
Open data in global environmental research: Findings from the community

Dr. Birgit Schmidt
Niedersächsische Staats- und Universitätsbibliothek
Göttingen

Delivering knowledge to support human action and adaptation to global environmental change



Established to foster global environmental change research;
Initiated June 2009 by NSF and NERC, building on the work
of the IGFA



(Courtesy Jialiang Gao www.peace-on-earth.org, Wikimedia Commons)

What is the Belmont Forum?

Funding Opportunities
Research Matching Registration
is Closed

News



E-infrastructures and Data Management CRA

“...the need to address global environmental challenges requires a more coordinated approach to the planning, implementation, and management of data, analytics and E-Infrastructures” through international collaboration.

Belmont Forum, New Delhi, February 2013



Work packages

1. Data Integration for Multidisciplinary Research
2. Improved Interface Between the Computation and Data Infrastructures
3. Harmonization of Global Data Infrastructure for Sharing Environmental Data
4. Data Sharing
5. **Open Data**
6. Capacity Building (spans Work Packages 1–5)

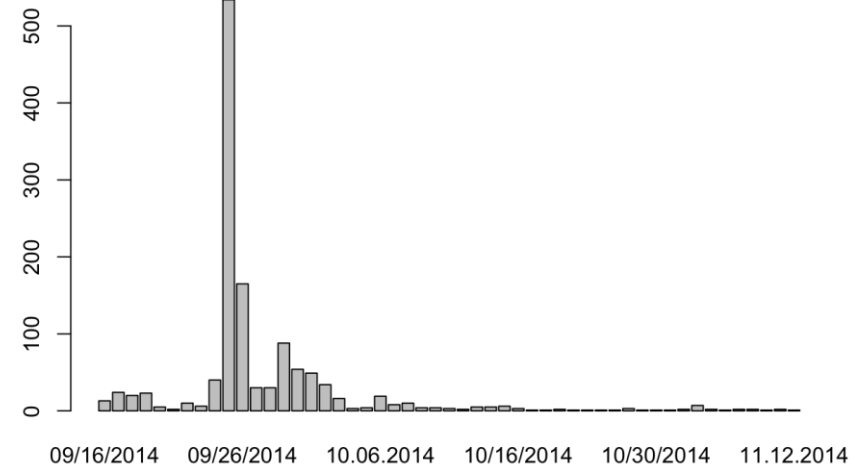
Open Data Survey

The main aim was to learn about

- Key open data initiatives in communities dealing with global environmental change
- Reasons where users' desire to share can be enhanced
- Barriers to open data sharing from a user perspective

Dissemination of the survey

- c 20 disciplinary mailing lists (geosciences, life sciences, social sciences, etc.)
- Copernicus Publications (2014/09/25)

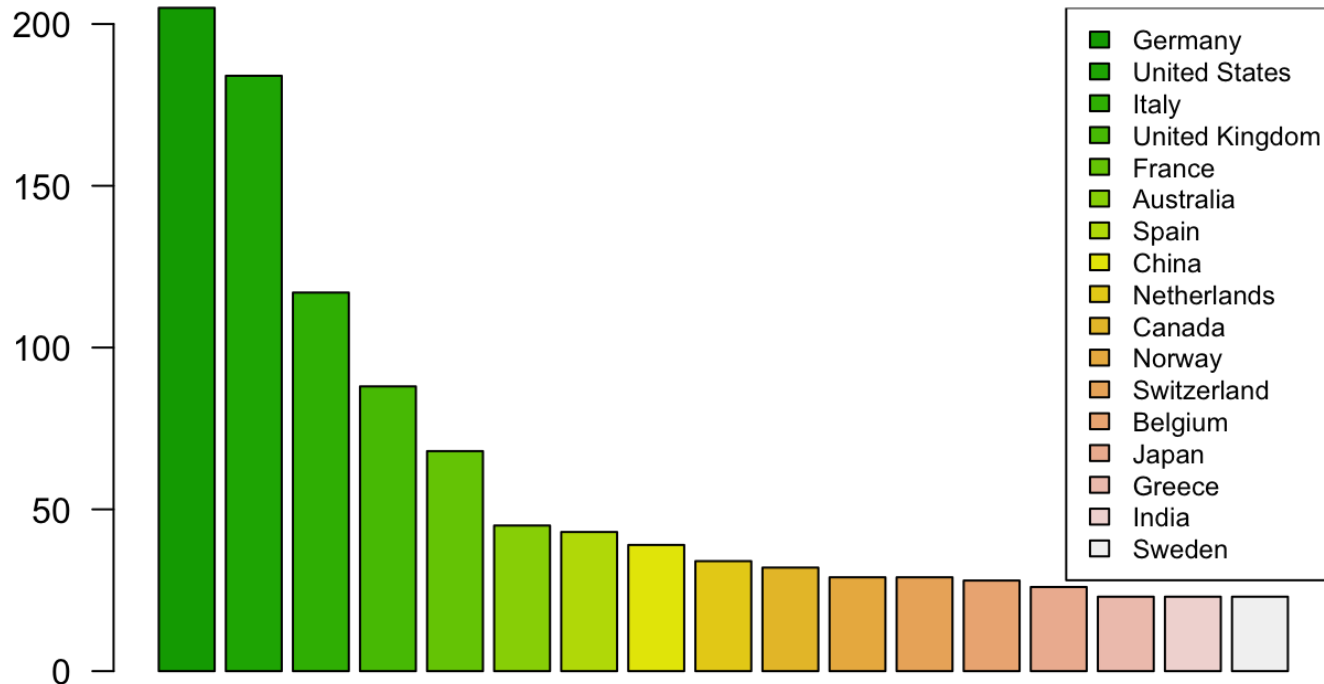


Open Data Survey: Responses

- 16 September – 12 November 2014
- 1330 responses (1253 used for analysis) from 80 countries
- 70.1% (878) academic, 17.9% (224) government, 5.6% (70) non-profit institutions, 2.6% (32) business, 0.3% (4) media, 3.6% (45) other roles
- 82.3% (1025) data users, 57.6% (718) data providers, 25.3% (315) data managers (multiple answers allowed), 5.3% (66) other roles
- Data published at: doi: 10.5281/zenodo.16384
- Schmidt B, Gemeinholzer B, Treloar A (2016) Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. PLoS ONE 11(1): e0146695. doi:10.1371/journal.pone.0146695

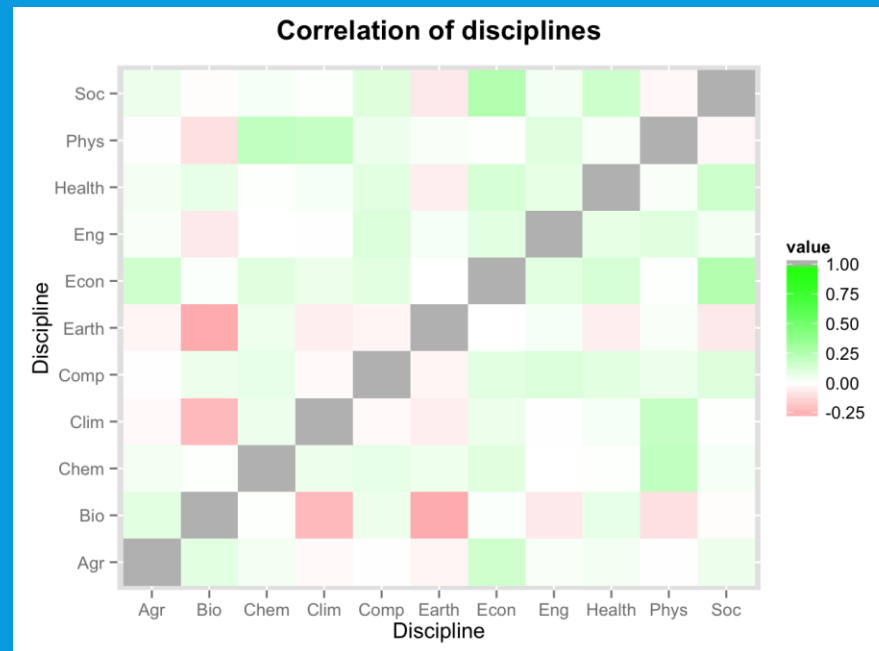
Countries with more than 20 answers

Which country are you from?

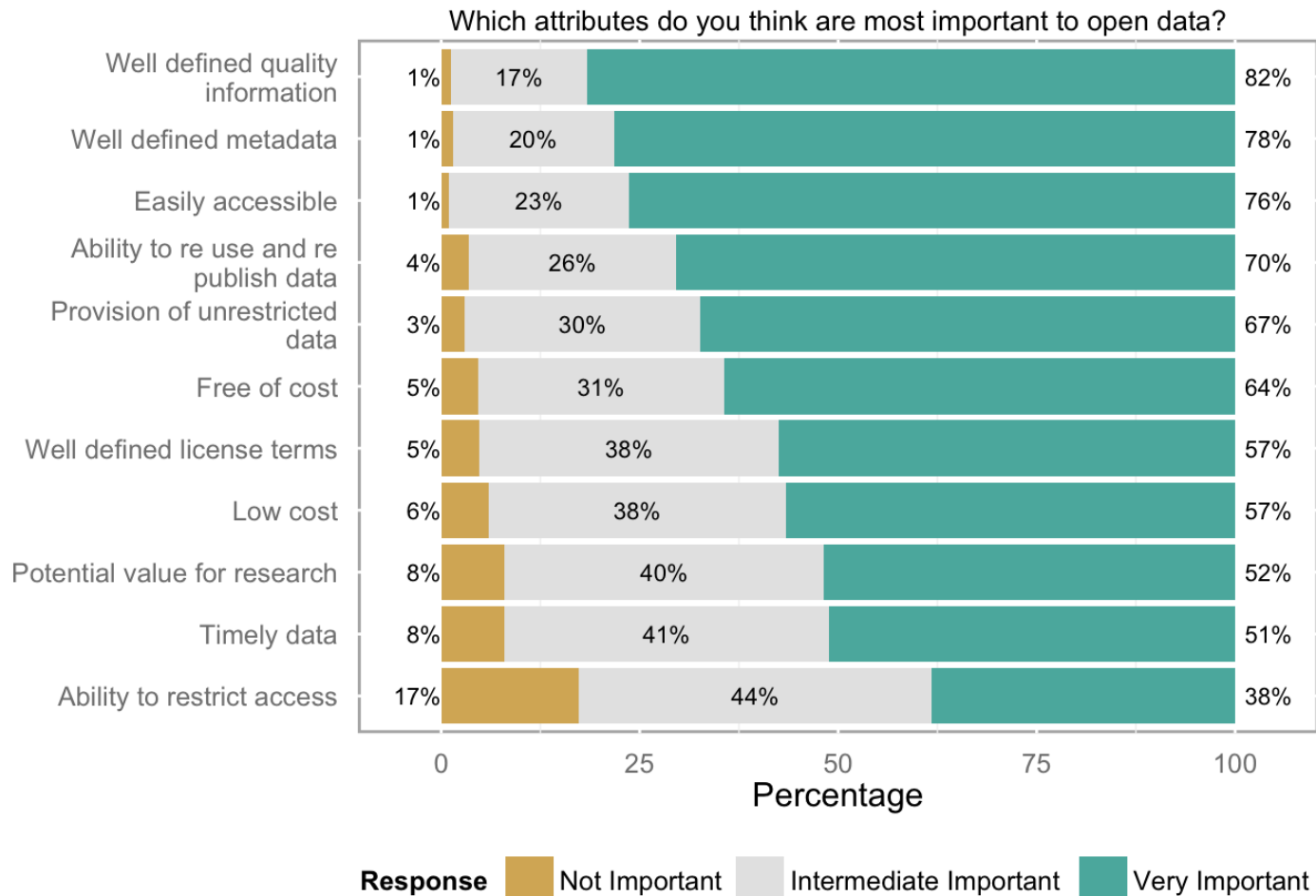


Responses by discipline (N=1253, multiple answers allowed)

- earth and environmental sciences (68.7%, 846 answers)
- climate and atmospheric sciences (31.3%, 386 answers)
- biological sciences (20.6%, 258 answers)
- physical sciences (12.9%, 162 answers)
- engineering (7.1%, 88 answers)
- computer sciences (6.9%, 85 answers)
- social sciences (5.4%, 66 answers)
- agricultural and veterinary sciences (4.3%, 53 answers)
- chemical sciences (4.1%, 50 answers)
- other discipline (3.2%, 40 answers)
- health sciences (1.8%, 22 answers)
- economics (1.7%, 21 answers)

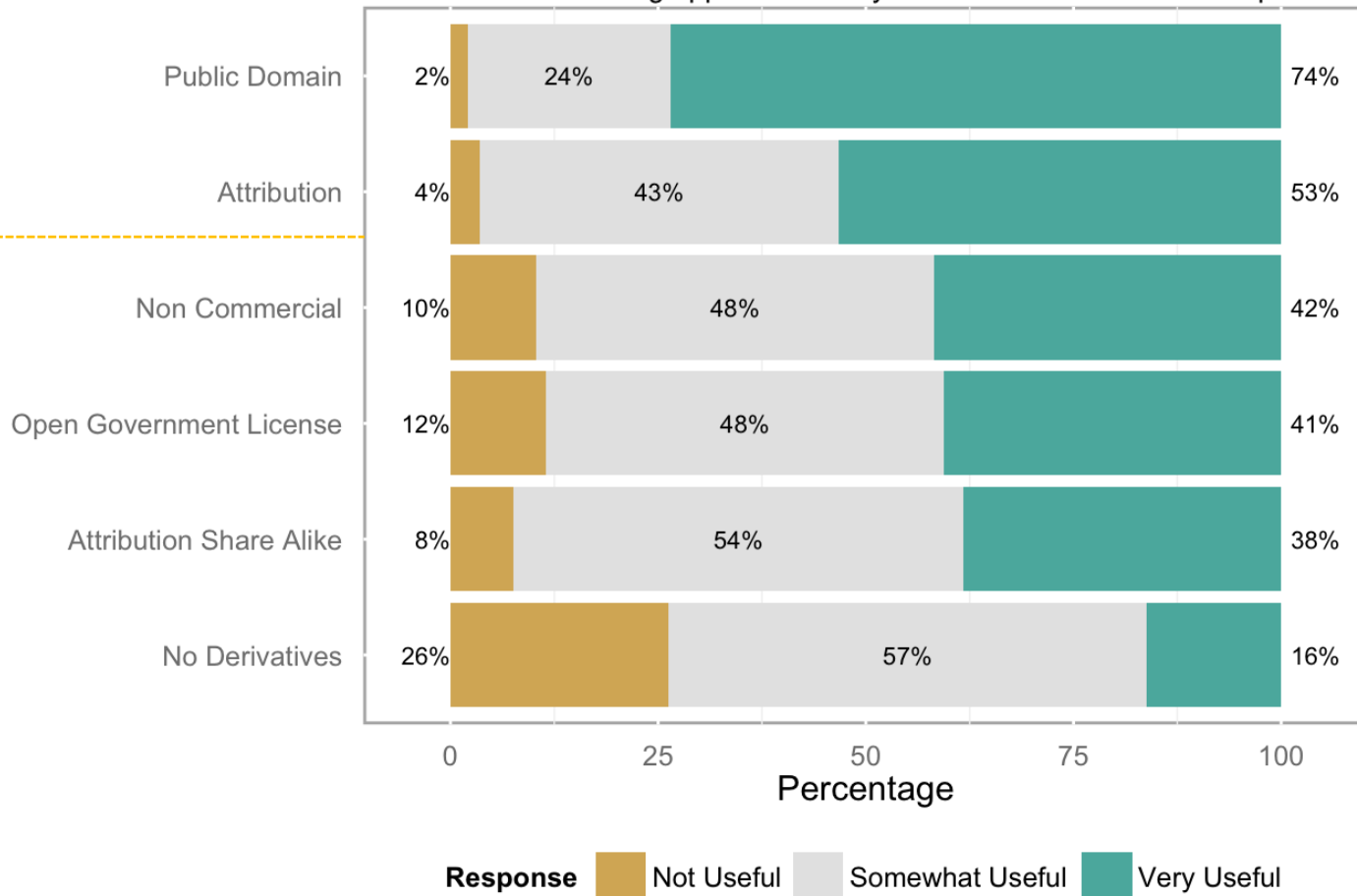


What properties do they expect for open data?



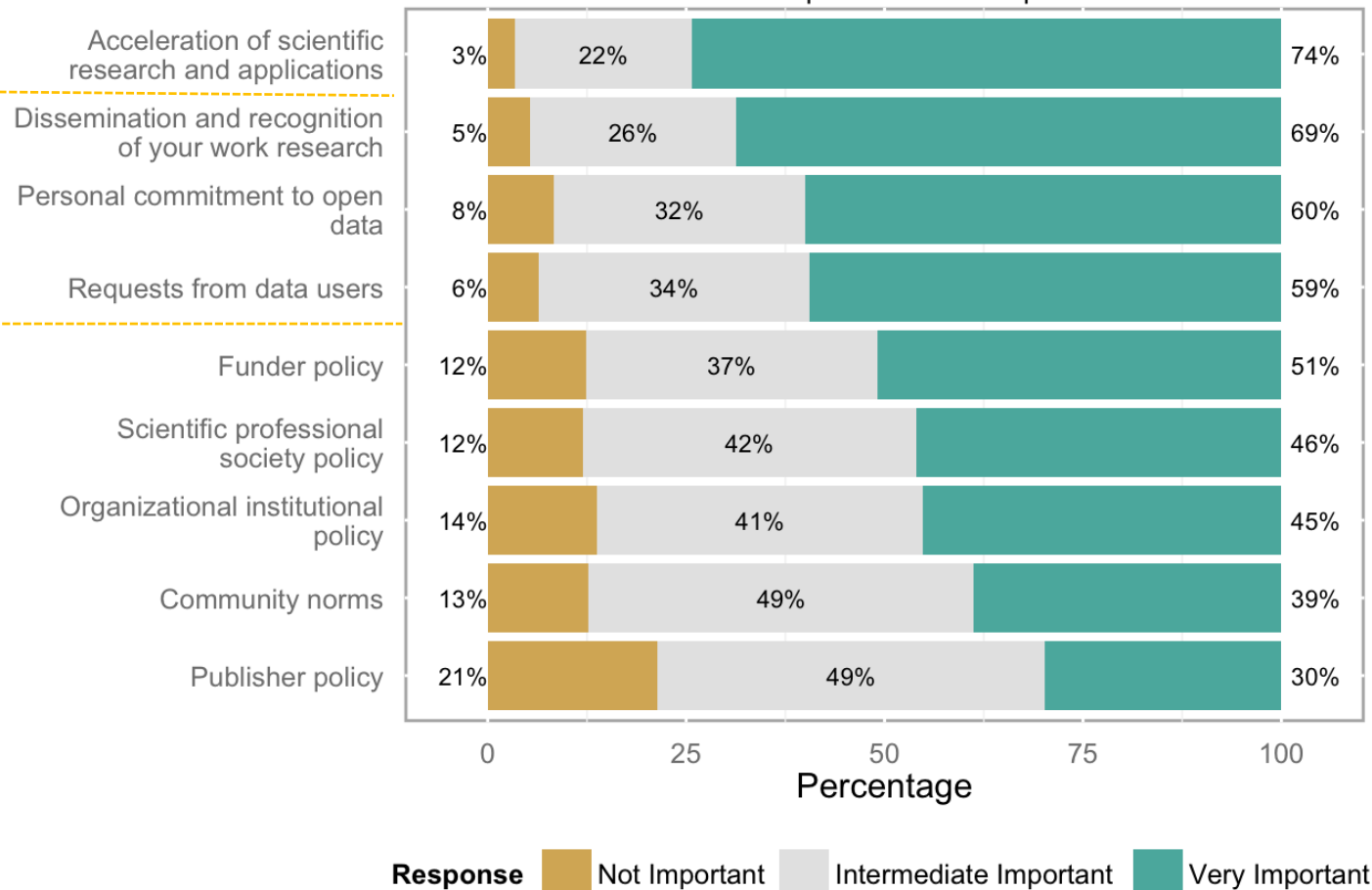
Access and licensing conditions

What access or licensing approaches do you consider most useful for open data?



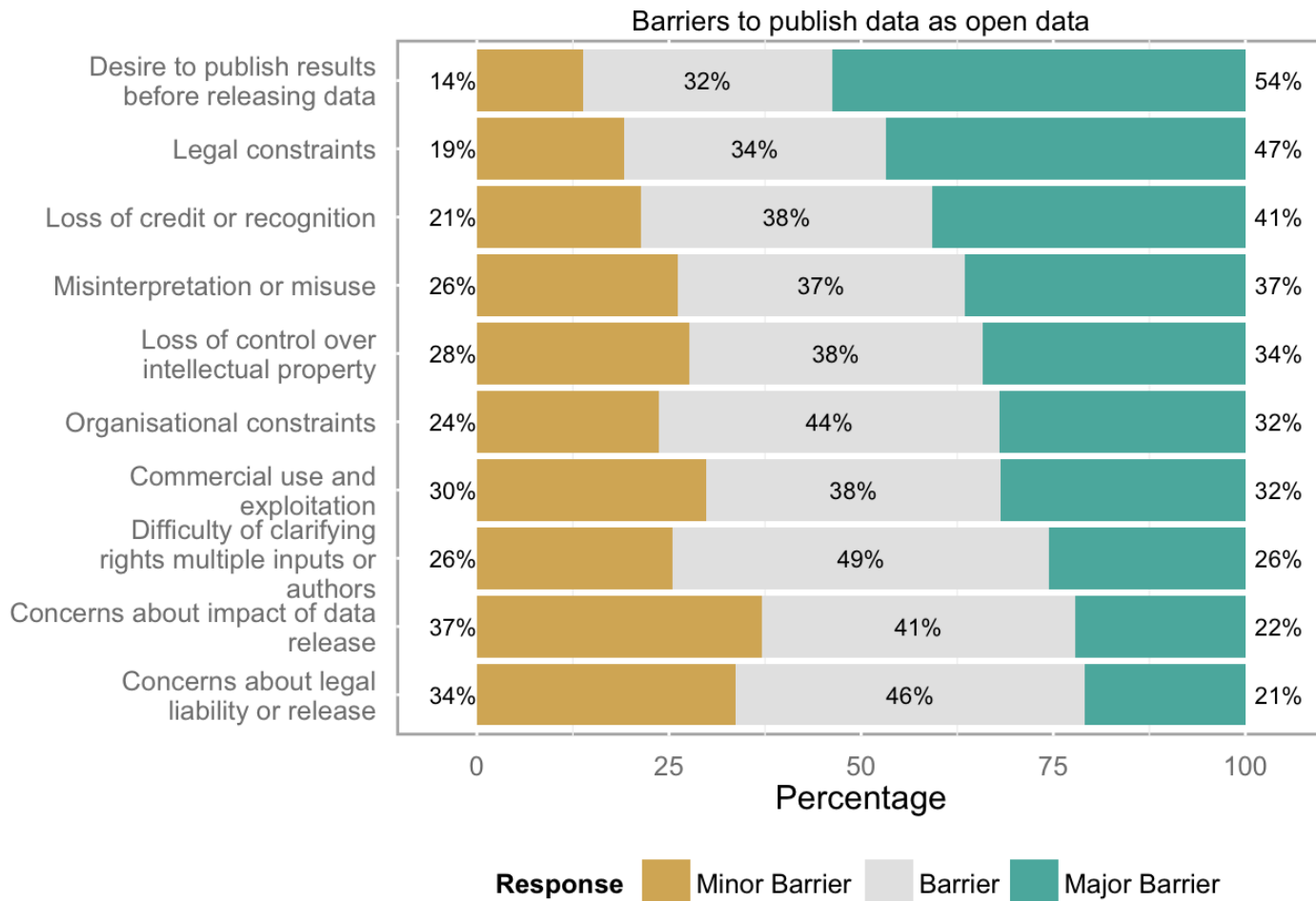
Motivators to publish data as open data

Motivators to publish data as open data

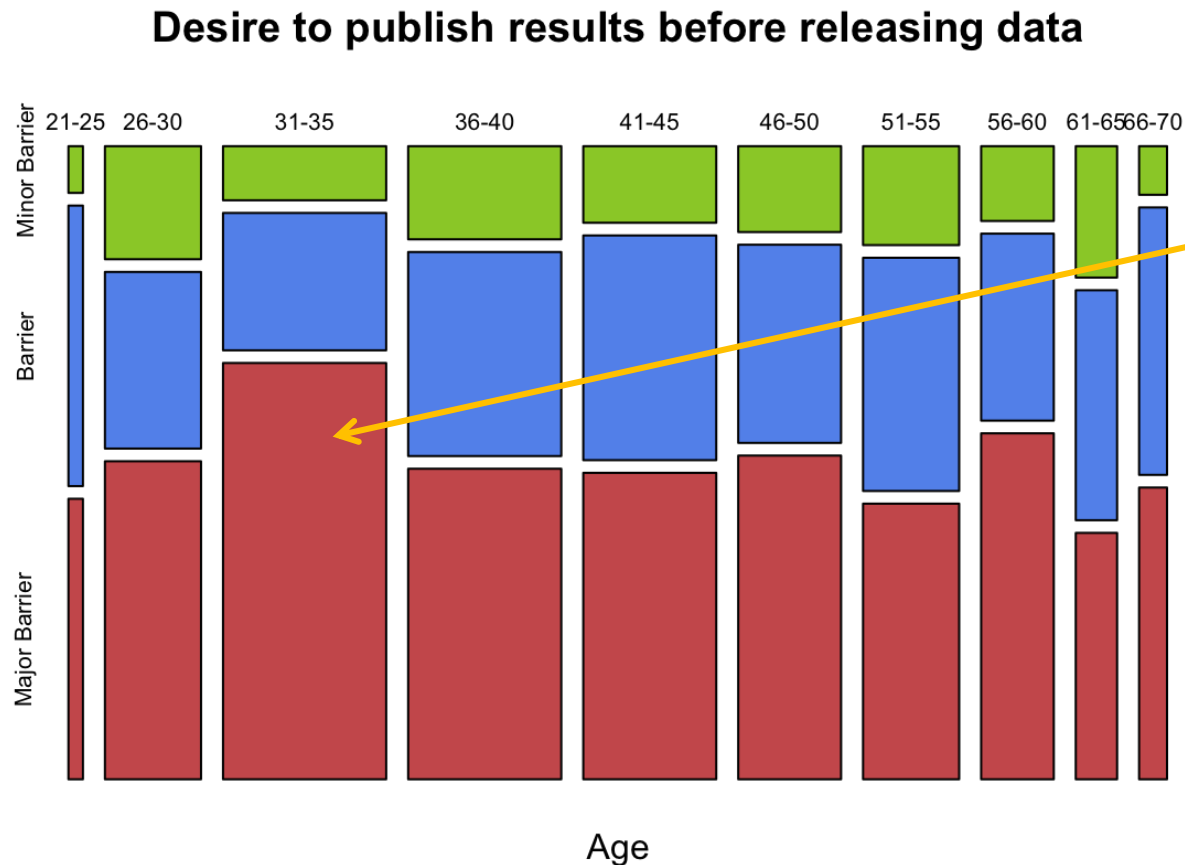


Data managers who contributed to the survey were significantly more committed to open data than all other data professionals ($p < 0.001$).

Barriers to publish data as open data

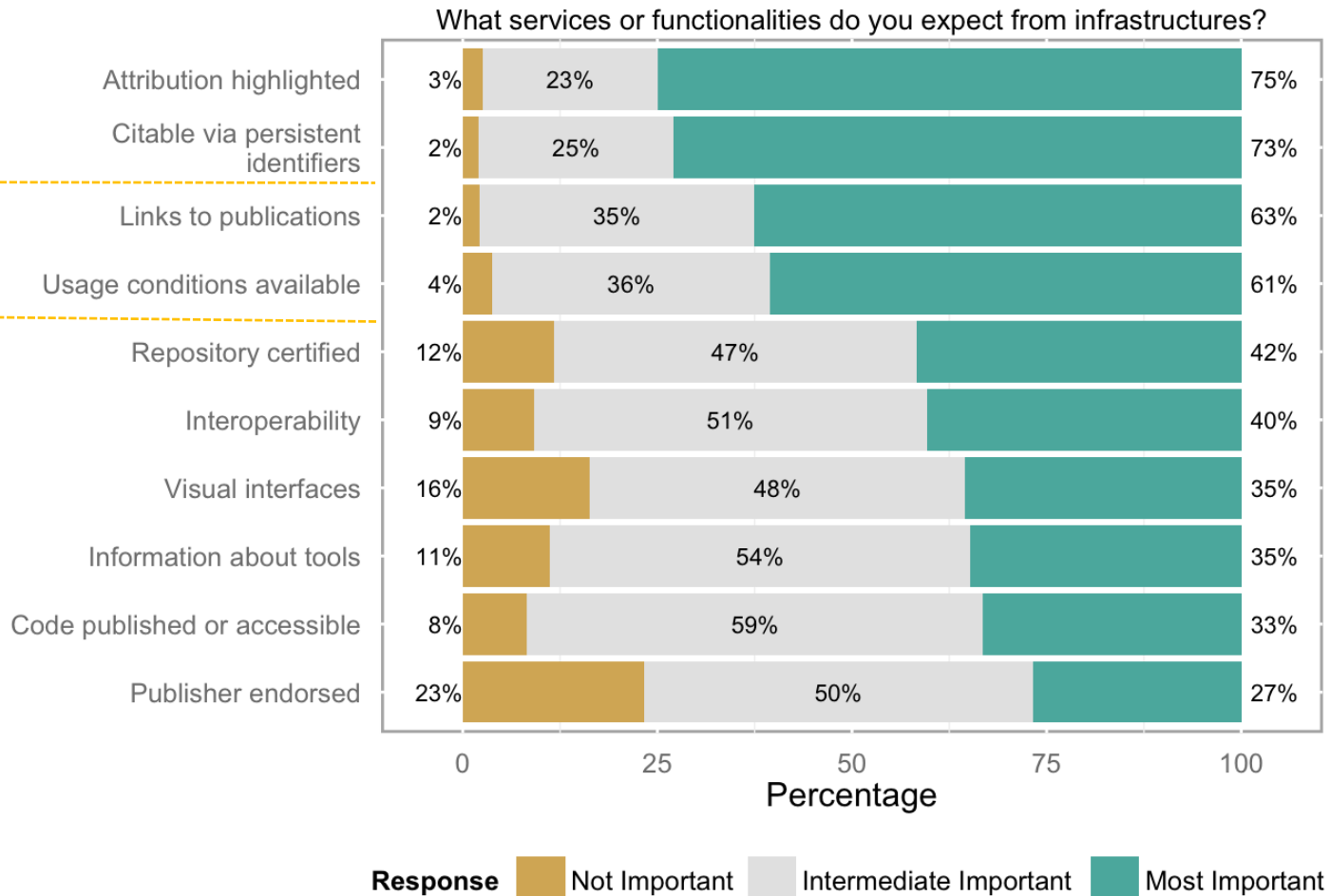


Some barriers relate with age / career stage



31–35 year-olds expressed a significantly higher desire to publish results before releasing data.

Infrastructure expectations



Discovery of data

How do you normally discover data?



Discovery route	Frequency	Percentage
References in journal articles	622	79.8
Web search engines	549	70.5
Searching in specific data repositories	492	63.2
Direct requests to data providers	314	40.3
Newsletters or other publications	201	25.8
Government or institutional announcements	164	21.1
Directories or catalogs	136	17.5
Social media	70	9.0
Blogs	63	8.1
Other discovery	36	4.6

Table shows frequencies and valid percentages for each discovery route, multiple answers were allowed (n = 779).

What do they value in data repositories?

„Good for heterogenous datasets, also long-tail“

„easy to obtain quality data“

„great to have automatic DOIs“

„independent, trustful,
recommended by several publishers“

„well-defined data quality
and metadata“

„large data files, managing
copies close to computing“

„visualization
interface“

„The site and tools are terrible,
but the data is important“

„sound international reputation“

„highest quality with excellent links to other databases“

What is missing?

- “For me, **long-tail research datasets** are most important, and I would therefore like to see more repositories supporting these (i.e. institutional ones)”
- “Every discipline in the natural sciences needs to be able to **access its heritage data** (those in analogue forms) which cannot presently be accessed electronically (observations are still, and only, in their virgin forms on paper, forms, books, photographs, unreadable early mag tapes, etc.).”
- “Some types of data **still lack community standards** that would allow creation of open data resources.”
- “the open publication of **source codes or scripts** of simulation programs or analytical / numerical solutions is extremely important”

Conclusions & recommendations

Based on the findings of the survey, we have made the following recommendations to the Belmont Forum:

- that funders should make open data archiving mandatory, to take into account the main motivators revealed by the survey,
- to strengthen support and training activities,
- to further facilitate interoperability between data infrastructures, and
- to support the long-term sustainability of archives and data infrastructures.

Thank you for your attention!

www.bfe-inf.org

Acknowledgements:

- To my co-authors Birgit Gemeinholzer, University of Giessen, and Andrew Treloar, ANDS
- Slides 2-3 by Belmont Forum colleagues
- All members of the Open Data WG

Birgit Schmidt, University of Göttingen, bschmidt@sub.uni-goettingen.de