



# Data Discovery Paradigms IG

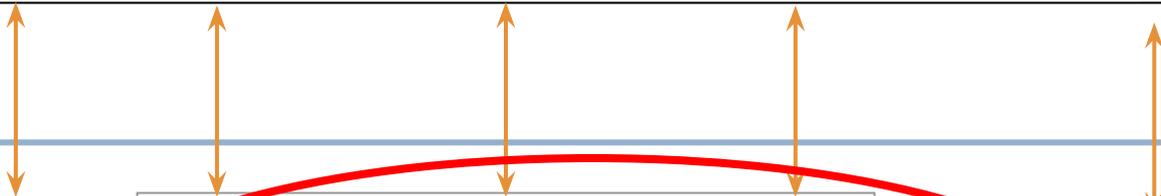
*Relevancy Ranking Task Force*

*RDA Global Adoption week  
18 June 2020*

**research data sharing without barriers**  
**[rd-alliance.org](http://rd-alliance.org)**



Search for Data Q Search



### System view

Crawling/  
Aggregating

Parsing

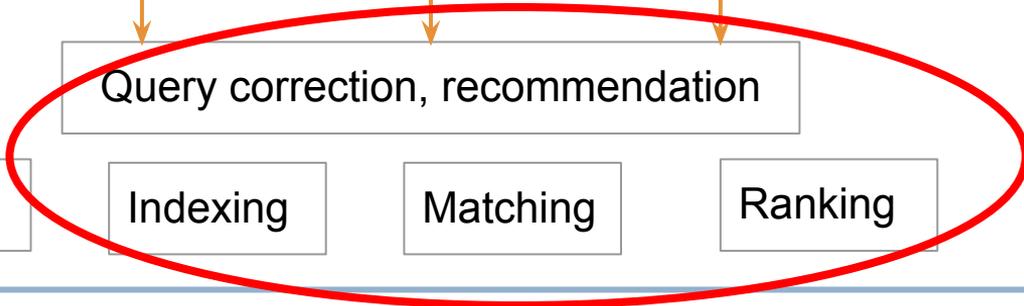
Query correction, recommendation

Indexing

Matching

Ranking

Snippet  
generation



data.gov



earthdata.nasa.gov



World Wide  
Science



Research  
Data  
Australia



DataCite



Dryad



Zenodo

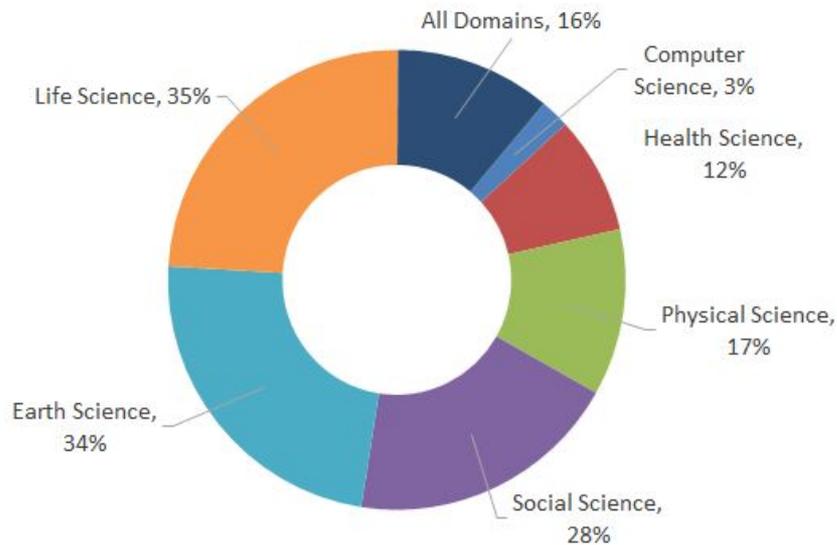
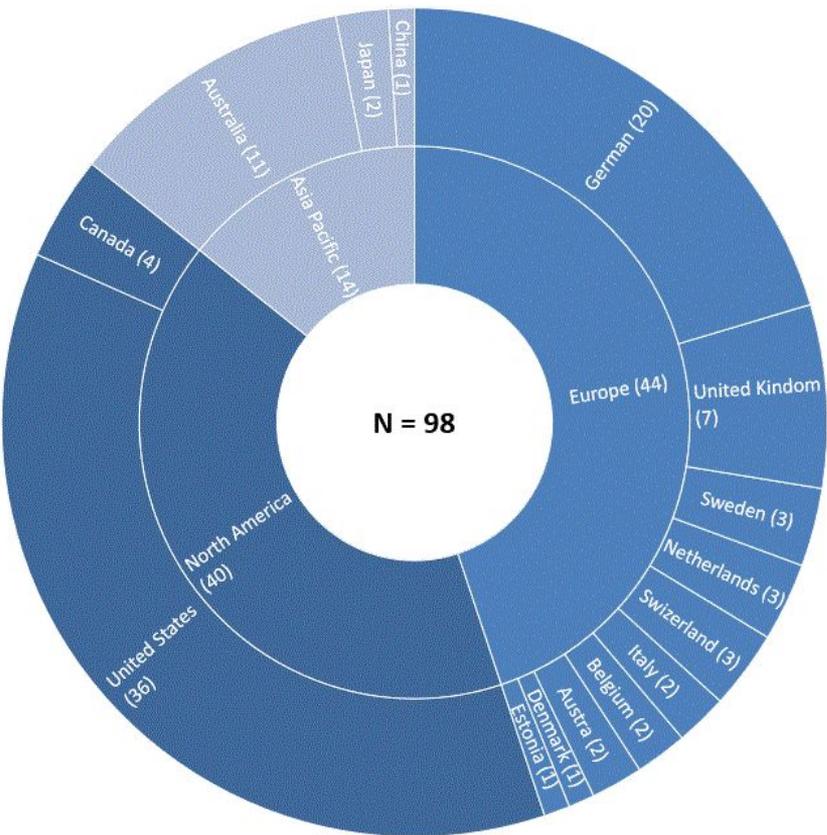
# Survey goals

- Investigate what data search systems and ranking models have been deployed.
- Serve as a benchmark to be looked back on in future to assess how much and in what ways data search has improved.
- Identify potential collaborative projects from the Survey

# Survey Design (33 Questions)

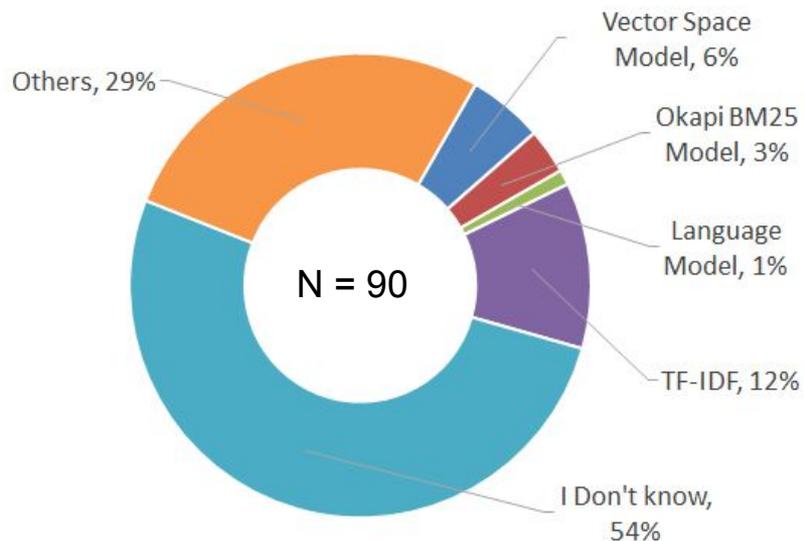
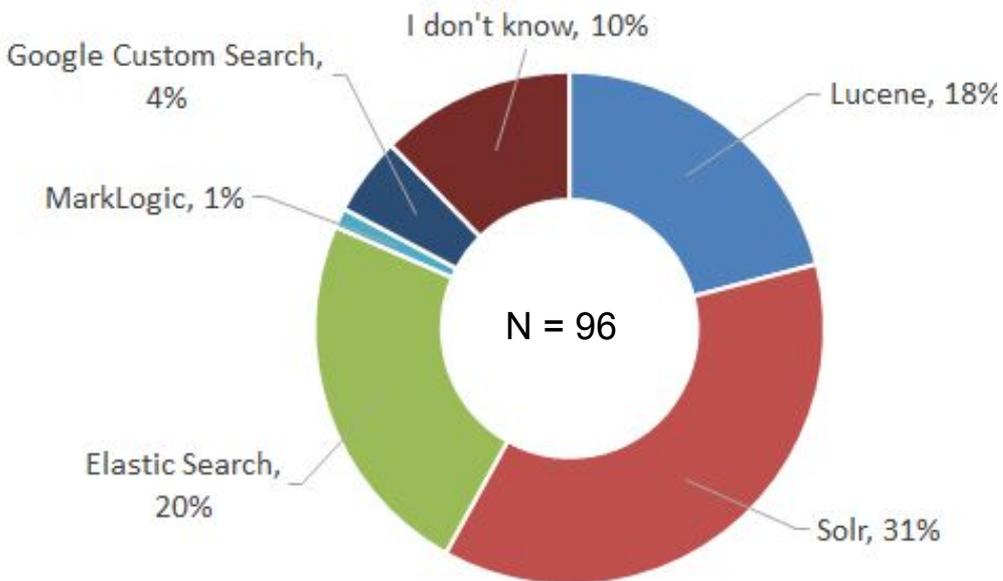
1. What are characteristics of each repositories? (5)
2. What are system configurations (e.g., ranking model, index methods, query methods)? (7)
3. What are evaluation methods and benchmark? (10)
4. What methods have been used to boost search-ability to web search engines? (2)
5. What other technologies or system configurations have been employed? (5)
6. Wish list for future activities for the RDA relevance task force (2)

# Participants background

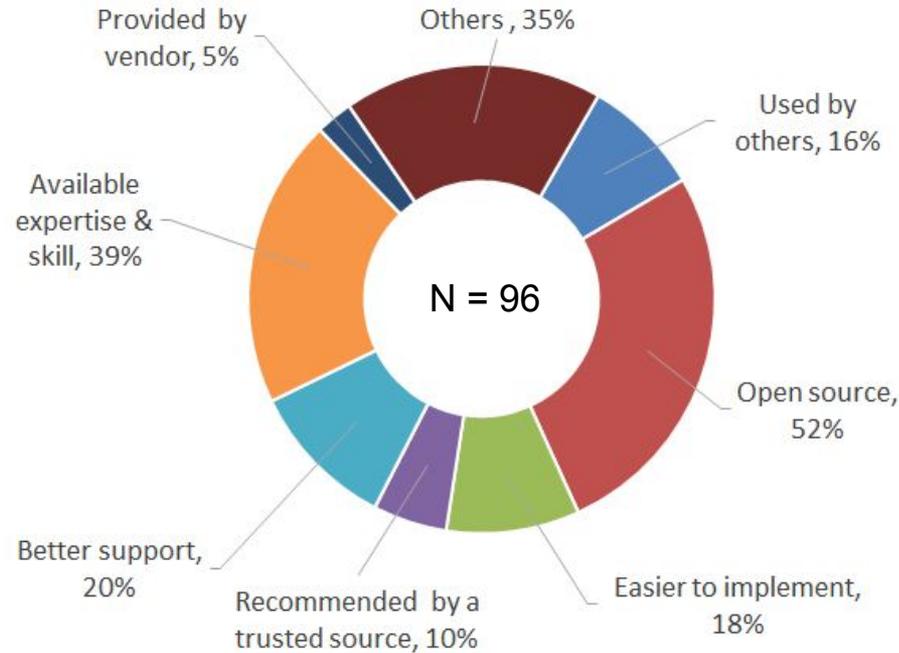


# Survey result highlights ...

# Data repositories use common search systems

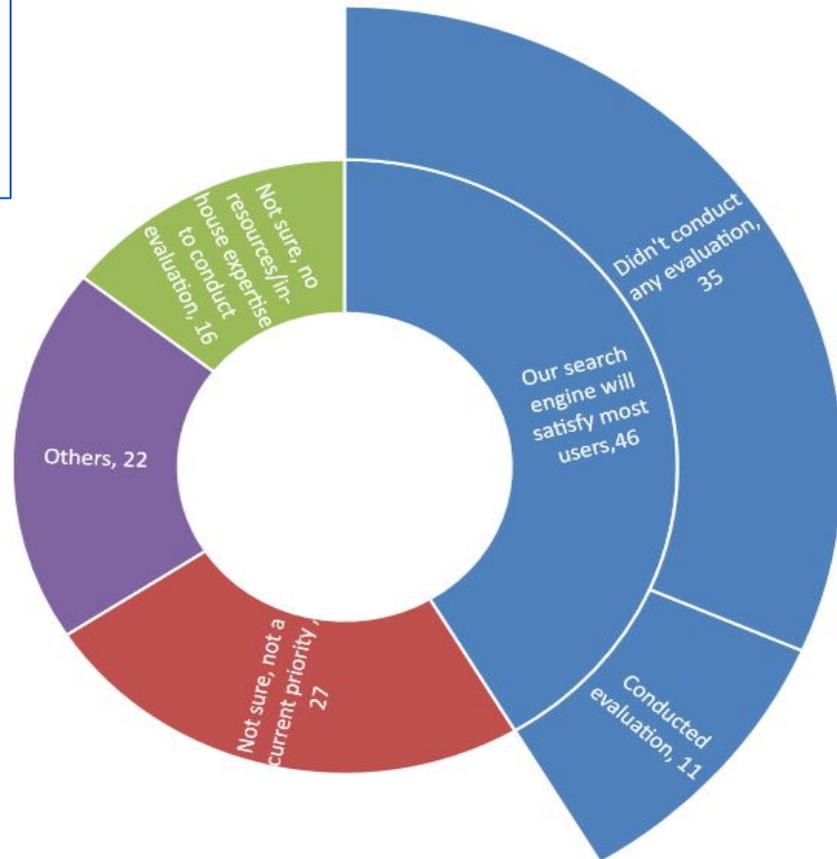
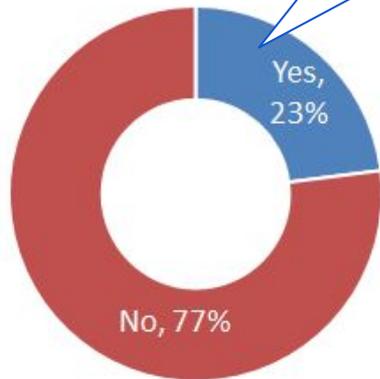


# Open source and available skills are top reasons for choosing a search system



# Majority didn't conduct any kind of evaluations

- 9 Created test collection
- 11 Informal evaluation
- 6 Log analysis
- No performance measure was provided



- Repositories desire guidelines for improving relevancy ranking in their data search system, with small repositories having the greatest need.
- Repositories understand that their search systems need to be evaluated and improved, but often lack the resources (time and/or expertise) to explore and evaluate the available options.
- The study concludes that there is an opportunity for people working in the search space to collaborate, to build test collections and other efforts that offer the greatest improvements in data search services.

Khalsa, SiriJodha; Cotroneo, Peter; Wu, Mingfang (2018), "A survey of current practices in data search services", Mendeley Data, v1 <http://dx.doi.org/10.17632/7j43z6n22z.1>

# Thank you

## Contact:

[fpsom@certh.gr](mailto:fpsom@certh.gr)

[mingfang.wu@ardc.edu.au](mailto:mingfang.wu@ardc.edu.au)

[sjsk@nsidc.org](mailto:sjsk@nsidc.org)

<https://www.rd-alliance.org/groups/data-discovery-paradigms-ig>