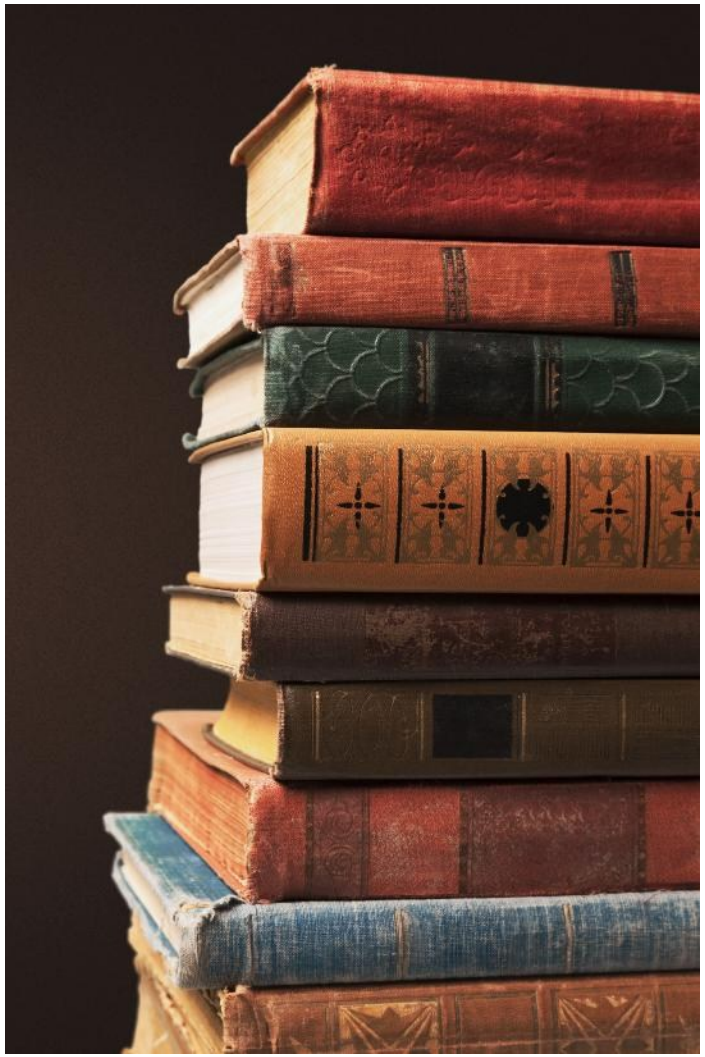
Navigating the Data Deluge: The Evolving Role of Librarians and Data Scientists in Strategic Decision-Making

Webb Myers, Nils Newman (Search Technologies), Erica Wiseman (NRC)

A Little History for Context…

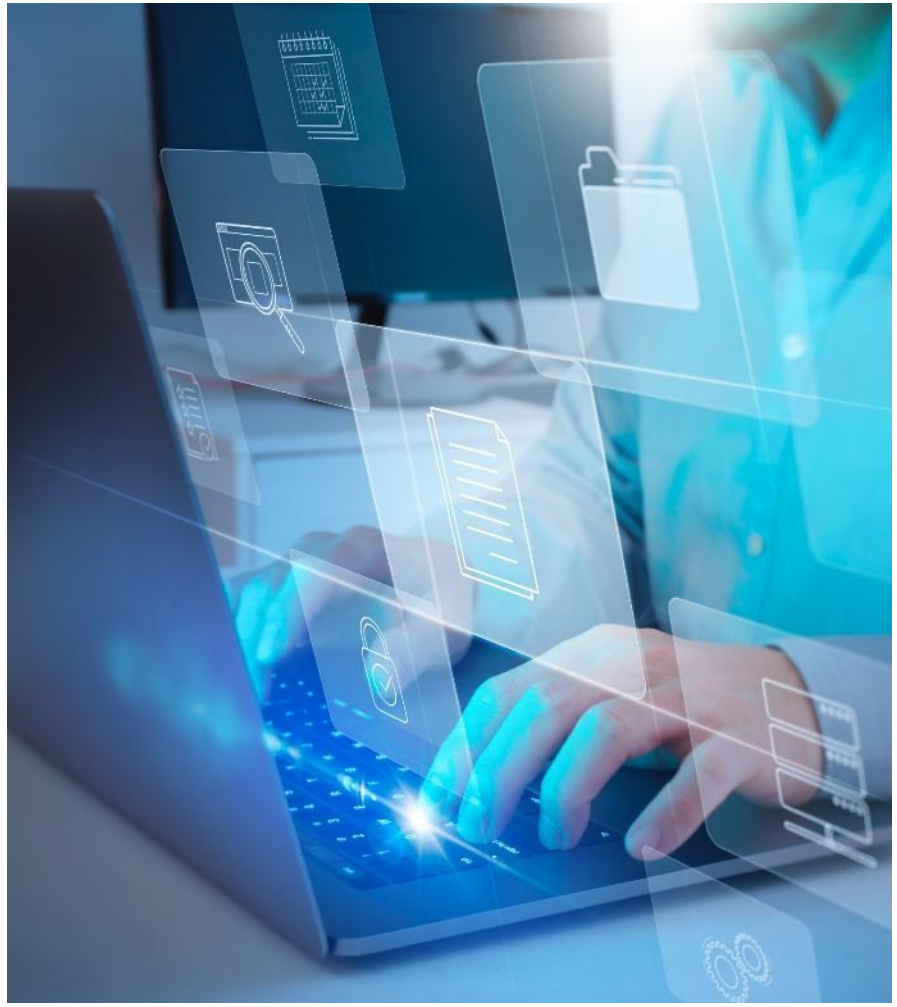
- 1980's – Books were actual books. Some specialist bibliographic databases were available (DWPI, Medline, SCI, INSPEC, EI, etc..) as were federated collections (Dialog [1966], SilverPlatter [1983], STN [1984], etc..). These resources required specialized skill to use (such as BRS) and specialized people (Reference Librarians).

- 1990 – Seismic change began in 1993 with the World Wide Web. Databases began to migrate from custom platforms to the Internet. That shift resulted in two major changes:
  - Searching perceived as a general-purpose skill
  - Data costs plummeted



More History….

- 2000 – The change speeds up….
  - Companies began digitizing everything while at the same time eliminating their corporate libraries and opening up their corporate information to employees on intranets.
  - Internet search engines grew to dominate search while traditional search approaches were killed off. (Google slowly deprecated Boolean in Google search finally killing it around 2019.)
  - Information which had traditionally been stored as text files shifted to XML.



More History….

- 2010 – More IT change…
  - Cost of storage plumets and storage capacity skyrockets.
  - Data Lakes emerge as a challenger to SQL (Big Data).
  - API's start becoming effective.
  - JSON starts replacing XML.
  - Developments in Machine Learning and AI advance "*natural language searching*" supplanting traditional query. (BERT and others…)

"Let's shrink Big Data into Small Data ... and hope it magically becomes Great Data."

Today

- The result of all this change has been a fundamental shift in the economics of information.
- Data historically was scarce resource requiring skills to use and was sold as such – per record pricing with few people using the resource but many records used.
- The data are now a commodity – the goal of providers is to sell access to as many people as possible.
  - *Having access to data is not special.*
  - *What you do with it is.*
- However, figuring out the best, most effective way to add value is a challenge.

Who Are We and What Do We Do with the Data Today

National Research Council Canada | Conseil national de recherches Canada

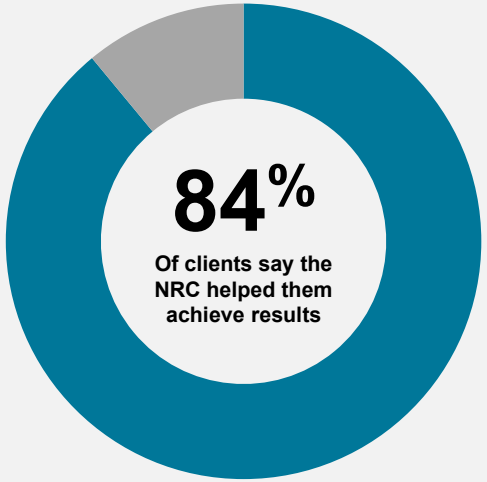
Research performer (national labs)

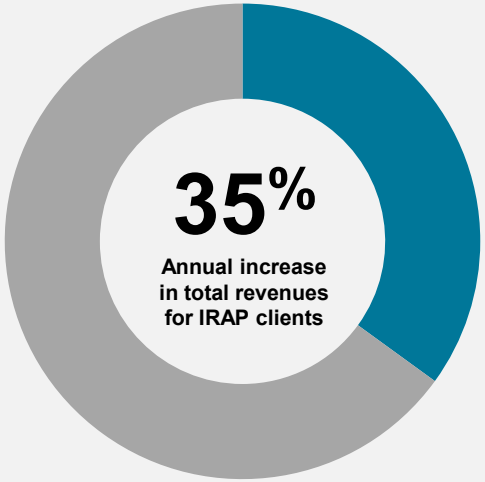
- **14 research centres**
- **24 laboratory sites**
- **113 major R&D facilities**
- **$1,526M total expenditures**
- **1,277 peer-reviewed publications**
- **175.9M total revenues**
- **956 R&D clients**
- **569 collaborative R&D project**
- **1,971 active patents (462 patent families)**

Advice and funding to SME innovation projects

Industrial Research Assistance Program (NRC IRAP)

273	113	$468M	3,262	6,148
ITAs	locations	Gs &Cs funding invested in SMEs	total firms funded	Unfunded firms receiving advisory services





Data for 2023-3034

Library and Information Management Services

Accessing Collections

Information Management

Research and Business Intelligence

Publishing, Preserving Research and Data

Access to Information and Privacy (ATIP)

Intelligence and Analytics (I&A)

→Market dynamics

→Key players in research and industry

→Collaboration opportunities & due diligence

→Scientometric studies

→Citation analysis

→Training materials for using the NRC Library

→Advice to formulate complex search strategies, identify useful resources

→Prompt engineering

→Assess the impact of a research group or subject area using bibliometrics

Typical Uses of I&A Services

Research

→ research and industry landscape

→ Identify new directions

→ benchmarking

→ research impact

and beyond

→ Guide and evaluate program

→ buy-in

→ collaboration agreements

→ consortia

→ funding

→ value chains

→ policy directions

Working with NRC Intelligence & Analytics

Scientometric Studies

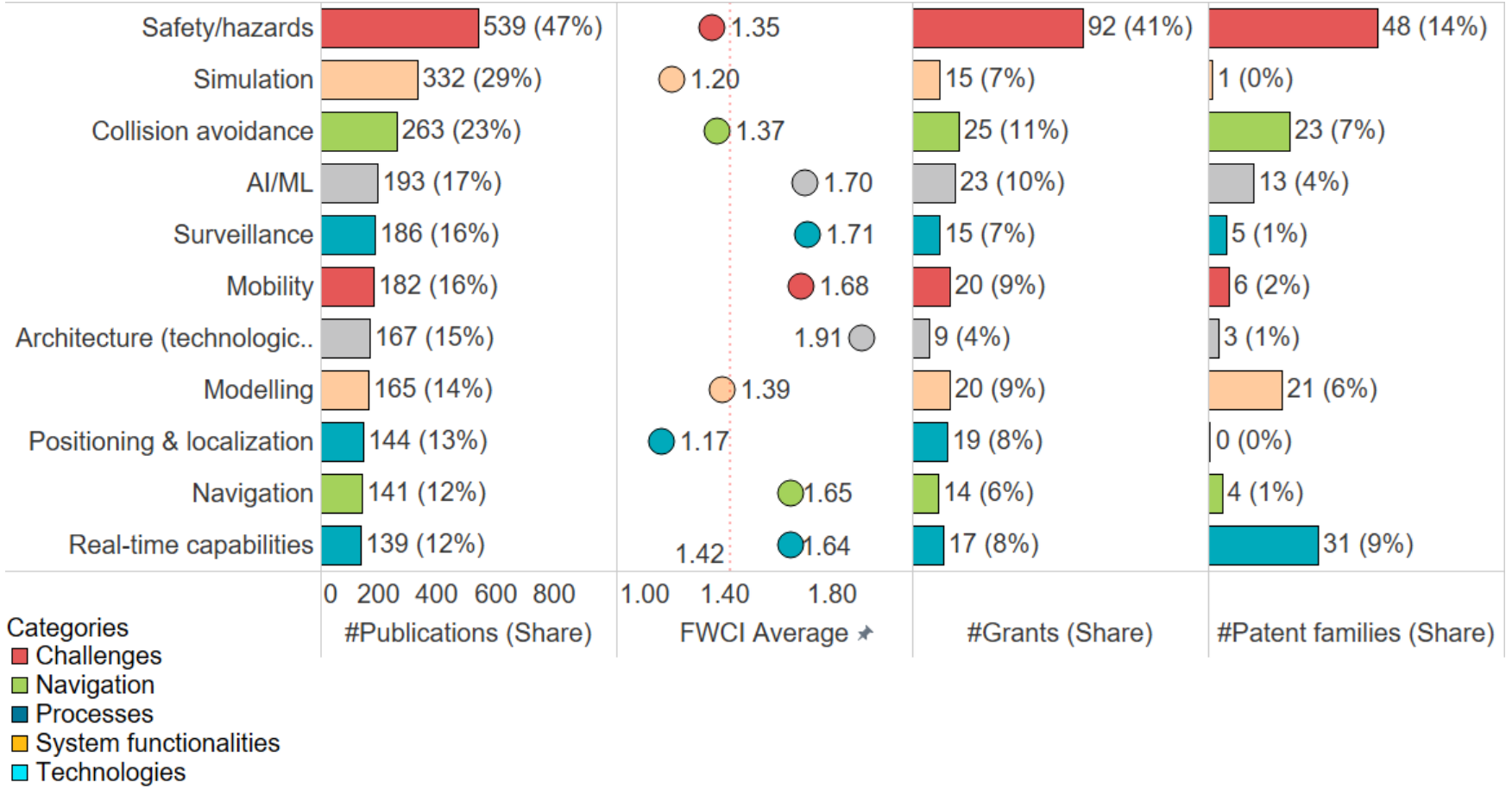
Impact

Typical Scientometric Questions

Strategic Planning

→ Barriers and drivers
→ Applications

Identify Research Topics



Tells you…

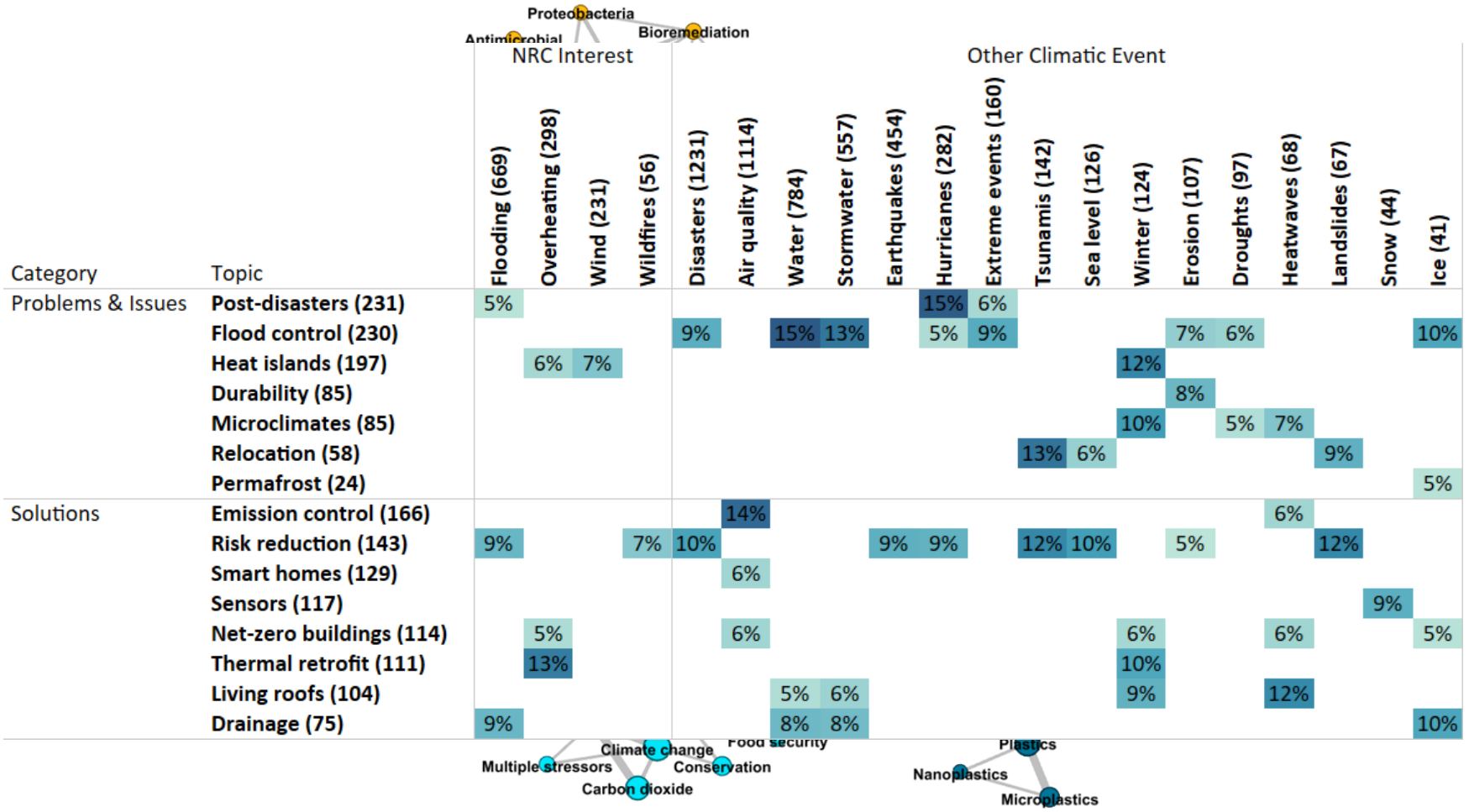
→ Volume of records per topics across 3 datasets

Impact…

→ Identify topics with high research impact

→ Insight into Technology Readiness Level (TRL)

Identify Relationships Between Topics



Tells you…

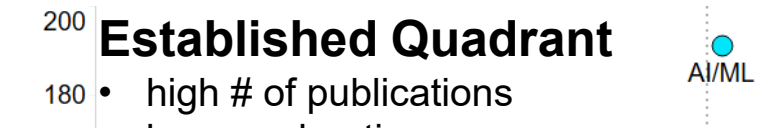
→ Relationships among topics in a domain

Impact…

→ Identify closely related topics to investigate

→ Highlight cross-disciplinary research

Identify Emerging Trends in R&D



Established Quadrant

- high # of publications
- low # of publications
- low acceleration

Notable topics from a research perspective

Topic category	Topic subcategory	Topic groups	Awards	Publications	Patents
Applications	Heavy vehicles	Trucks	↗	↗	↓
	Light/Midsize vehicles	Automobiles	↑	↑	↓
Challenges	Fabrication	Degradation	→	→	↓
		Durability	↗	↗	
	Governance/Human factors	Adoption	↑	→	↓
		Governments	↑	↗	
	Infrastructure/Operation	Remote areas	↗	↗	
		Smart cities	↗	↗	
		Urban areas	→	↑	↓
	Performance	Reliability	→	→	
	Safety/Security	Cybersecurity	↗	↗	↓
Features	Analytics	Modelling	→	→	↓
		Simulation	↑	→	↓
	Operation	Fault tolerance	↗	↗	
		Robustness	→	→	↓
	Safety/Security	Fire resistance	↗	↗	↓
Materials	Battery materials	Cobalt	↗	↗	↓
		Manganese	↗	↗	↓
		Nickel	↗	↗	↓
	Battery/Vehicle component..	Composites	→	↗	↓
Technologies	Charging	Fast charging	↑	→	↓
	Energy distribution	Smart grids	→	→	↓
	Vehicle components	Heating, ventilation and air co..	↗	↗	↓

Quadrant
→ Established
↑ Hot
↗ Emerging
↓ Disappearing/New

Hot Quadrant

- high # of publications
- high acceleration

Notable topics from a commercial perspective

Topic category	Topic subcategory	Topic groups	Awards	Publications	Patents
Applications	Heavy vehicles	Heavy/Service vehicles	↓	↓	↑
Challenges	Fabrication	Vibration	↓	↓	↗
	Governance/Human factors	Standards	↓	↓	↗
	Safety/Security	Battery overheating	↓	↓	↗
		Security	↓	↗	↑
Features	Design	Miniaturization	↓	↓	↗
	Operation	Automation	↓	↓	↑
		Thermal management systems	↓	↓	↗
Materials	Battery materials	Zinc		↓	↗
	Battery/Vehicle component..	Aluminum	↗	↓	↑
Technologies	Charging	Wireless power transfer	↓	↓	↗
	Energy distribution	Distributed energy	↓	↓	↗
	Energy equipment	Transformers	↓	↓	↑
	Vehicle components	Regenerative braking	↓	↓	↑

Quadrant
Established
Hot
Emerging
Disappearing/New

Tells you…

→ What topics are established, hot, emerging or brand new/declining

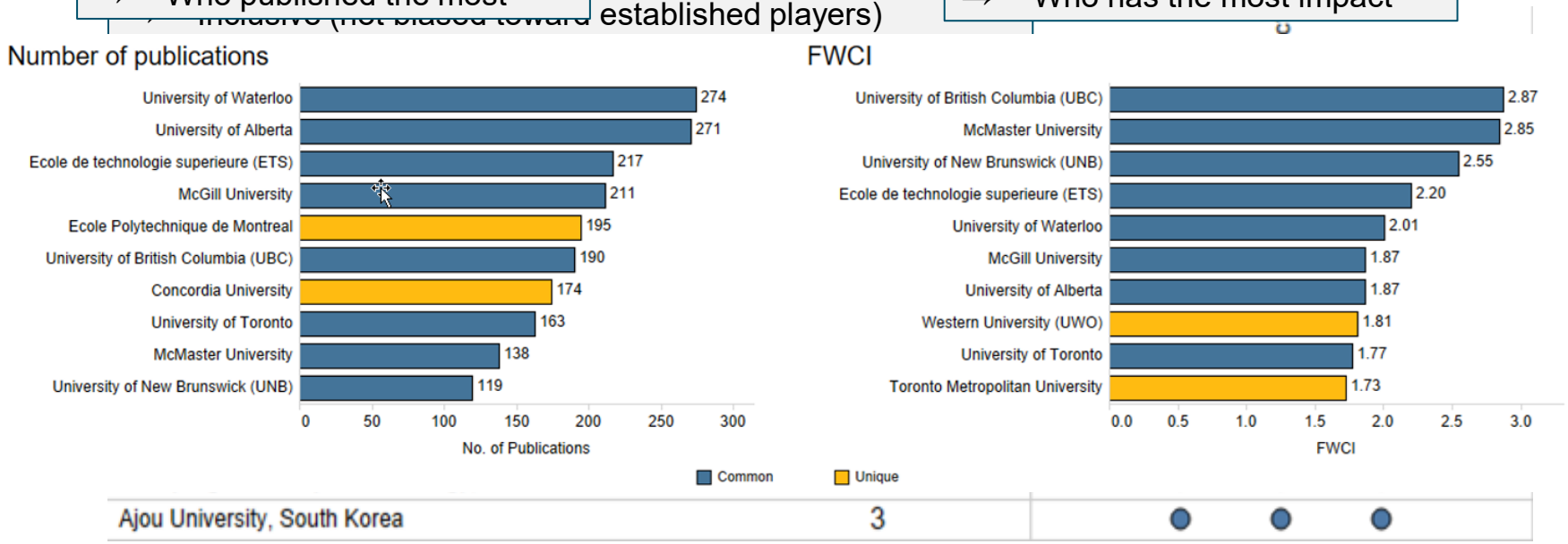
Impact…

→ Validate assumptions
→ Identify new research areas
→ Understand the research momentum of topics in your field
→ Focus efforts on emerging technologies

Identify Key Players (People or Organizations)

Volume (Quantity)
→ Who published the most

FWCI (Impact)
→ Who has the most impact



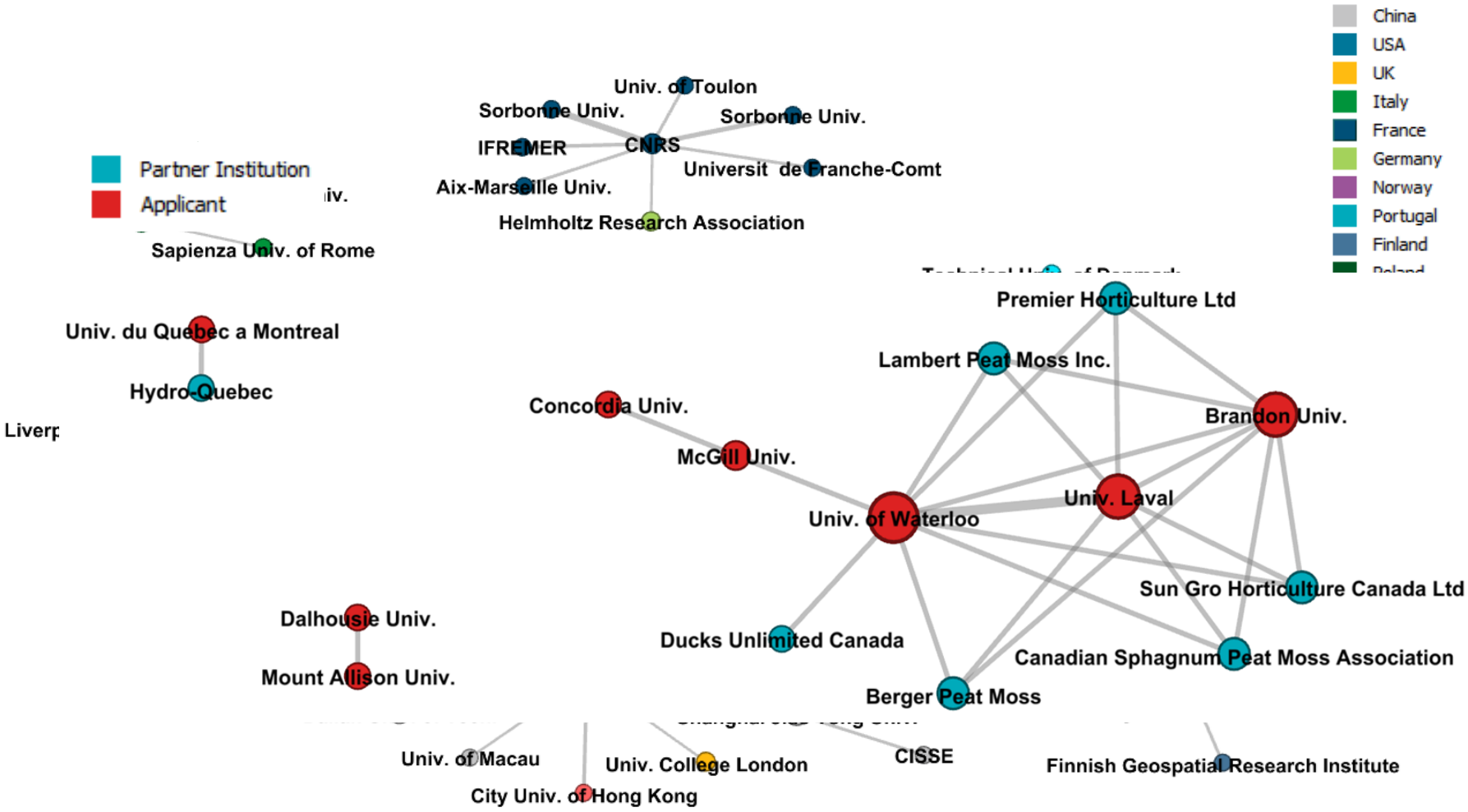
Tells you…

→ What key players operate in an area of interest based on customizable criteria

Impact…

→ Identify potential collaborators, clients competitors & licensees
→ Identify where R&D is happening
→ Rank key players (ex. by scientific impact)

Identify Collaboration Networks



Tells you…

→ Who is working with who
→ Role of collaborators (gatekeepers, most connected, etc.)

Impact…

→ Understand the collaborative research landscape
→ Identify R&D networks and clusters
→ Identify key potential collaborators/competitors based on influence or scientific impact

Identify Areas of Expertise

			Topics						
			Collision avoidance	Detect and avoid	Global navigation satellite systems ..	Localization	Mapping	Navigation	Trajectory planning
NASA, USA	Publications	93	22% (20)	3% (3)	5% (5)	1% (1)	1% (1)	18% (17)	10% (9)
German Aerospace Center (DLR), Germany	Grants	11	18% (2)						
	Publications	37	19% (7)	3% (1)	5% (2)		3% (1)	22% (8)	5% (2)
RMIT University, Australia	Publications	26	19% (5)	4% (1)	15% (4)	4% (1)		50% (13)	15% (4)
Cranfield University, UK	Grants	10	10% (1)				20% (2)	10% (1)	
	Publications	24	8% (2)		4% (1)		4% (1)	8% (2)	13% (3)
Nanyang Technological University, Singapore	Publications	24	29% (7)		8% (2)			25% (6)	8% (2)
Beihang University, China	Publications	22	36% (8)					9% (2)	5% (1)
Delft University of Technology, Netherlands	Publications	20	10% (2)				5% (1)	5% (1)	10% (2)

Tells you…

→ Who is actively working on topics of interest

Impact…

→ Identify who has unique areas of expertise

→ Identify experts in emerging topics

→ Identify who is working in topics of interest for potential collaboration

Supporting Security Checks

B	C	D
NRO found	**Number of co-publications with NROs**	**Number of publications funded by NROs**
Harbin Institute of Technology, China	3	
Tianjin University, China	3	
Beijing Computational Science Research Centre, China	1	
Harbin Institute of Technology, China	1	
Hefei National Laboratory for Physical Sciences at Microscale, China	3	
Northwestern Polytechnical University, China	1	
Nanjing University of Science and Technology, China	2	
Tianjin University, China		1

Tells you…

→ Whether a potential partner is associated with an affiliation of concern

Impact…

→ Ensure compliance with the Policy on Sensitive Technology Research and Affiliations of Concern

→ Protect NRC innovations and projects

What Clients Say

"I feel it's important to stress the benefit of this…service and how it supports NRC's unique mandate to work with industry to develop the technologies, innovations and other solutions that will be needed to create a new and prosperous climate economy"

The Future

- AI-based “natural language query” will be ubiquitous (bye-bye Boolean).
- LLM’s will touch all of our lives.
- Downstream prompt engineering will become a general purpose skill. (As will data science.)
- Traditional databases will continue to be under pressure.

- AI “Answer Machines” will become increasingly available.