persistent Identifier für n=1312
(Mehrfachnennung möglich, gesamt = 1360)



License type für n=1312 (Mehrfachnennung möglich, gesamt = 1170)



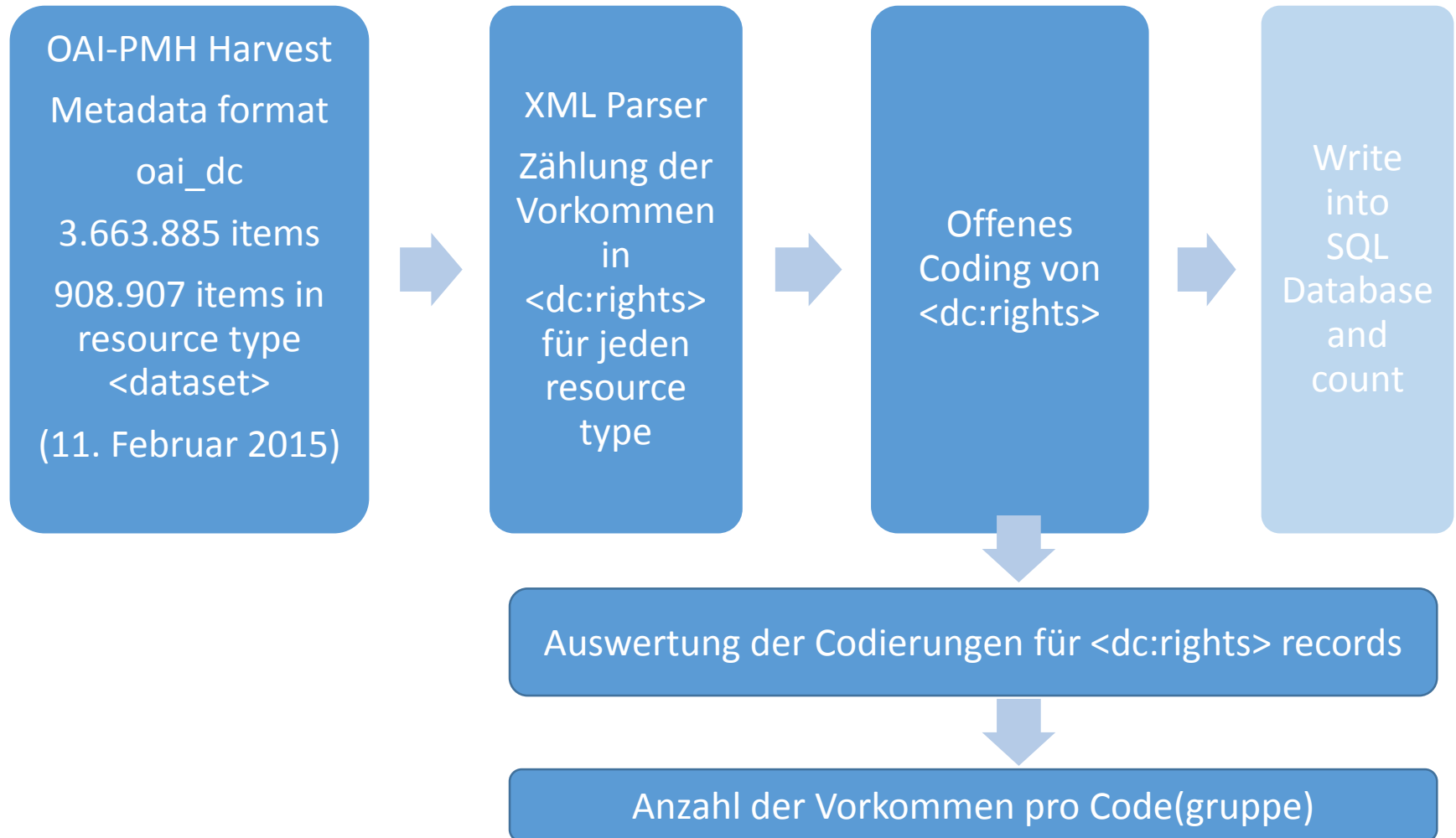
API type für n=1312 (Mehrfachnennungen möglich, gesamt = 767)



Offenheit und Nutzbarkeit von Forschungsdaten in Repositorien



RDR = Research Data Repository; MD = Metadata; DS = Data Set



Gruppierung auf Basis des Codesystems	Anzahl der Vorkommen
Leeres Feld	18
Open Data Commons	613
Creative Commons	604.689
OGL	23
Rights reserved	183
Terms of Use	13.912
Text	3.127
Gesamt	622.565

Zahl der items
in resource type
„data set“
908.907

Gruppierung im Codesystem (Anzahl der vergebenen Codes)	Variante	CC: durchschnittlich 6,4 Vorkommen pro Lizenzvarianten
Creative Commons Licenses + CC0 (34)	< 100	
Digital Peer Publishing Licenses (3)		11
info:eu-repo Licenses (4)		5
Open Government Licenses (2)		3
Open Data Commons Licenses (3)		14
Open Source Licenses (25)		74
Rights reserved/Copyrights statements (6)		> 5.000
Terms of Use (5)		86
Text statements (12) Sprache!		95
Codesystem gesamt (94)		> 5.617

Creative Commons cc:by:nc:nd 3.0 international: 23 verschiedene Varianten der Lizenzangabe

- Abhängigkeit von (Zahl) der Content-Provider
und der Zahl der items mit dieser Lizenzangabe
- Abhängigkeit bei offeneren Kategorien
(Text/Terms of Use) auch vom Codesystem

The data may be used for scholarly
research only.

Editura Medical? Universitar?[...]

Open access for scientific and
educational purposes under
requirement of correct citation
(see above).

20 Minuten

„Der Vertragstext bei CC-Lizenzen sieht vor, dass LizenznehmerInnen bei weiterer Verwendung der Inhalte immer eine Kopie der Lizenz oder die Uniform Resource Identifier (URI) der jeweiligen Lizenz angeben.“ (S. 56)

Voigt, Micheala (2014) Nutzungsrechte bei institutionellen Repositorien in Deutschland – Analyse der Rechteübertragung und Metadatenangaben.
urn:nbn:de:kobv:11-1002283262014

Codierung (open)	Varianz <dc:rights> records mit URI (von gesamt)	Anzahl Vorkommen <dc:rights>-records mit URI (open)
cc:by 3.0 international	2 (4)	201.743
cc:by:nc 3.0 international	1 (3)	1.546
cc:by:nc:sa 3.0 Ireland	1 (1)	360
CC0 1.0 Public domain dedication	3 (9)	206
cc:by:nc:nd 3.0 international	1 (2)	6
cc:by:sa 3.0 international	1 (2)	5
cc:by:nc:sa 3.0 international	1 (2)	1
cc:by:nd 3.0 international	1 (2)	1
cc:by 3.0 Australia	1 (1)	1
Open Data Commons BY 1.0	1 (4)	1
Gesamt	13 (98)	203.870 (201.951)

Mehr offene Forschungsdaten als wir wissen können (?)

Ein Blick in die Metadaten lohnt.

Metadatenangaben zur Offenheit und Nachnutzbarkeit müssen normiert, standardisiert, maschinell lesbar aufbereitet werden.

- Unterstützung von Initiativen wie:
www.rightsstatements.org

Grenzen und Möglichkeiten:

- Mehr Metadateninformationen über oai_dc (Datenformate, Datentypen, Metadatenstandards).
 - Service Provider oder Dataprovider?
 - Metadatenmapping zu oai_dc
 - Verknüpfung mit Datentypen/Repositorien.
- Weitere Samples (re3data.org RDR und BASE in Vorbereitung).
- Mehr Schnittstellen.
- Was ist ein Forschungsdatenrepositorium?
Was beschreiben Metadaten? Was ist ein Datensatz?

Danke!



Vielen Dank für die Aufmerksamkeit!
Fragen und Details: maxi.kindling@hu-berlin.de

Mein Dank für den Support geht an:
Alexander Petrus und Stephanie van de Sandt (HU Berlin),
das Projektteam von re3data.org,
Frauke Ziedorn (DataCite)
und vorab an Friedrich Summann (BASE)

