<head>
<meta name = "DC.Title"
content = "Deposit Behavior in Institutional Repositories: A Case Study of ORBi">

<meta name = "DC.Creator"
content = “Behrooz Rasuli, IranDoc, rasuli@irandoc.ac.ir">

<meta name = "DC.Creator"
content = “Paul Thirion, Université de Liège, Paul.Thirion@uliege.be">

<meta name = “DC.Presenter"
content = “Fatemeh Seyfzadeh, Université de Liège, f.seyfzadeh@uliege.be ">

<meta name = "DC.Date"
content = “Nov 7, 2024">

<meta name = “DC.Event"
content = "ULiège Open Science Day">

<meta name = "DC.Language"
content = "en">
</head>

Agenda

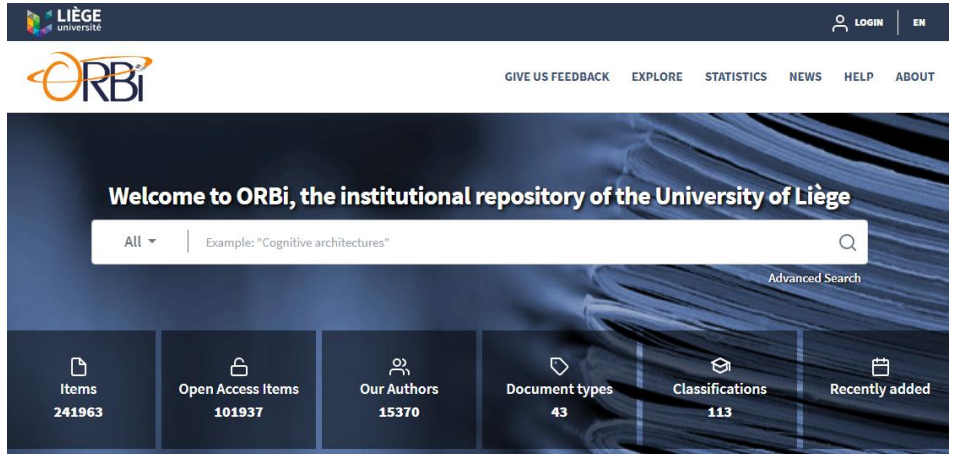
- Research problem
  - ORBi Case Study
  - Nature and types of Metadata
  - Factors affecting metadata quality
- Methods and data
- Findings
- Conclusion
- Recommendations

Research Problem

- Metadata Practices
  - Differences in metadata practices across disciplines
- Metadata quality and its recognized importance
  - The impact of metadata completeness on the visibility and usage

Context of this study - ORBi

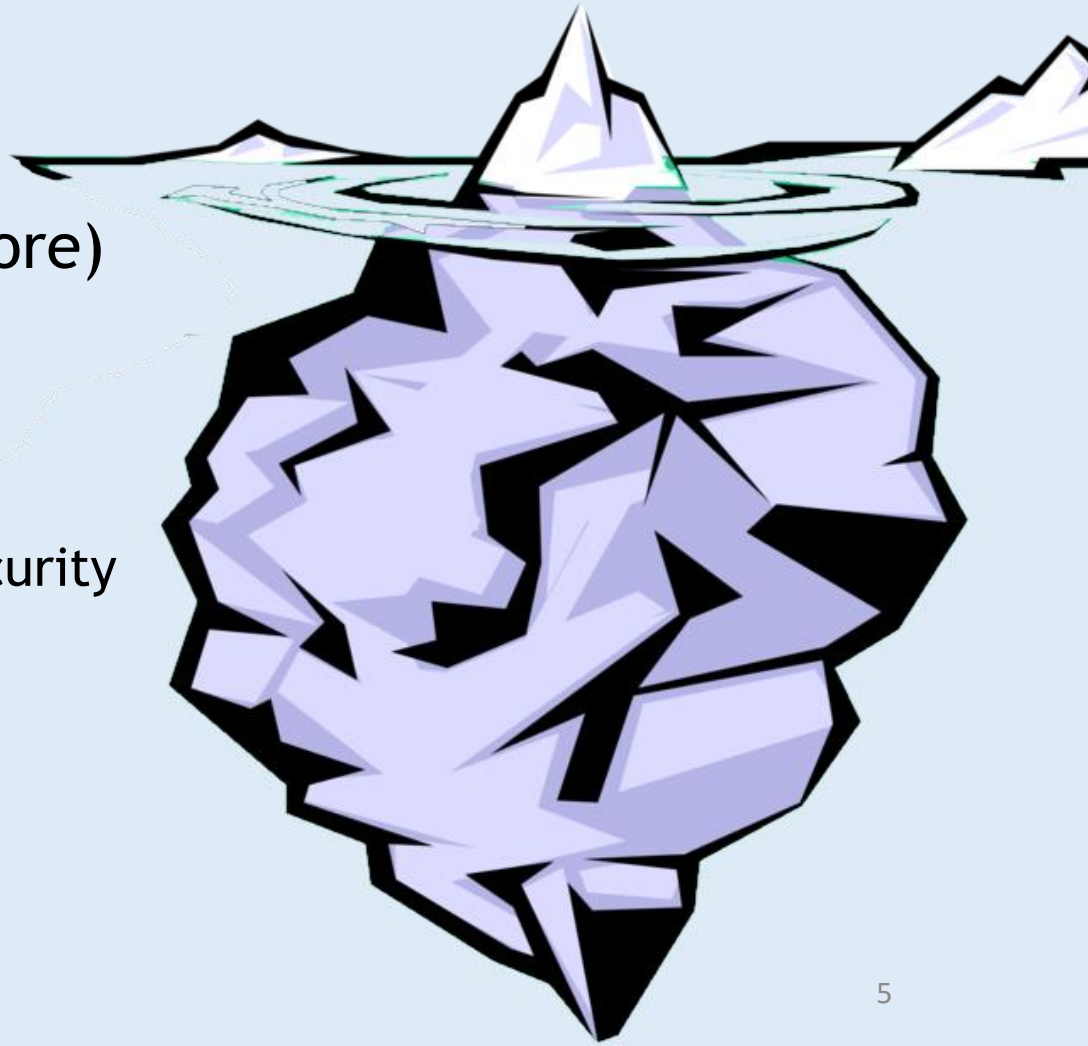
- ORBi (Open Archive and Bibliography)
  - The institutional archive of the University of Liège (EST. 2008)
  - Within the context of its implementation, a compulsory submission mandate
  - Principal objectives: To increase the visibility & usability; strengthen the Open Access practices; facilitate internal evaluation; and provide the public with free and fast access; institutional values

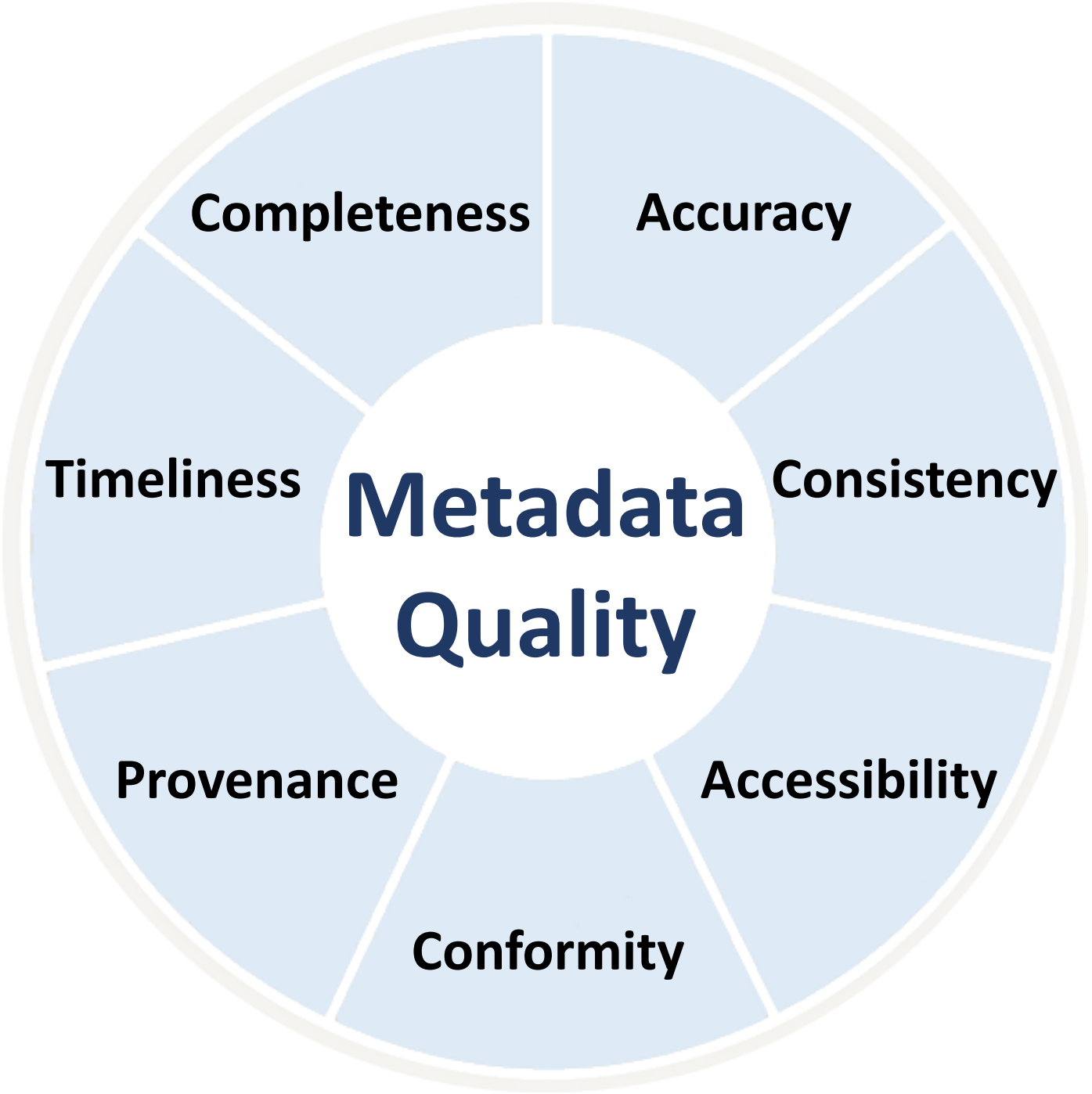


- Internal reviews dependent on only publications and communications available on ORBi

Metadata

- Data about data
- Recognized metadata schema (e.g., Dublin Core)
  - Fields, tags, number of items
  - Different purposes (e.g., findability, archiving, security management)
  - Different across disciplines





Metadata standards (schema) and the completeness of metadata are closely connected!

Study Rationale

- Increasing importance of IRs:

Impact of open science paradigm; archiving as an institutional asset; increased discoverability

- Role of metadata quality:

the backbone of the discoverability of publications

- More metadata fields:

comprehensive completion of metadata fields, facilitating better indexing by search engines and IRs, enabling users to locate and use relevant works

- Informed user engagement:

Detailed information about publications, higher engagement rates (e.g., views and downloads)

Research Methods

1

Crawl and collect data on 240,758 records from ORBi

till the end of October 2024

https://orbi.uliege.be/handle/2268/XXXXX

Unique ID for each record

2

Data Extraction

Record ID, Discipline, Date of availability, Number of completed metadata fields, Abstract/Title word count, Number of downloads/views

3

Data Organization

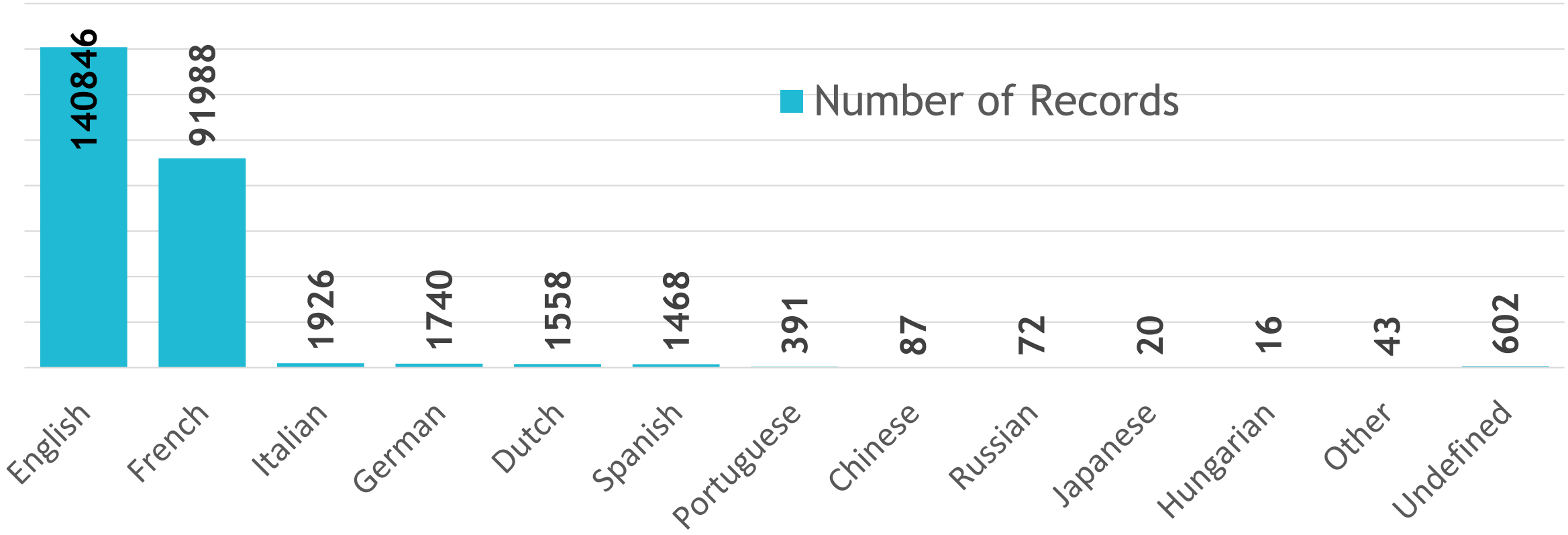
Excel file/SPSS

4

Data Analysis

Descriptive & Inferential Statistics

ORBi – Statistics: Language



33 Languages

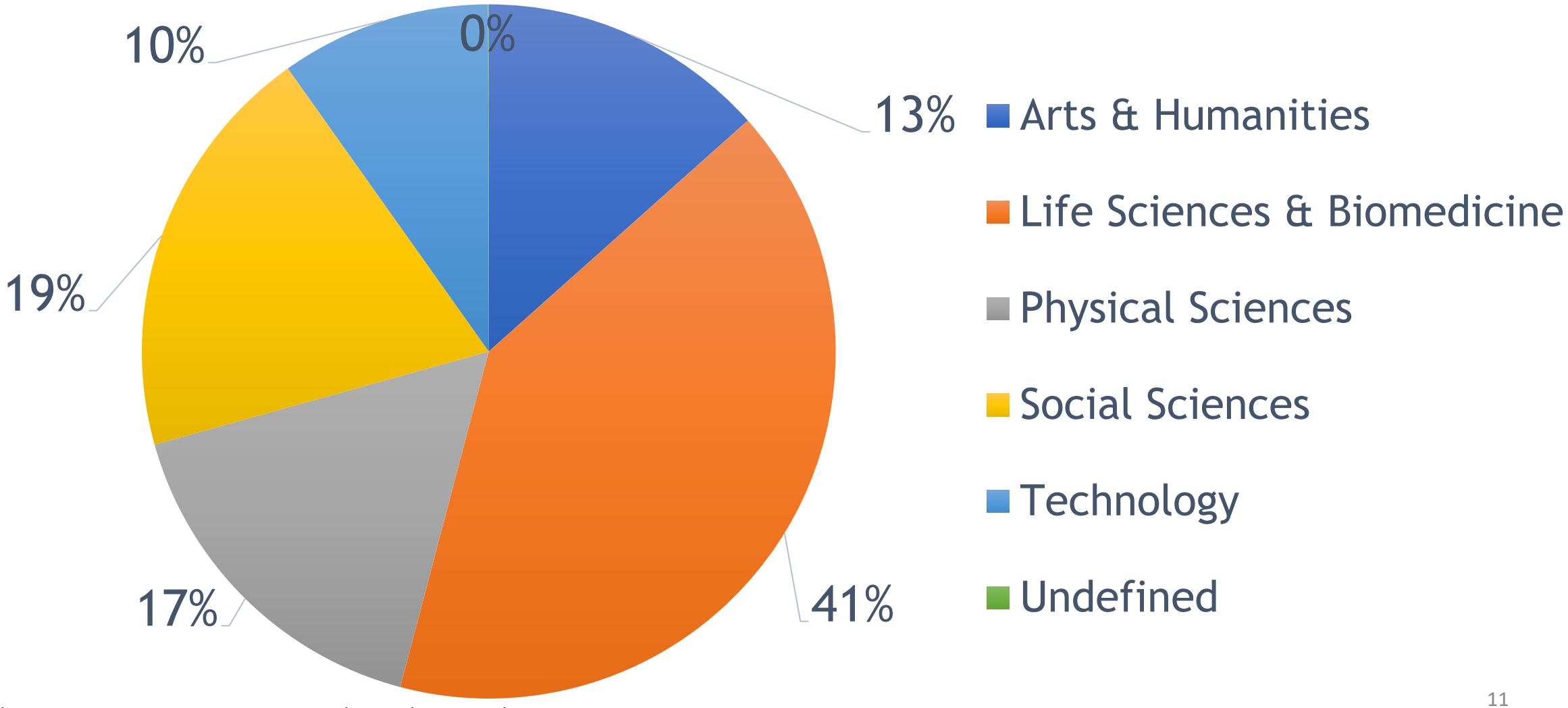
ORBi – Statistics: Disciplines

Discipline	Records
Earth sciences & physical geography	8452
Space science, astronomy & astrophysics	8168
Veterinary medicine & animal health	7516
Biochemistry, biophysics & molecular biology	6315
Literature	6155
Education & instruction	5633
Physics	5388
Chemistry	5001
History	4578
Agriculture & agronomy	4555
Sociology & social sciences	4552
Environmental sciences & ecology	4476
Political science, public administration & international relations	4476
Pharmacy, pharmacology & toxicology	4166
Civil engineering	4137
Philosophy & ethics	3938
Mathematics	3904
Languages & linguistics	3713
Neurosciences & behavior	2996
Art & art history	2896
Treatment & clinical psychology	2654

Discipline	Records
Endocrinology, metabolism & nutrition	2602
Classical & oriental studies	2521
Neurology	2381
Phytobiology (plant sciences, forestry, mycology...)	2237
Arts & humanities: Multidisciplinary, general & others	2216
General & internal medicine	2211
Electrical & electronics engineering	2190
Theoretical & cognitive psychology	2176
Oncology	2161
Engineering, computing & technology: Multidisciplinary, general & others	2121
Cardiovascular & respiratory systems	2102
Archaeology	2058
Public health, health care sciences & services	2007
Animal production & animal husbandry	2000
Entomology & pest control	1936
Geological, petroleum & mining engineering	1863
Zoology	1787
Anatomy (cytology, histology, embryology...) & physiology	1767
Computer science	1759
Performing arts	1736
Microbiology	1712
Food science	1691

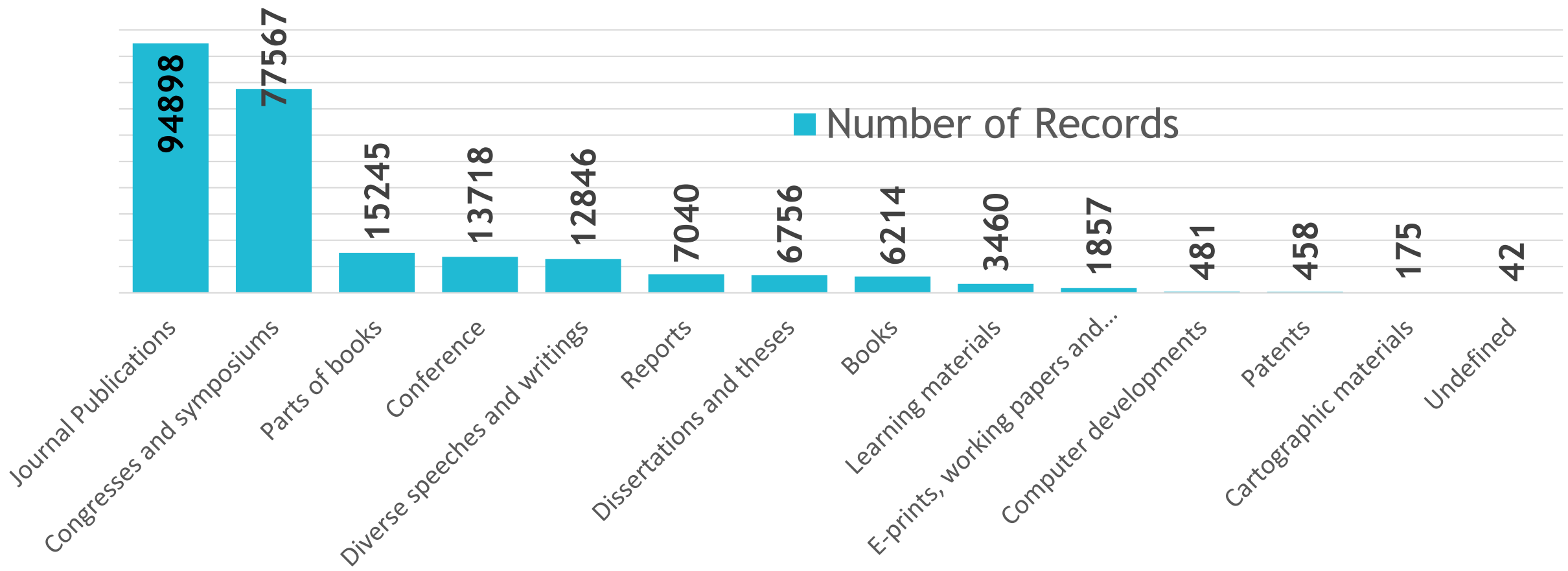
Discipline	Records
Architecture	1664
Aquatic sciences & oceanology	1658
Materials science & engineering	1638
Mechanical engineering	1622
Anesthesia & intensive care	1616
Chemical engineering	1484
European & international law	1483
Orthopedics, rehabilitation & sports medicine	1479
Civil law	1472
Laboratory medicine & medical technology	1441
Anthropology	1356
Genetics & genetic processes	1266
Hematology	1234
Public law	1222
Dermatology	1220
Surgery	1206
Life sciences: Multidisciplinary, general & others	1199
Energy	1187
ChemistryMaterials science & engineering	1131
Urology & nephrology	1097
Aerospace & aeronautics engineering	1094
Tax law	1089
Rheumatology	1083
Physical, chemical, mathematical …	1081
Biotechnology	1070
Dentistry & oral medicine	1003

ORBi – Statistics: Categories



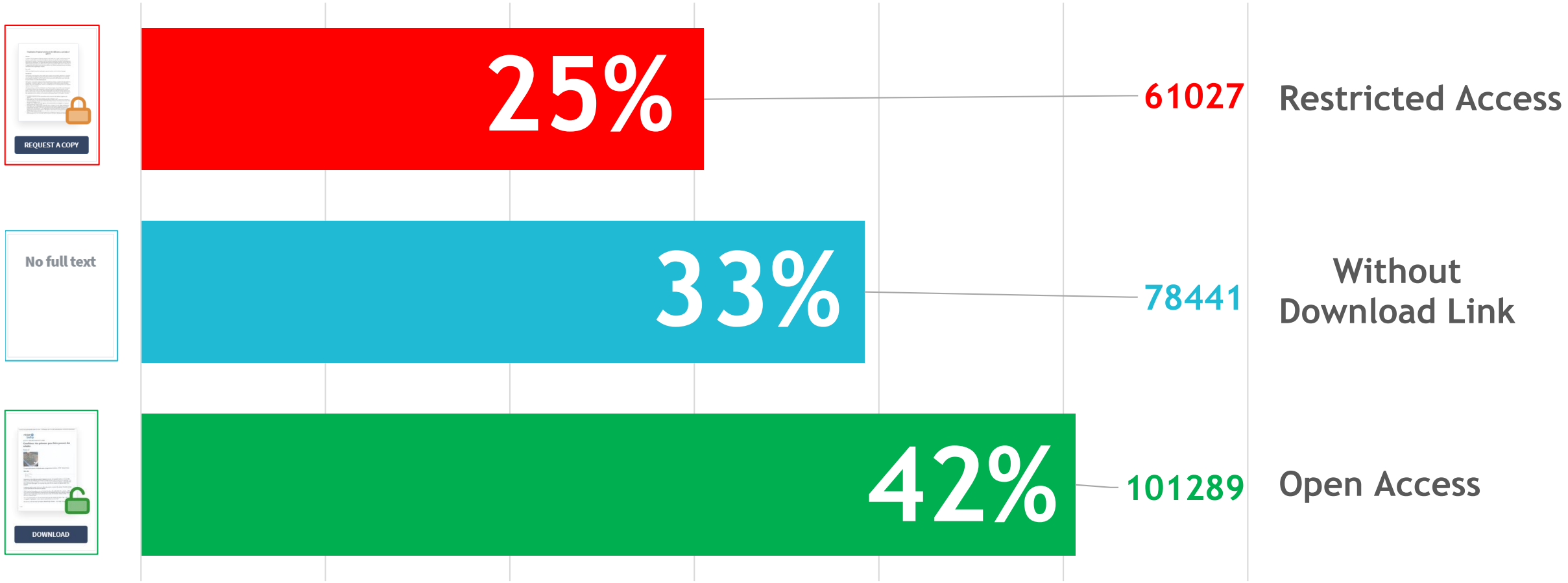
https://webofscience.help.clarivate.com/en-us/Content/research-areas.html

ORBi – Statistics: Document Types

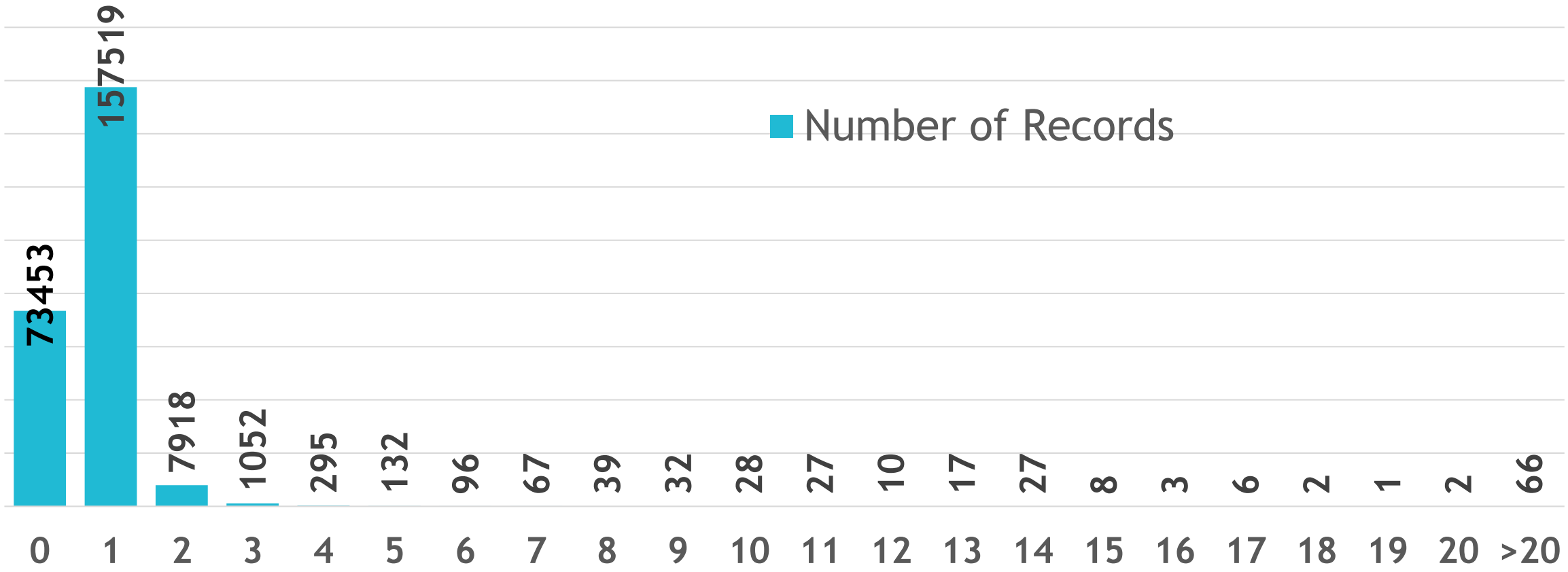


13 Document Types

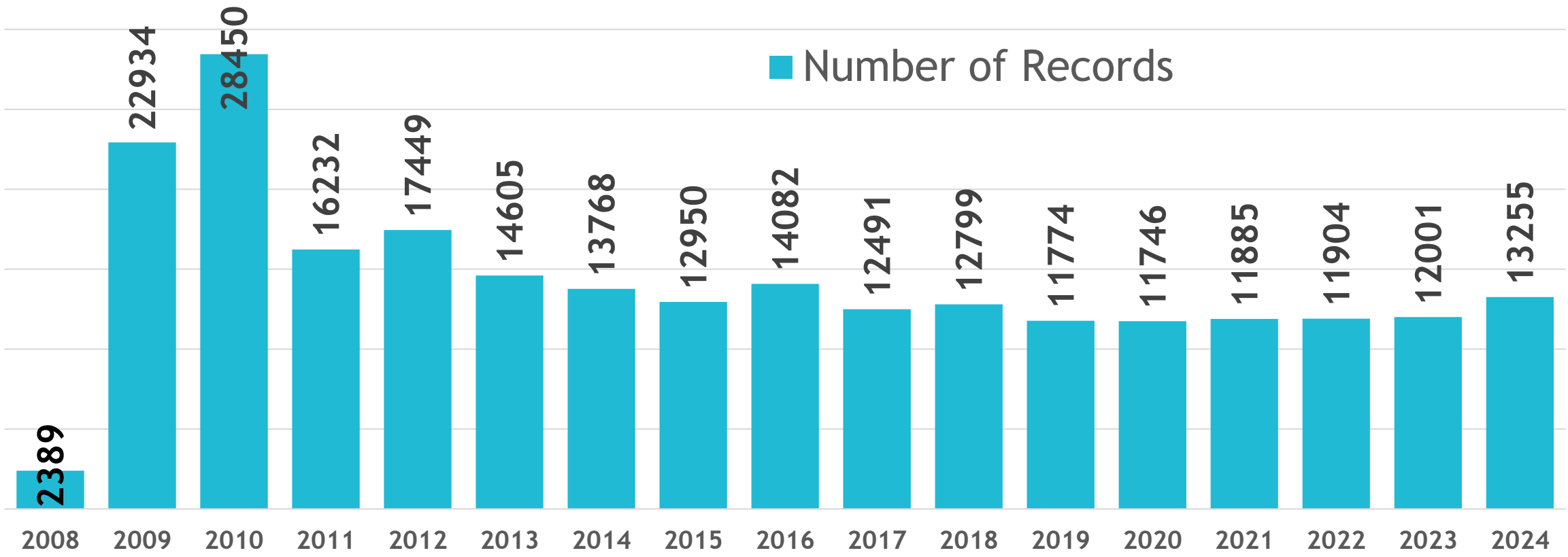
ORBi - Statistics: Access Type



ORBi – Statistics: Attached Files

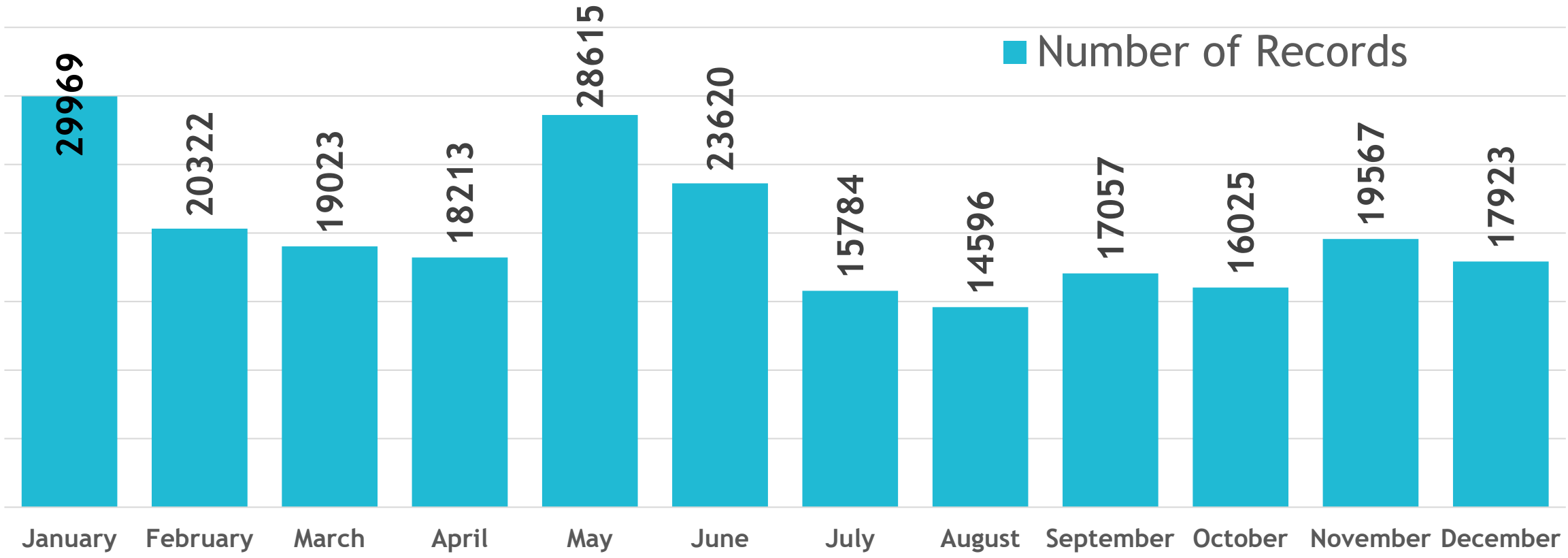


ORBi – Statistics: Publication Year



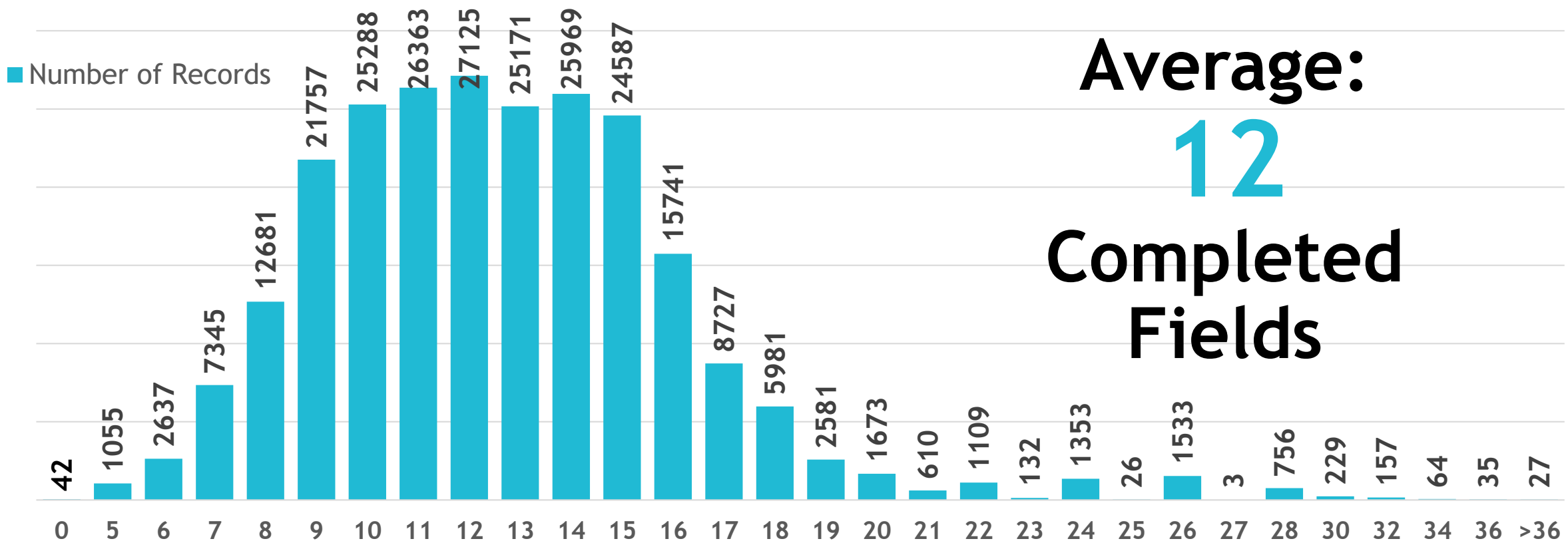
Publication Year

ORBi – Statistics: Publication Month



Publication Month

Findings: Completed Metadata Fields

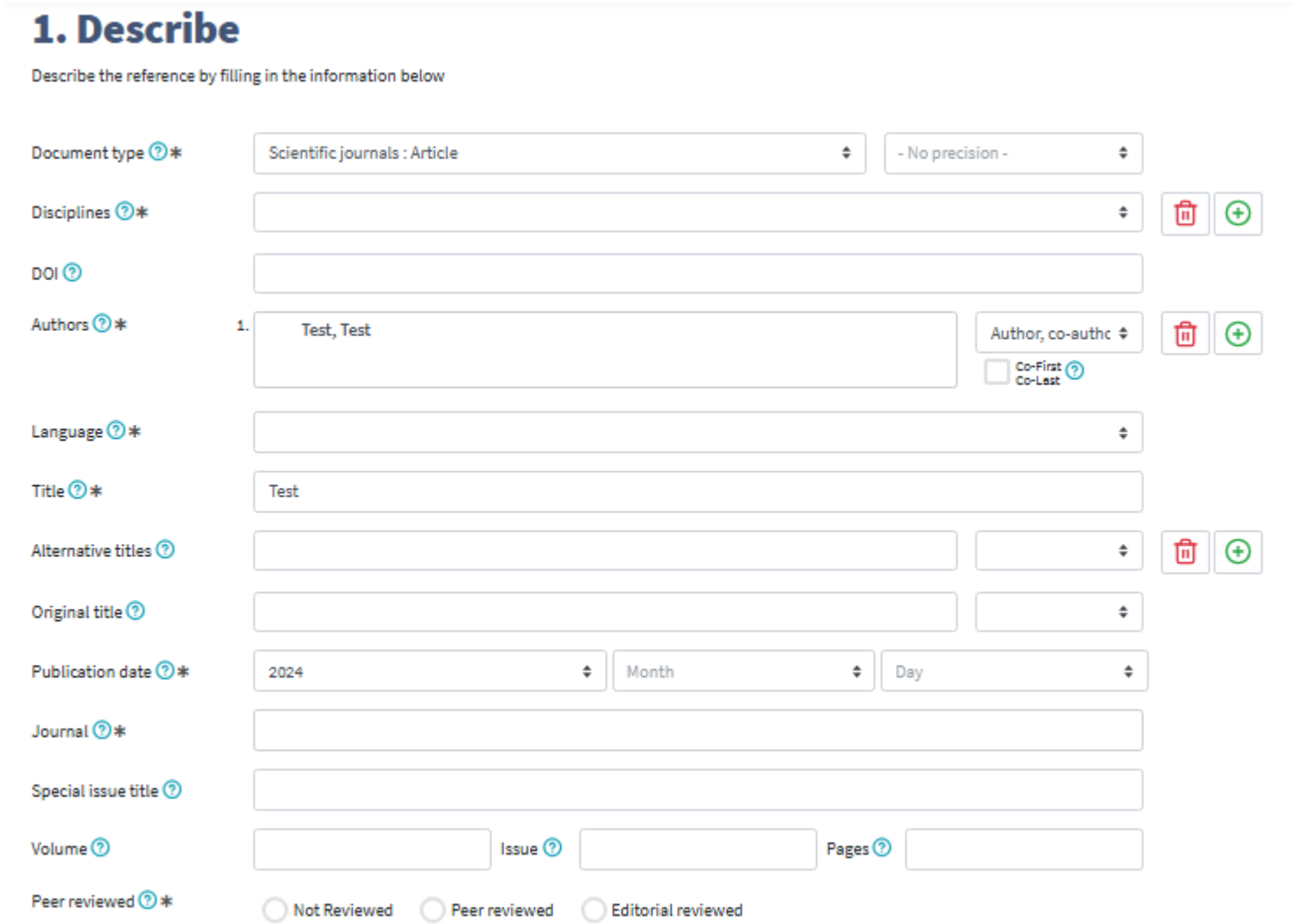


Average:

12

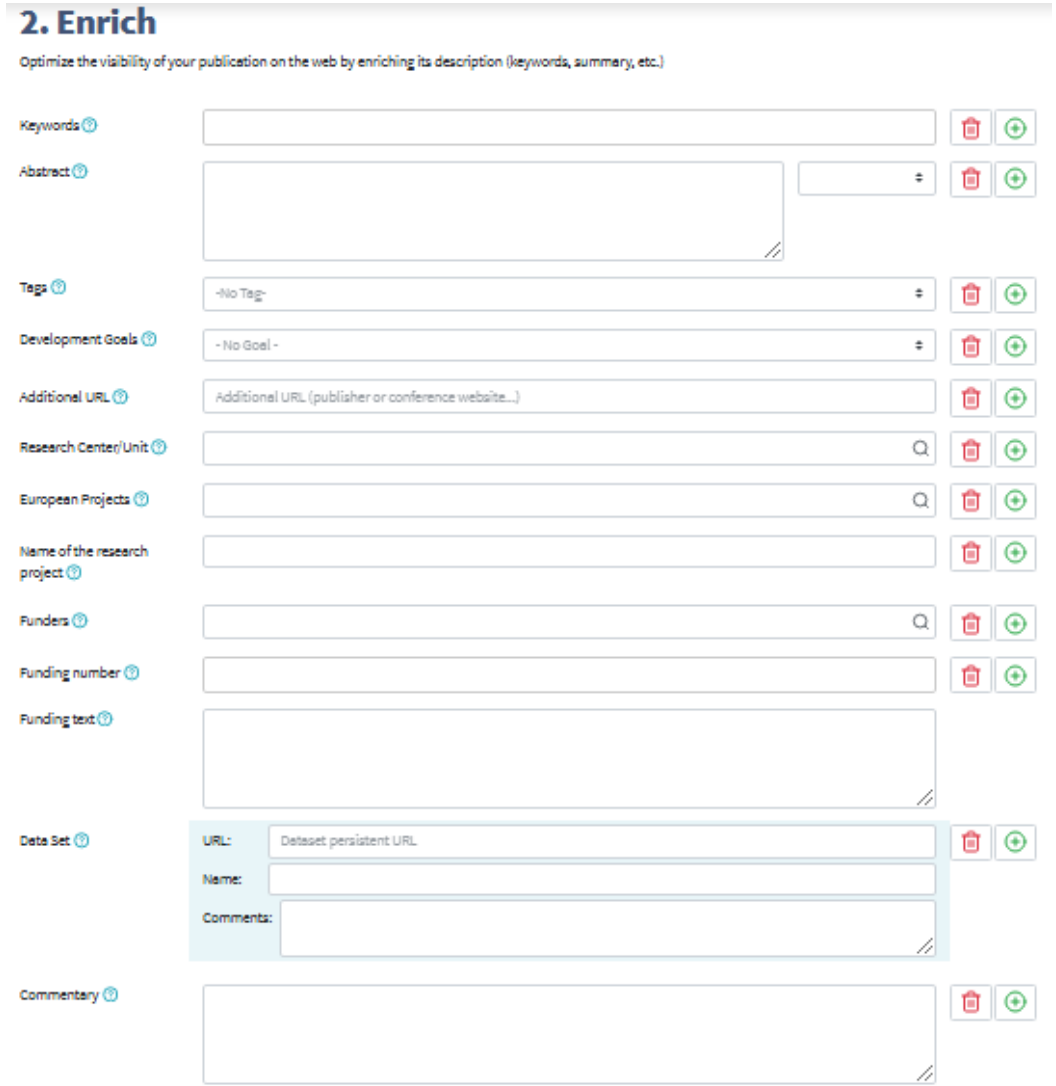
Completed Fields

Findings: Metadata Fields - Journal Article



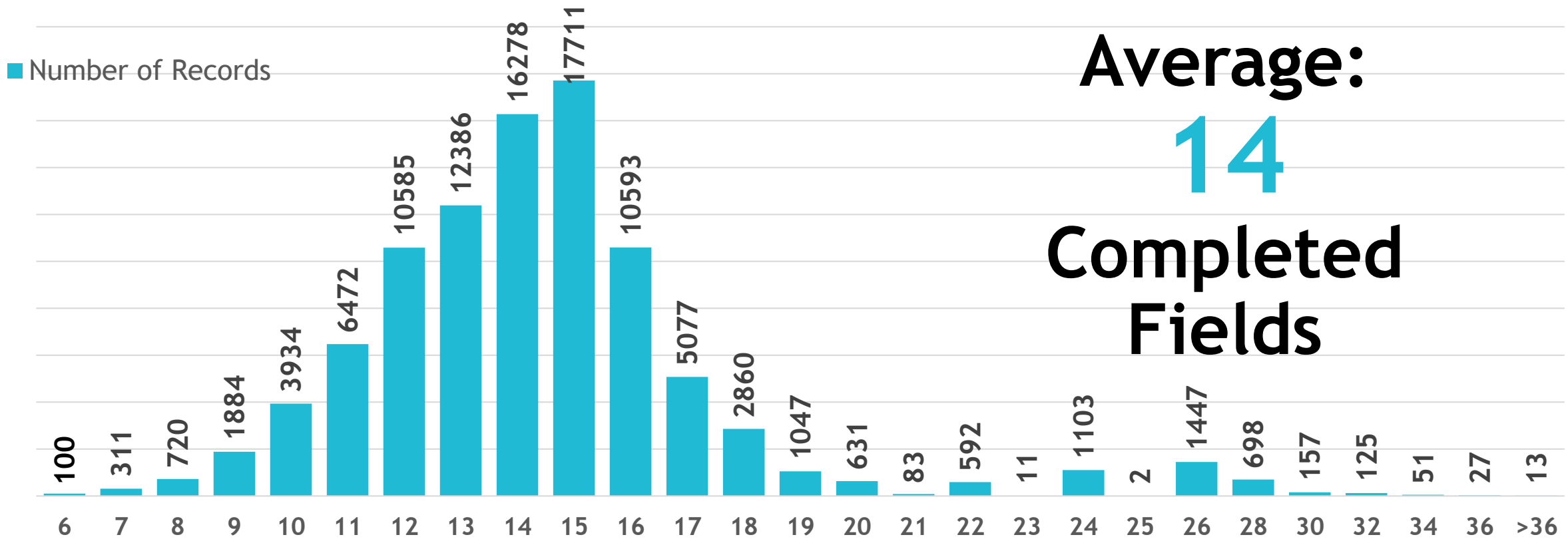
8 Required Fields

Findings: Metadata Fields - Journal Article



20 Optional Fields

Findings: Completed Fields - Journal Publications



Average:

14

Completed Fields

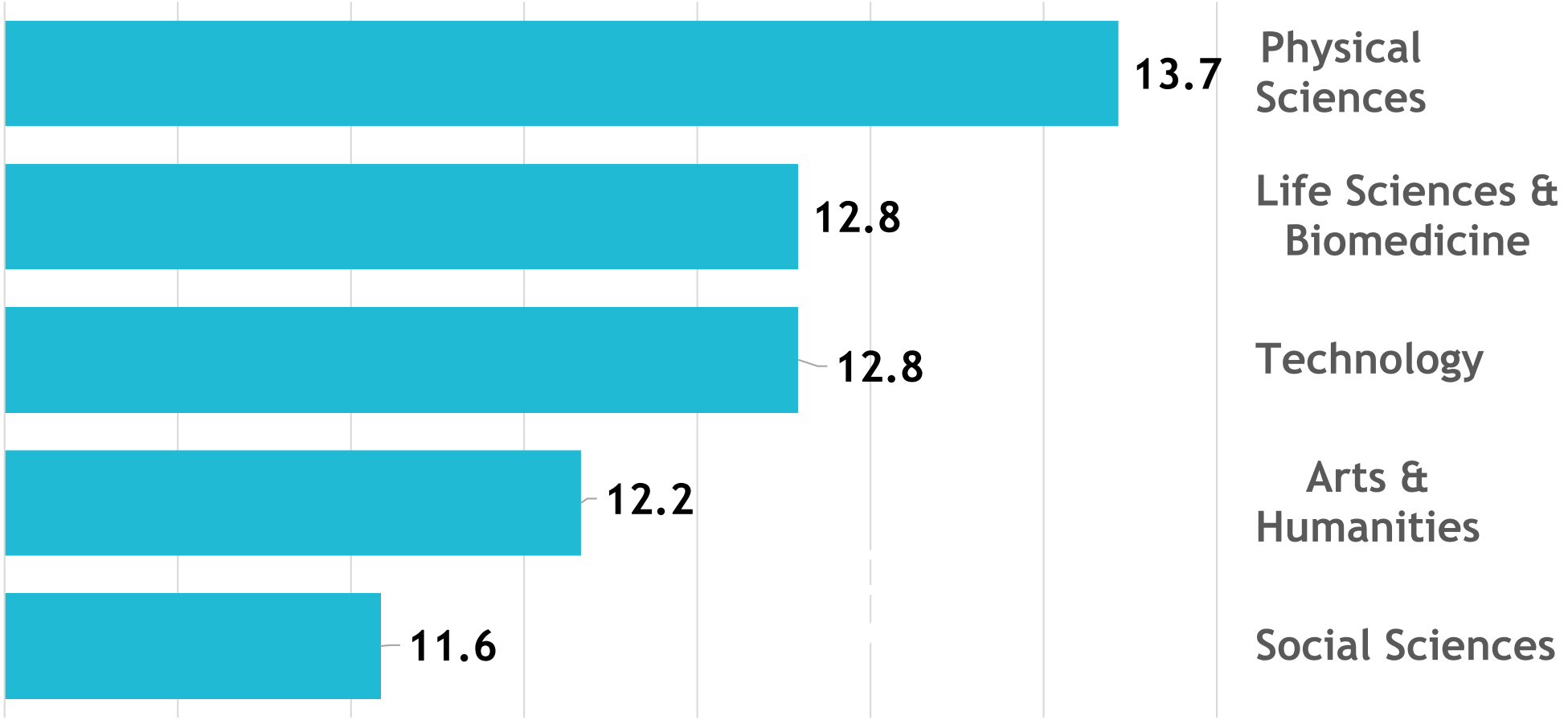
Findings: Average Completed Fields by Discipline

Test Statistics[a,b]

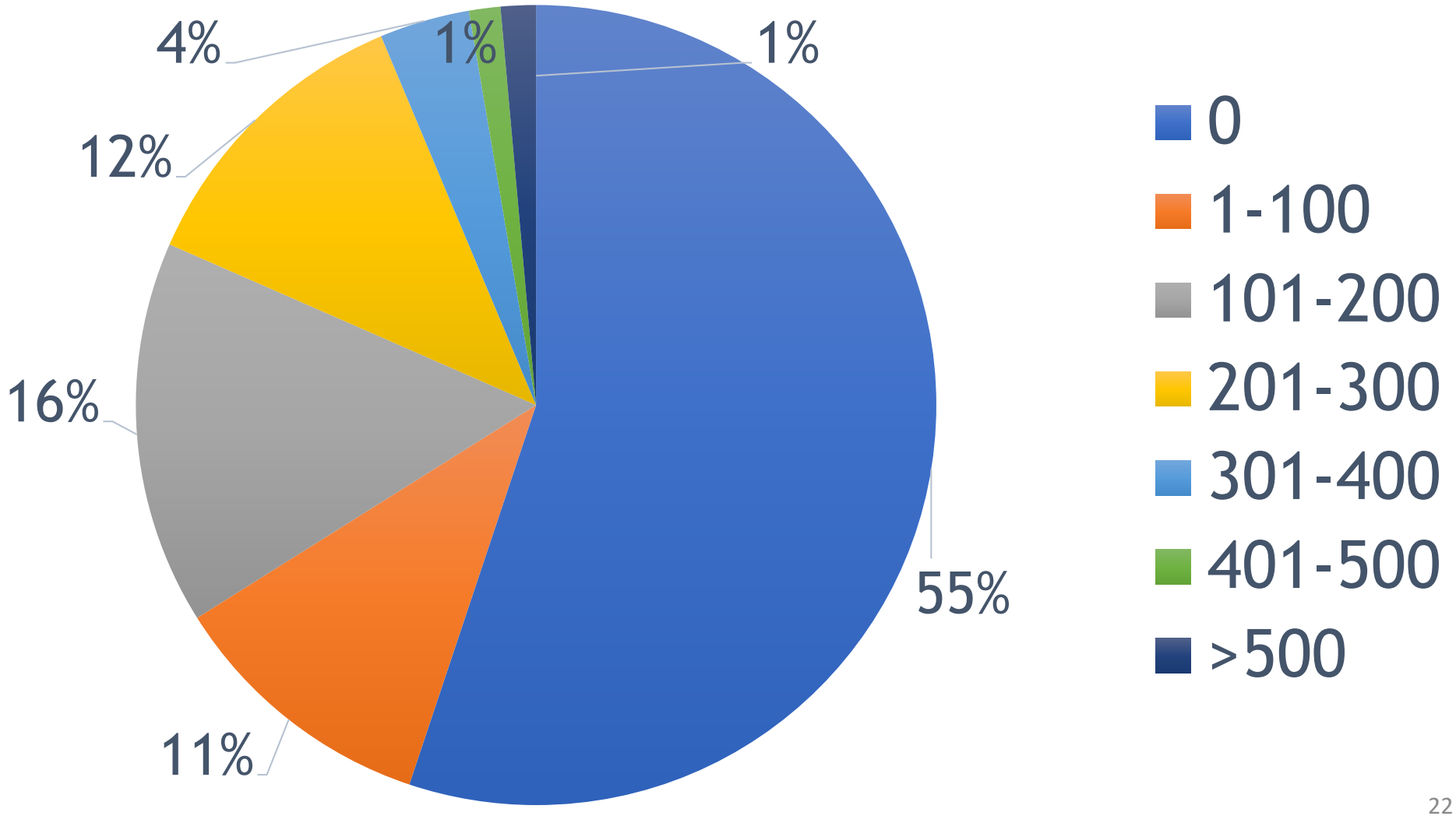
	Metadata_count
Kruskal-Wallis H	8513.546
df	4
Asymp. Sig.	.000

a. Kruskal Wallis Test

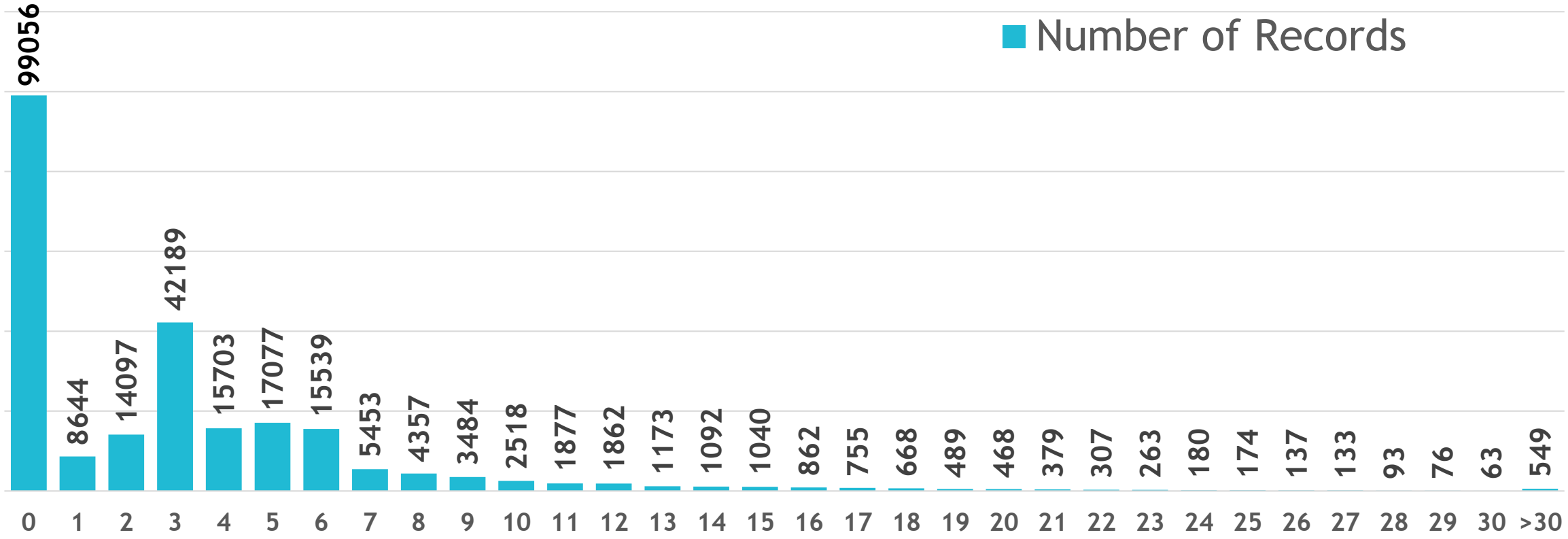
b. Grouping Variable: Categories



Findings: Abstract Word Count



Findings: Number of Keywords



Number of Keywords

Normalizing Visibility and Download

Normalized Visibility Number:
view-based measure of the visibility of one record
relative page view performance of a publication when compared to similarly-aged publications in its research category
number of views divided by the geometric mean of views for similarly-aged publications in the same research category (normalization of the number of page views across different research categories and years)

Normalized Download Number:
download-based measure of the visibility of one record and it indicates
relative download performance of a publication when compared to similarly-aged publications in its research category
number of downloads divided by the geometric mean of downloads for similarly-aged publications in the same its research category (normalization of the number of downloads across different research categories and years)

Average of views and downloads

Row Labels	Average of Normalized View	Average of Normalized Download	Average of Number of Views	Average of Number of Downloads
Arts & Humanities	1.45	2.91	94.08	99.11
Life Sciences & Biomedicine	1.59	5.43	112.15	182.84
Physical Sciences	1.49	3.36	80.85	105.82
Social Sciences	1.58	3.80	135.21	207.47
Technology	1.58	4.90	128.44	271.41

Correlation: Completed Fields and Normalized View

Correlations

			Metadata_count	NormalizedView
Spearman's rho	Metadata_count	Correlation Coefficient	1.000	.223**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedView	Correlation Coefficient	.223**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation: Completed Fields and Normalized Download

Correlations

			Metadata_count	NormalizedDownload
Spearman's rho	Metadata_count	Correlation Coefficient	1.000	.329**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedDownload	Correlation Coefficient	.329**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation: Completed Keywords and Normalized View

Correlations

			Keywords_Count	NormalizedView
Spearman's rho	Keywords_Count	Correlation Coefficient	1.000	.276**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedView	Correlation Coefficient	.276**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation: Completed Keywords and Normalized Download

Correlations

			Keywords_Count	NormalizedDownload
Spearman's rho	Keywords_Count	Correlation Coefficient	1.000	.278**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedDownload	Correlation Coefficient	.278**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation: Abstract Word Count and Normalized View

Correlations

			Abstract_Word_Count	NormalizedView
Spearman's rho	Abstract_Word_Count	Correlation Coefficient	1.000	.391**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedView	Correlation Coefficient	.391**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation: Abstract Word Count and Normalized Download

Correlations

			Abstract_Word_Count	NormalizedDownload
Spearman's rho	Abstract_Word_Count	Correlation Coefficient	1.000	.275**
		Sig. (2-tailed)	.	.000
		N	240757	240714
	NormalizedDownload	Correlation Coefficient	.275**	1.000
		Sig. (2-tailed)	.000	.
		N	240714	240714

**. Correlation is significant at the 0.01 level (2-tailed).

Findings

Out of the +20 optional metadata fields on ORBi:

- ~4 optional fields completed by researchers;
- The average completeness of fields across the records = 12,
- Metadata of publications on ORBi not fully complete, suggesting potential areas for improvement in metadata quality

Discussion and Recommendations

- Metadata as a fundamental pillar in the information-centric world
- Facilitating the efficient organization, retrieval, and utilization of information in line with FAIR principles
- Higher usability of IRs, requiring high-quality metadata for digital items
- Significant differences across disciplines in terms of completing metadata on ORBi
- More comprehensive metadata entry to be encouraged and more training needed
- Incorporating tools to improve metadata quality

Final Remark

- Key role of metadata completeness in enhancing user engagement on ORBi

- Metadata practices influencing visibility and discoverability of research outputs in line with Open Science Paradigm

Thank You

Behrooz Rasuli

Paul Thirion