Data Archiving and Networked Services

DANS

Discovering Data:
Practices of Data Search and Reuse
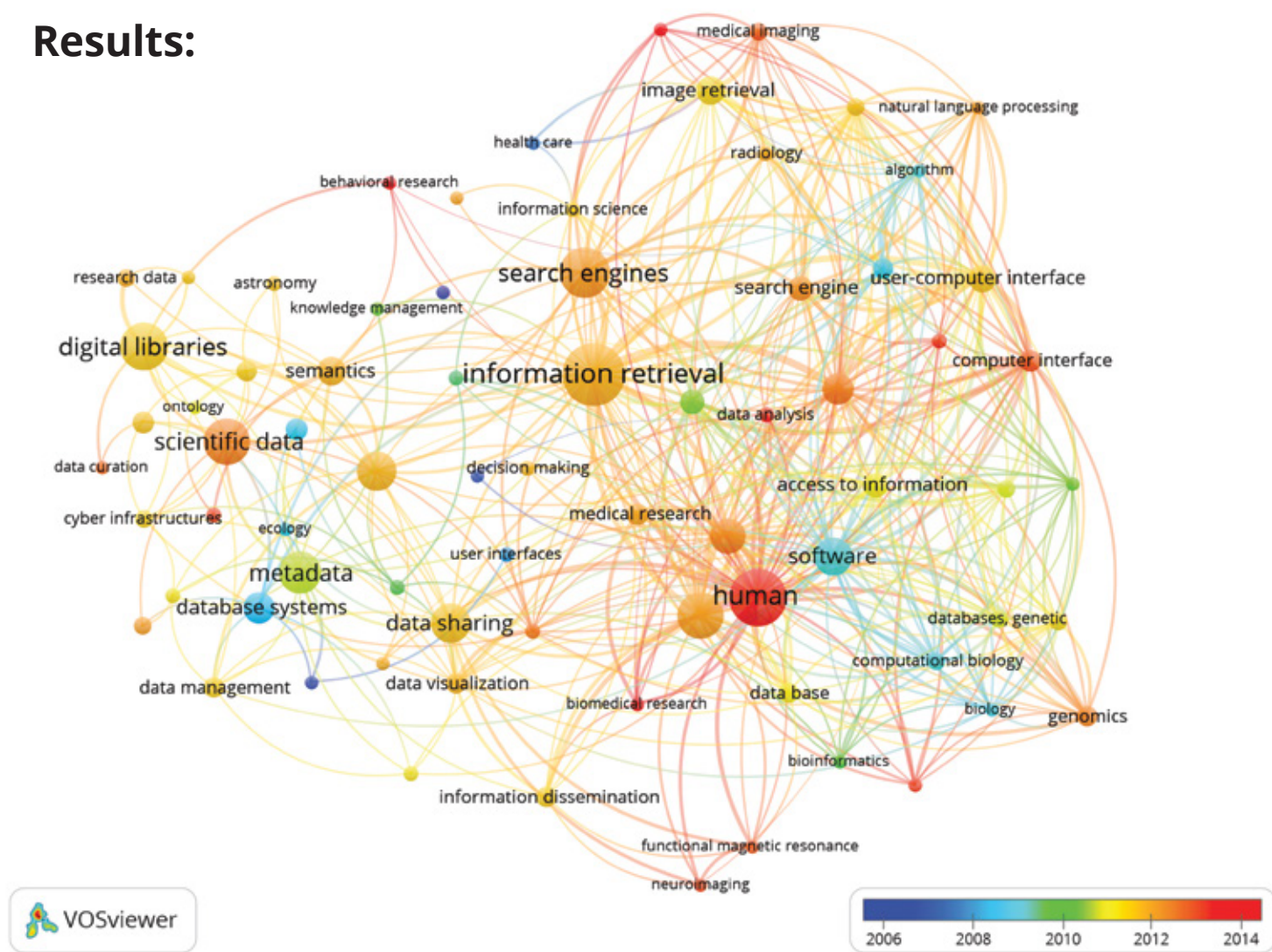
Kathleen M. Gregory

Introduction:

One of the promises driving both data sharing and policy development is that shared research data can and will be discovered and reused. Exactly how researchers engage in data discovery remains a topic that is under-explored. My research examines practices of data search and is centered around the following questions:

- How do people discover, make sense of, and (re)use research data that they do not create themselves?

- How are these practices related to community norms, infrastructures and data supplies?

- How can these insights inform the design of search solutions for research data?

Methods:

- Literature survey of data search practices
  - Keyword searches related to information retrieval and data practices, data search and discovery, primarily in Scopus

- Semi-structured interviews with data seekers across disciplines (n=22)
  - Interview protocol developed based on literature survey and consultation with data discovery systems designers from Elsevier Data Search

- Cross-disciplinary survey of authors indexed in Scopus (n=1677)
  - Majority of respondents were researchers; librarians/data professionals also responded
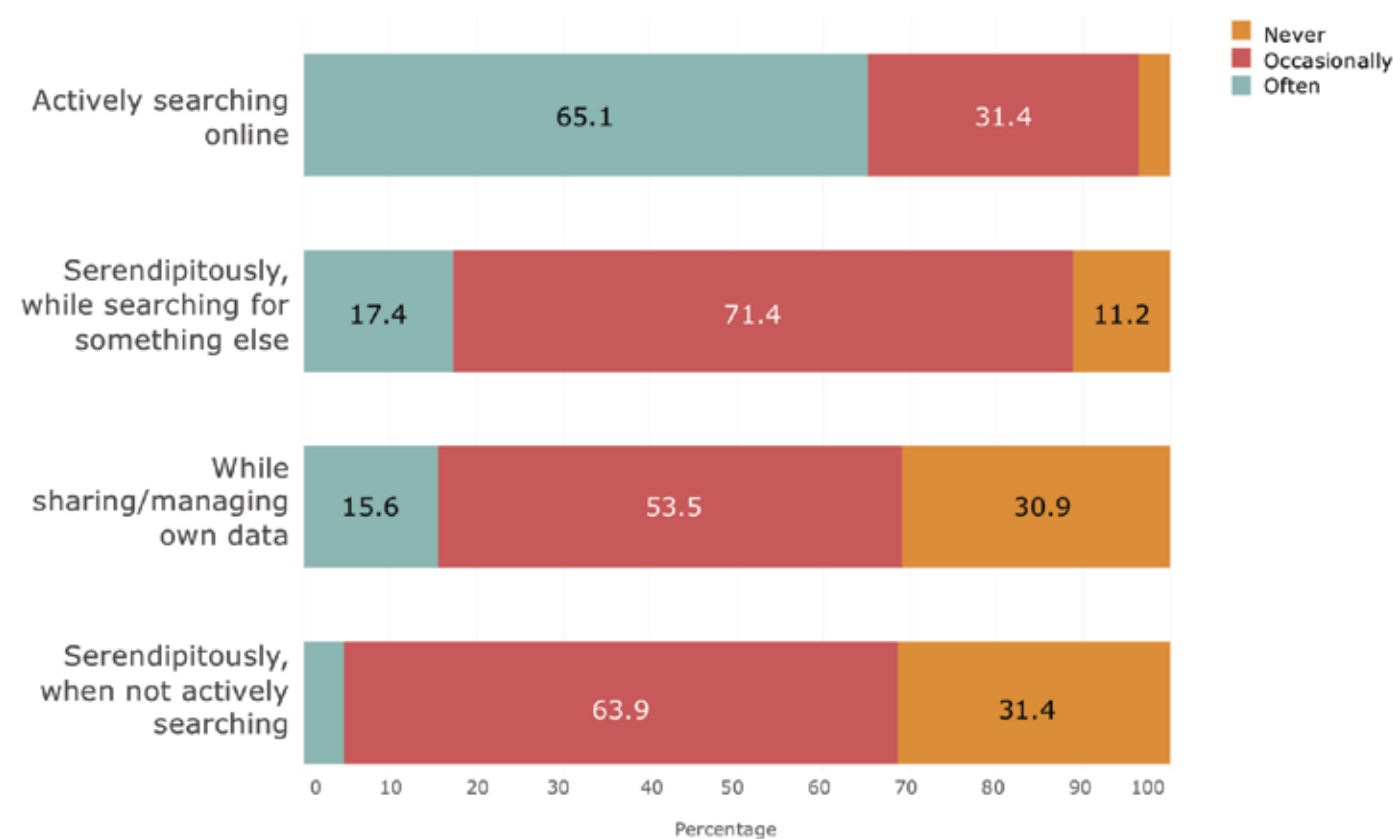
Results:

The discourse is spread across only a few disciplines and has a technical predominance.

"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)."

How frequently do you find data in the following ways?

	Never	Occasionally	Often
Actively searching online		65.1	31.4
Serendipitously, while searching for something else	17.4	71.4	11.2
While sharing/managing own data	15.6	53.5	30.9
Serendipitously, when not actively searching		63.9	31.4

Percentages represent the percent of all respondents (n=1677).

Conclusions:

Data search is a sociotechnical practice. As data seekers discover and make sense of data, their actions are situated within and formed by interactions between the social and technical spheres.

Data seekers do not only need what is commonly thought of as "research data." Findability and reusability are limited by existing infrastructures, unstandardized metadata creation practices and access to resources needed for sensemaking.

Even within disciplines, there is a great diversity in data search practices. Data needs and search practices are dynamic, influenced by changing contexts and developing technologies.

Data search practices are interwoven with other (re)search practices. They are shaped by existing resources and accustomed search strategies, particularly those for academic literature.

References:

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.; Groth, P; Cousijn, H.; 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

Funding:

Data Archiving and Networked Services (DANS)
Anna van Saksenlaan 51 | 2593 HW Den Haag
+31 70 349 44 50
info@dans.knaw.nl | dans.knaw.nl

Contact Info:
Kathleen Gregory, M.A., MSLIS
Researcher, PhD candidate
kathleen.gregory@dans.knaw.nl

DANS is an institute of KNAW and NWO

KNAW

NWO
Driven by data