

Findable and Reusable?

A sociotechnical perspective on data search and reuse

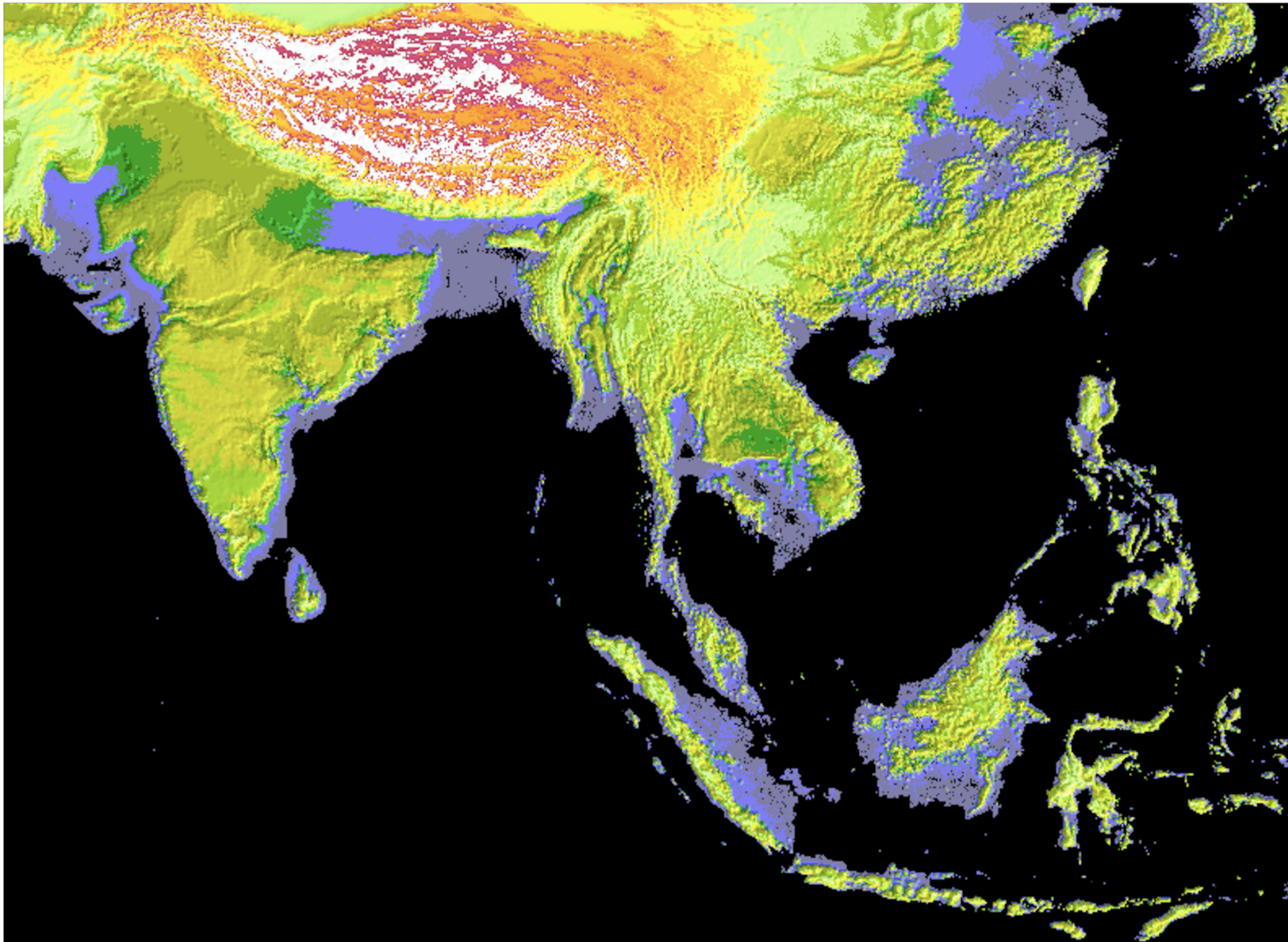
Kathleen Gregory

M.A., MSLIS, PhD candidate

Open Science Conference

March 20, 2019

A tale of searching data



Structure

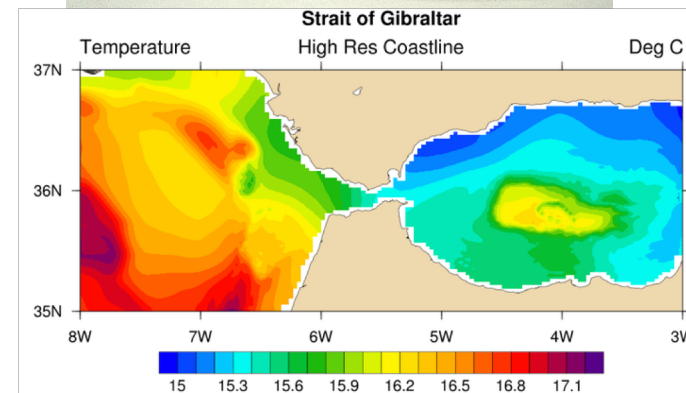
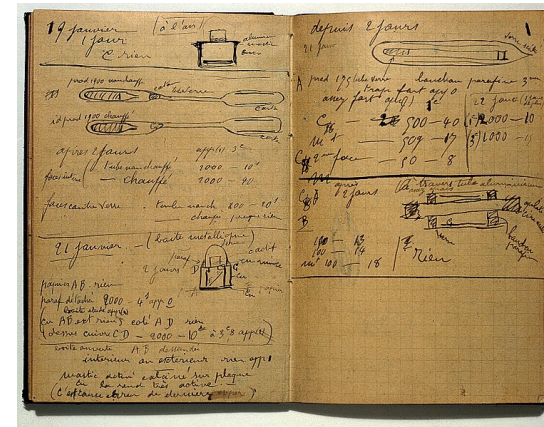
- Findable?
- Reusable?
- Applications and conclusions

Data are complex objects

Data are diverse.

Data do not stand alone.

Data are not always stable
and do not travel easily.



Borgman, C.L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press.

Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. *Science, Technology, & Human Values*, 42(2), p.191-202.

“Openness” is also not simple

- Technological, organizational and moral dimensions dominate policy discourse (Noorman, et al., 2017)
 - Actual practice does not always match how “openness” is portrayed in policy documents
- Dynamic practice of opening/closing (Levin & Leonelli, 2017)
 - What is open for whom, when and how?

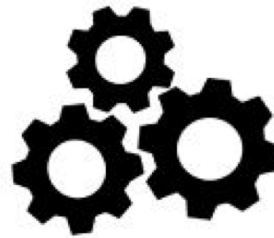
F
indable



A
ccessible



I
nteroperable



R
eusable



F
indable



A
ccessible



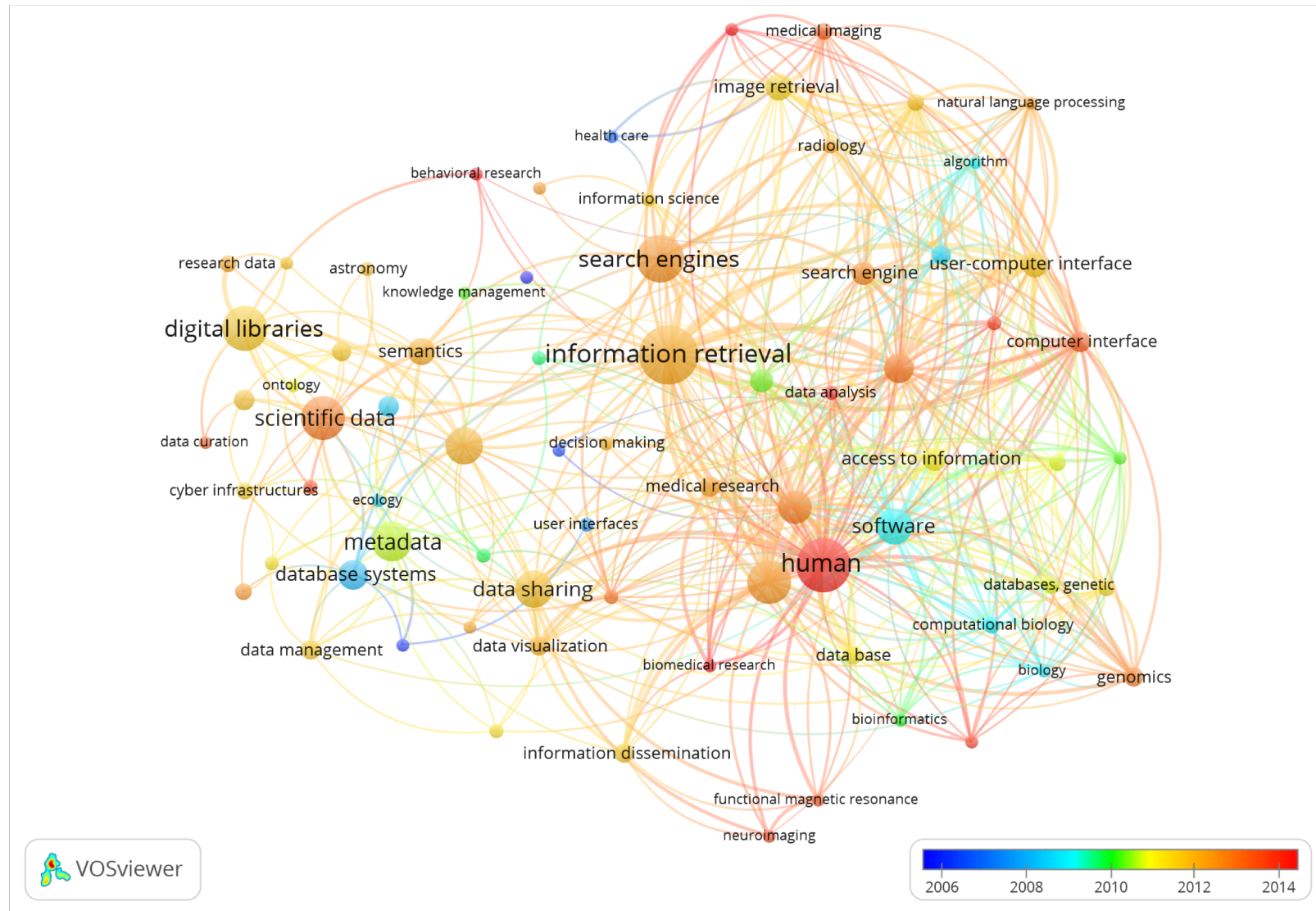
I
nteroperable



R
eusable

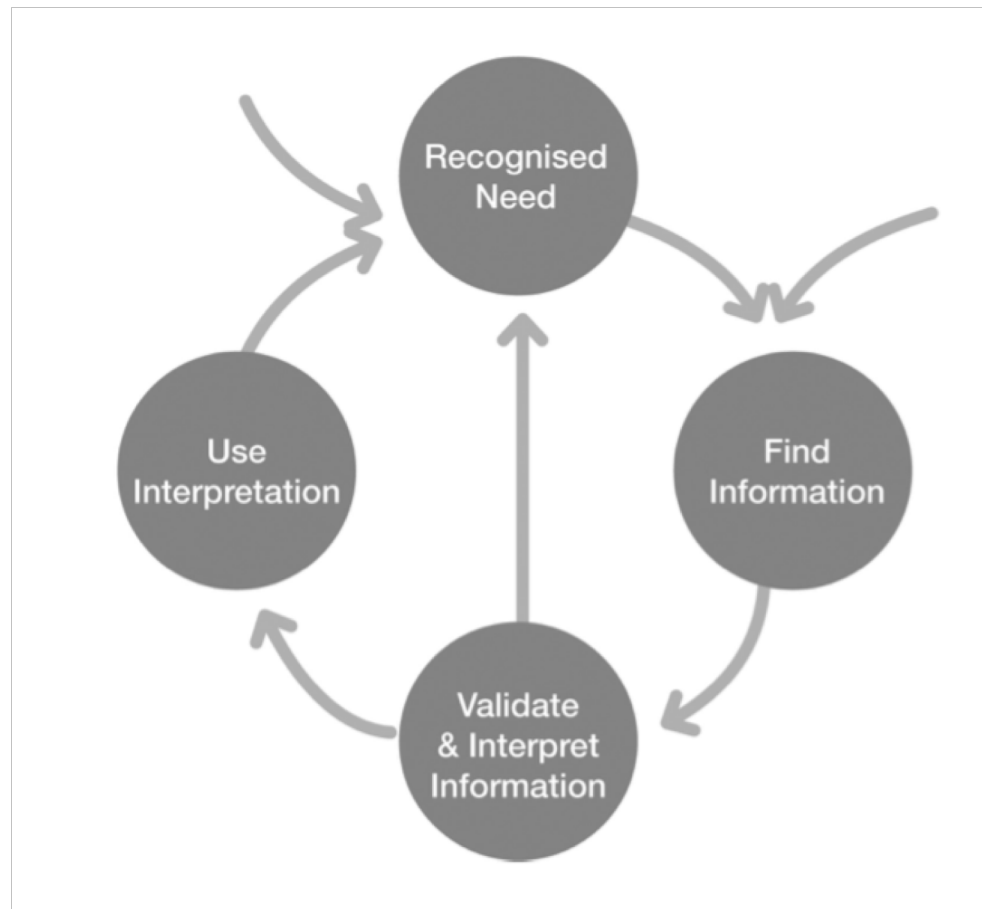


Data search – technical focus



Exploring the socio-technical

Starting point: frameworks of interactive information retrieval



Methodology

- **Primary:** Semi-structured interviews with data seekers across disciplines (n=22)
- **Next stage:** Multidisciplinary survey (n=1677, still in analysis phase)

Users and data needs

A broader understanding of the user is needed.

- Heterogeneity within disciplines
- The person seeking data is not always the person using the data

Users and data needs

A broader understanding of the data needed by users

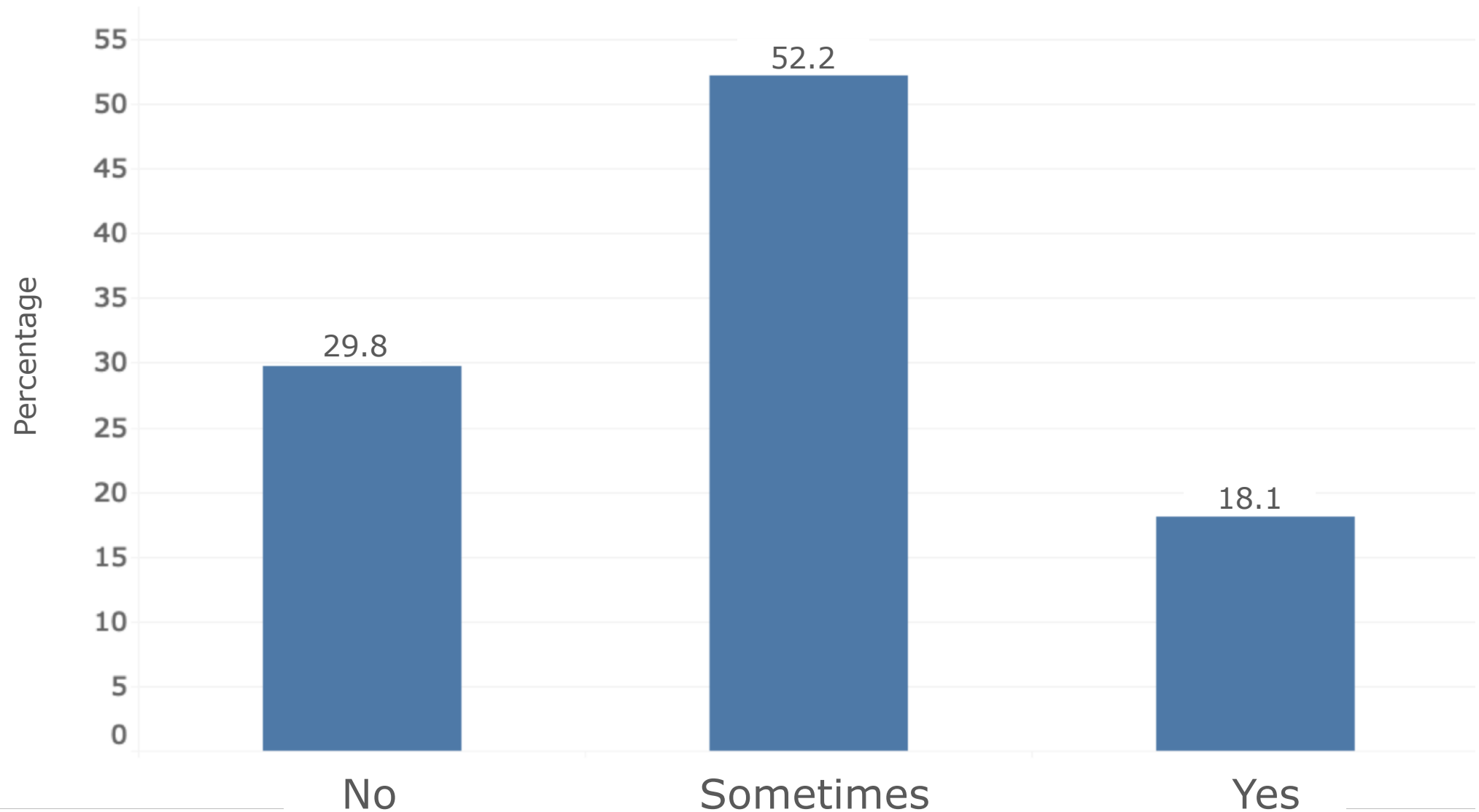
- Data needed for research are not always research data
- Numerous roles - data as hubs for collaboration and creativity

Search and discovery strategies

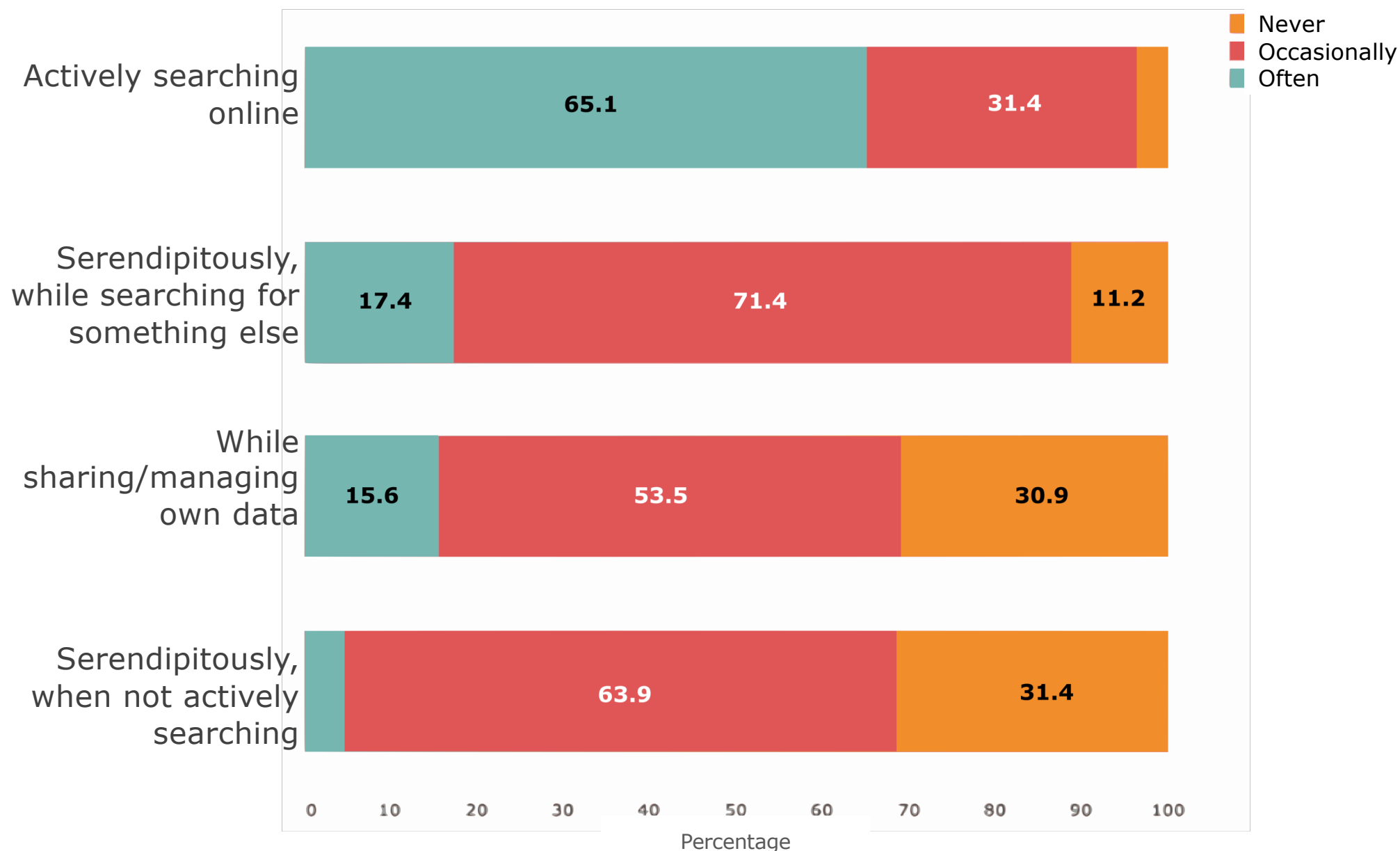
Multiplicity of strategies



Do you discover data differently than how you discover academic literature?



How frequently do you find data in the following ways?



Search and discovery strategies

Key role of social interactions

Actually, most of the times that I have looked for external data, it has been through (personal) connections (11).

The human network of contacts is still the best way to find the information you want, especially if it is a small group...that is the most powerful and accurate source of information that I use at this point. (17)

F
indable



A
ccessible



I
nteroperable



R
eusable



Evaluation and sense-making

Role of social interactions continues

*I think if there was a good search engine, then I could get the dataset directly. I would still get in touch with the data author anyway, both for social reasons - developing the network and eventual collaboration - and also **because most of the times the metadata are not enough to really understand the biology behind the species** (4).*

Evaluation and sense-making

Role of social interactions continues

*I am used to working with experts from different areas of knowledge. For me it is usual to have partners with different expertise: biology, agronomy, economy...I know the language of LCA (life cycle assessment), not of electronics or agricultural biology. **My limit is not the data that I cannot find, but people that can work with these data (16).***

Data search is a complex socio-technical process.

Data search is interwoven with other (re)search and data practices.

What does this mean for system design?

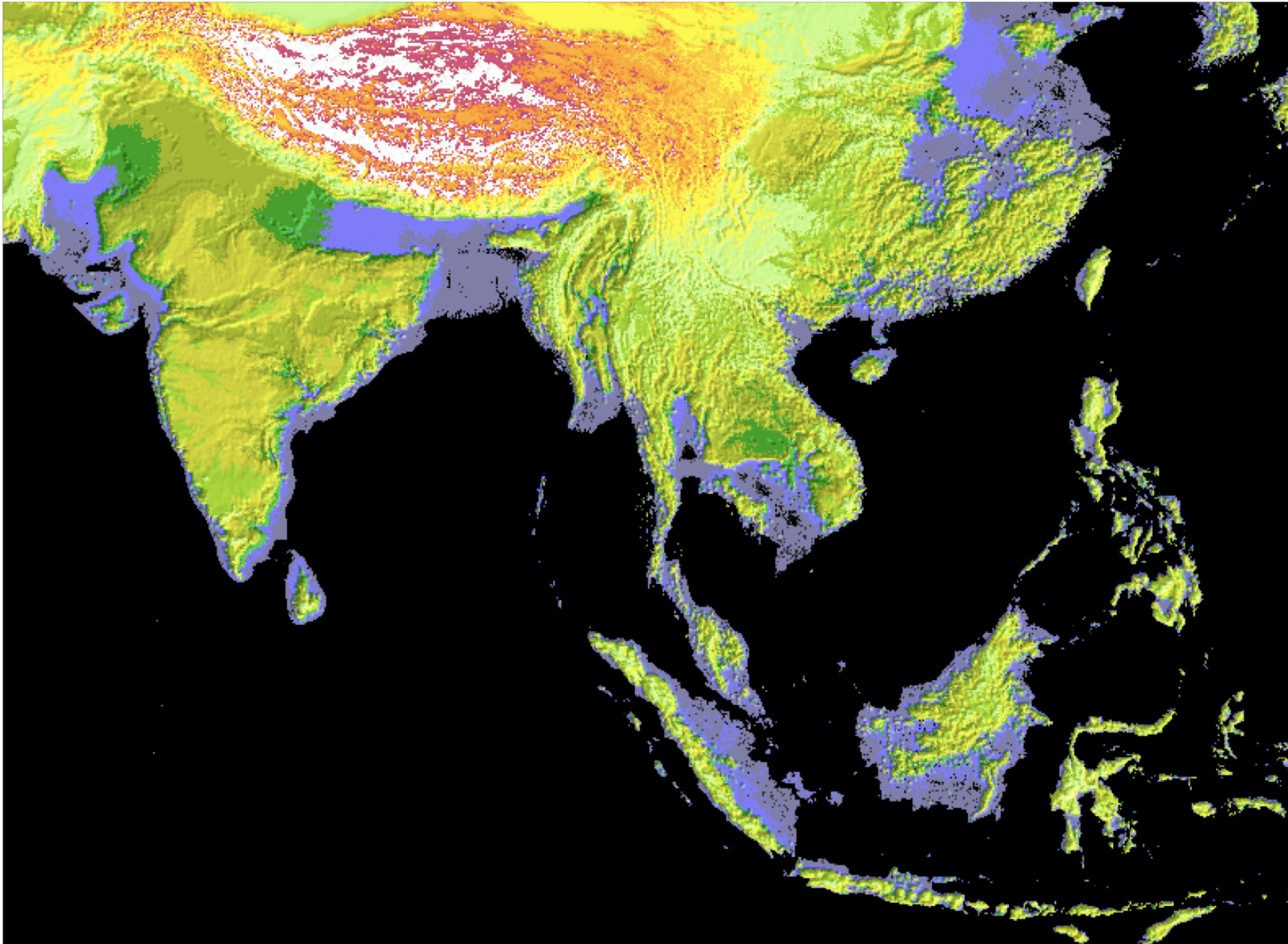
- **Consider how data are made available**
- **Consider entirety of data needs**
- **Consider diversity and overlaps**
- **Consider how to incorporate role of social interactions**

What does this mean for policy?

Make social nature of data practices explicit

- Importance of social communication in data discovery and reuse (and other practices)
- Interwoven nature of data practices with other (domain-specific) research practices
- Plurality of meanings of data, openness, FAIRness

A tale of searching data - revisited



References

- Blandford, A., & Attfield, S. (2010). *Interacting with Information. Synthesis Lectures on Human-Centered Informatics*. San Rafael, CA: Morgan & Claypool Publishers.
- Borgman, C.L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press.
- Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. *Science, Technology, & Human Values*, 42(2), p.191-202.
- Levin, N., & Leonelli, S. (2017). How does one “open” science? Questions of value in biological research. *Science, Technology, & Human Values*, 42(2), 280-305.
- Gregory, K.; Cousijn, H.; Groth, P.; Scharnhorst, A.; Wyatt, S. (forthcoming). Understanding Data Search as a Socio-technical Practice. *Journal of Information Science*. *arXiv preprint: arXiv:1801.04971*.
- Gregory, K.; Cousijn, H.; Groth, P.; Scharnhorst, A.; Wyatt, S. (2019). Searching data: A review of observational data retrieval practices in selected disciplines. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24165>
- Noorman, M., Wessels, B., Sveinsdottir, T. and Wyatt, S. (2018) Understanding the ‘open’ in making research data open: policy rhetoric and research practice. In: Saetan, A.R., Schneider, I. and Green, N. (eds.) *The Politics and Policies of Big Data: Big Data, Big Brother?* Taylor & Francis (Routledge), pp. 292-318.

Image credits (in order of appearance)

- NASA. Screenshot of map of Aceh area. <https://search.earthdata.nasa.gov/search>.
- Wellcome Images. [Marie Curie: Holograph Notebook](#). Under [CC BY 4.0](#). (on Slide 3: Data are complex objects).
- National Center for Atmospheric Research. [Example of Coastlines Graphics: Straits of Gibraltar](#). (on Slide 3: Data are complex objects.)
- SangyaPundir. [FAIR guiding principles for data resources](#). Wikimedia. Under [CC BY-SA 4.0](#). (on Slide 5)
- All other images available from pixabay.com under the [Pixabay License](#) and are free for commercial use, no attribution required.

Funding

This work is part of the project *Re-SEARCH: Contextual Search for Research Data* and was funded by the NWO Grant 652.001.002.



Questions?

Kathleen Gregory

kathleen.gregory@dans.knaw.nl