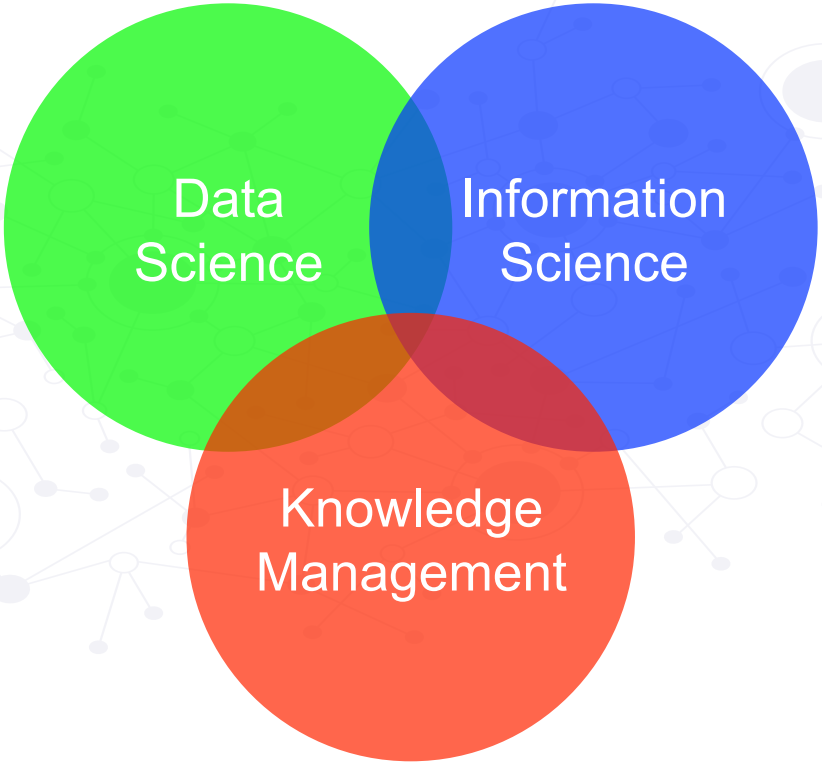
Synaptica

International Society for Knowledge Organization
Singapore Chapter

Straits Knowledge
Redefining corporate knowledge

Dave Clarke
Founder, Synaptica

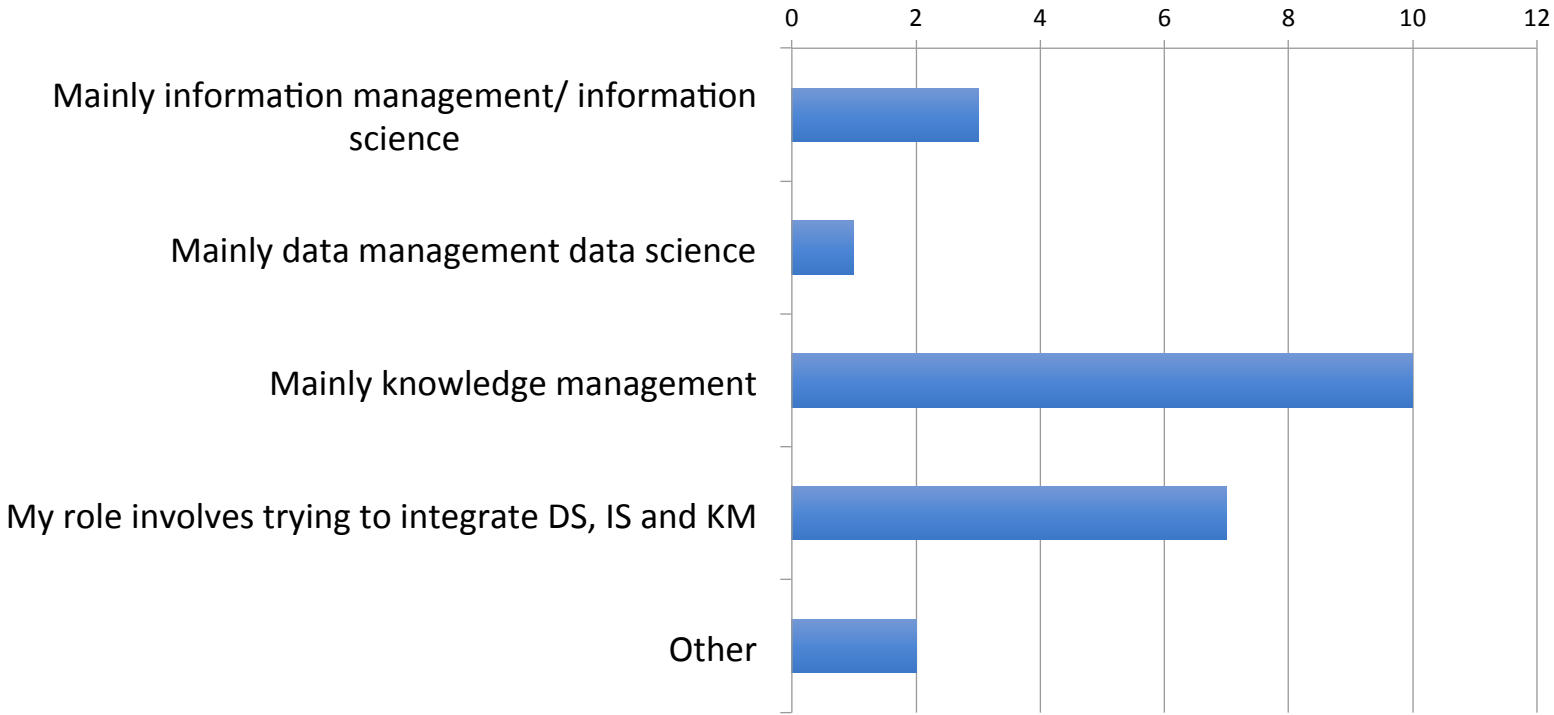


Patrick Lambe
Principal, Straits Knowledge

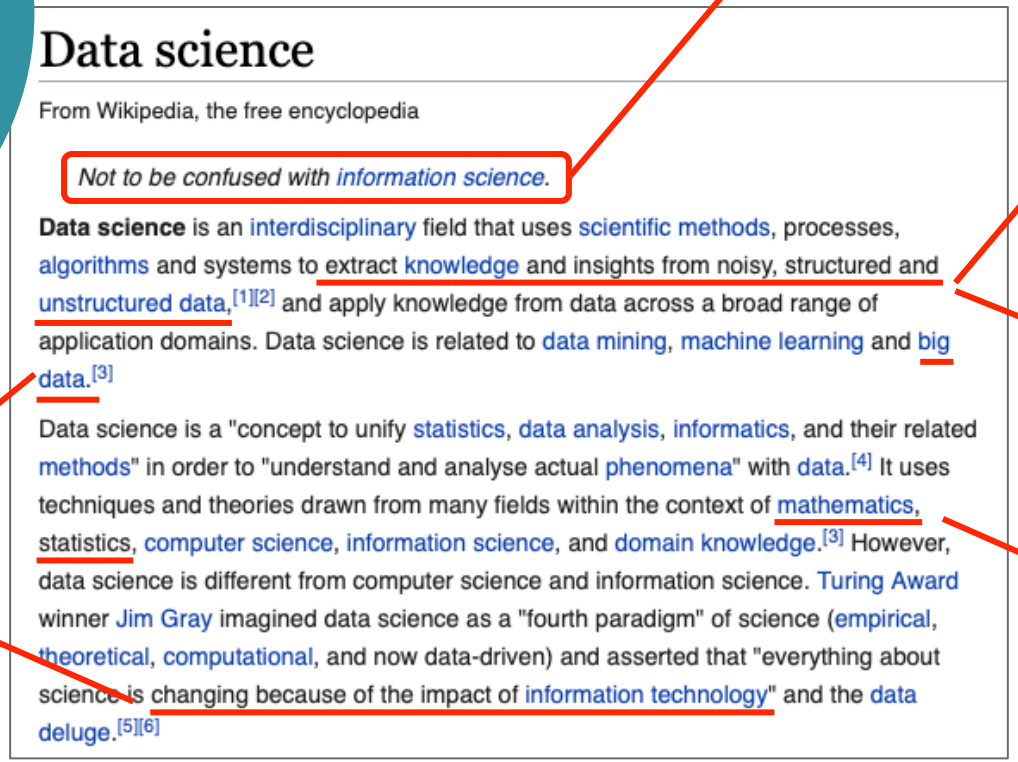
Agenda

- Poll: What's your main focus area?
- Data Science vs Information Science (Dave Clarke)
- How does Knowledge Management relate? (Patrick Lambe)
- Commonalities and differences (Dave and Patrick)
- Poll on alignment/coordination between DS, IS, KM
- When we fail to "mind the gap": issues around exploiting common ground and respecting differences (discussion)

Poll 1: What's your background?



Data Science



Data science

From Wikipedia, the free encyclopedia

Not to be confused with information science.

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from noisy, structured and unstructured data,[1][2] and apply knowledge from data across a broad range of application domains. Data science is related to data mining, machine learning and big data.[3]

Data science is a "concept to unify statistics, data analysis, informatics, and their related methods" in order to "understand and analyse actual phenomena" with data.[4] It uses techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, information science, and domain knowledge.[3] However, data science is different from computer science and information science. Turing Award winner Jim Gray imagined data science as a "fourth paradigm" of science (empirical, theoretical, computational, and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the data deluge.[5][6]

1 Not Information Science

2 Noisy data

3 Information extraction

4 Maths and statistics

5 Use of big knowledge graphs like Wikidata

6 IT capabilities driving a changing landscape

https://en.wikipedia.org/wiki/Data_science

Library and Information Science 1

Information Science

Classification and Retrieval 2

Taxonomy and Knowledge Organization Systems 3

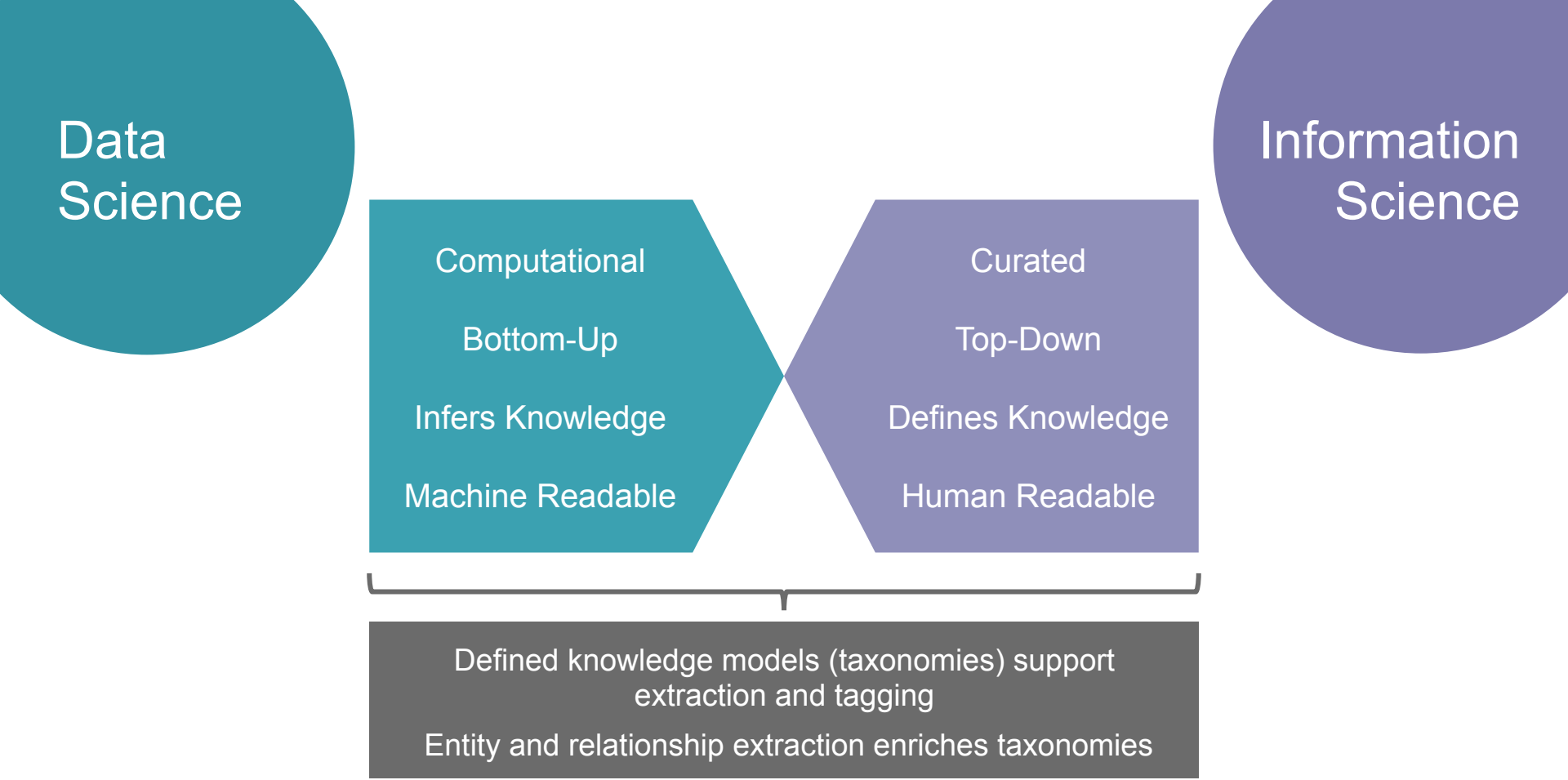
4 Organizational and human knowledge … overlap with KM

Information science

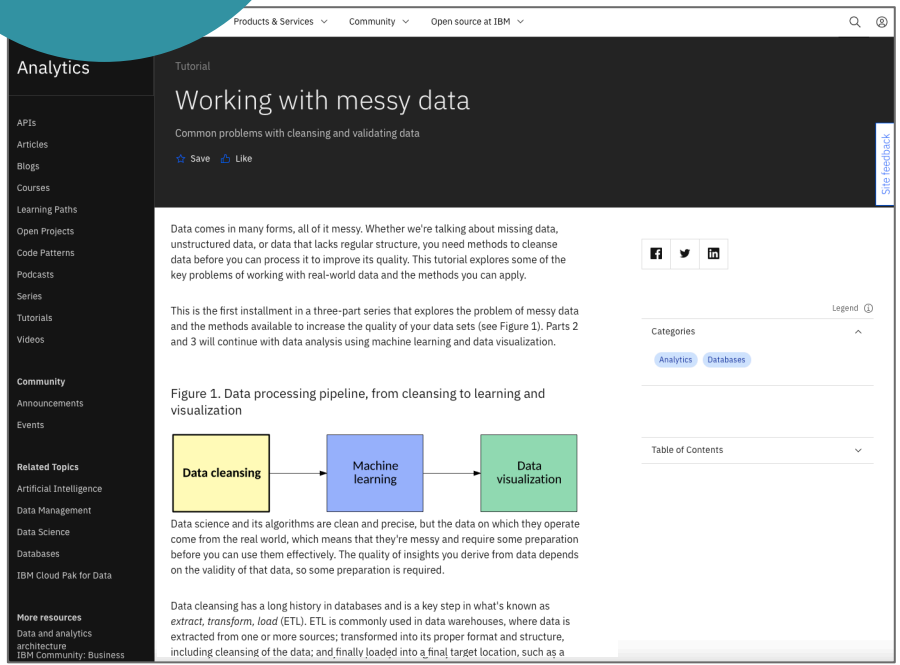
From Wikipedia, the free encyclopedia

Information science (also known as **information studies**) is an academic field which is primarily concerned with analysis, collection, classification, manipulation, storage, retrieval, movement, dissemination, and protection of information.[1] Practitioners within and outside the field study the application and the usage of knowledge in organizations in addition to the interaction between people, organizations, and any existing information systems with the aim of creating, replacing, improving, or understanding information systems.

Historically, information science is associated with computer science, data science, psychology, technology, library science, healthcare, and intelligence agencies.[2] However, information science also incorporates aspects of diverse fields such as archival science, cognitive science, commerce, law, linguistics, museology, management, mathematics, philosophy, public policy, and social sciences.

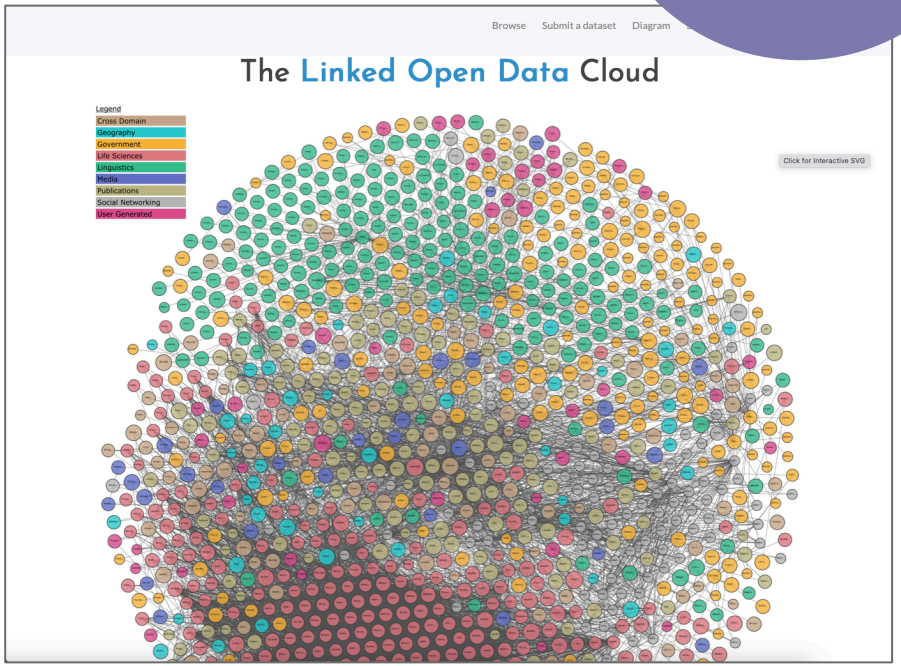

Data Science
Information Science
Computational
Bottom-Up
Infers Knowledge
Machine Readable
Curated
Top-Down
Defines Knowledge
Human Readable
Defined knowledge models (taxonomies) support extraction and tagging
Entity and relationship extraction enriches taxonomies

Messy Internal Content



https://developer.ibm.com/tutorials/ba-cleanse-process-visualize-data-set-1/

Messy External Data



https://lod-cloud.net

Information Extraction

now news on the web by ontotext

Best ever image of the universe revealed by Nasa's James Webb telescope

news.google.com • 12 hours ago

Nasa and the James Webb Space Telescope just peered billions of years further into the past than ever before, and brought photos home for all to see.

Around 6.20pm EDT, US President Joe Biden revealed the very first public image taken by the Webb telescope, a sea of distant galaxies and brilliant spiky stars, the most distant view yet of the Cosmos . The image was presented on a screen at the White House along with comments by Nasa Administrator Bill Nelson .

"Mr President, if you held a grain of sand on the tip of your finger at arms length, that is the part of the universe you are seeing," Mr Nelson said, a tiny portion of the sky magnified by Webb to reveal thousands of galaxies. "That light you are seeing on one of those little specks, has been traveling for over 13 billion years."

Nasa's Twitter account published a higher resolution version of the image than was projected at the White House.

Stars in the foreground appear as brilliant blue white spikes, while the most distant galaxies appear as orange-red smears, a

This article is about

James Webb Space Telescope

space observatory launched in 2021

Other mentions:

NASA Bill Nelson

Präsident der Vereinigten S

Twitter White Hous

Finding specific things mentioned in text is easier than determining overall document aboutness

Precision vs recall bias

Many links

Annotation types

Topic 48

Location 16

Organization 11

Person 10

Software 6

NLP most suited to extraction of novel **named entities** (people, places, organizations, etc.)

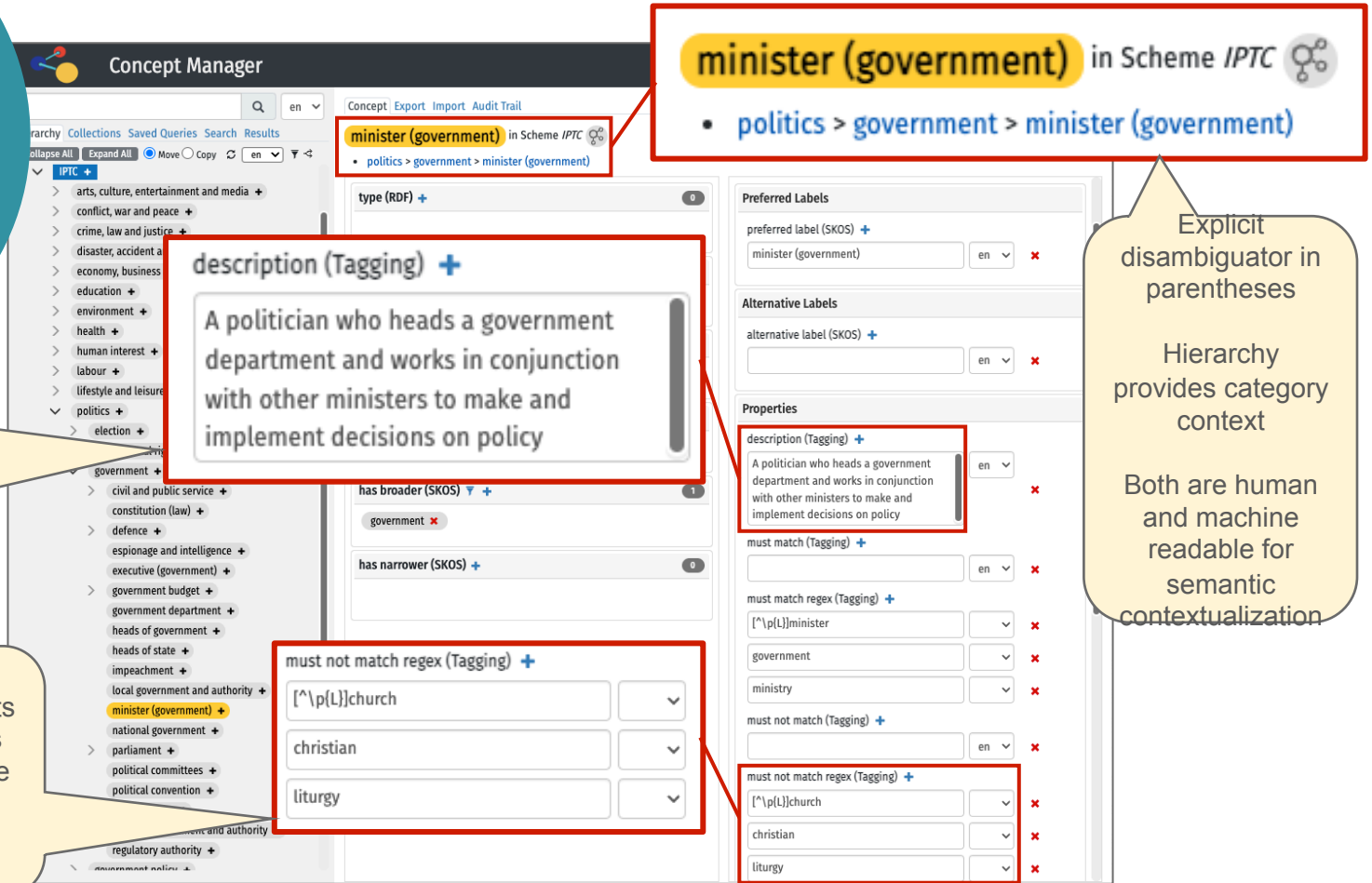
Extraction of abstract **topical concepts** is harder to achieve

NLP coupled with big knowledge graphs can help with the identification and disambiguation of entities

This NOW news demo uses NLP and Wikidata

https://now.ontotext.com/

Bridging the Human v. Machine-Readable Gap



Natural language descriptions / definitions are readable by humans and by NLP ML semantic similarity algorithms

To disambiguate ambiguous concepts simple NOT words and phrases can be added to help the machine eliminate false matches

Explicit disambiguator in parentheses

Hierarchy provides category context

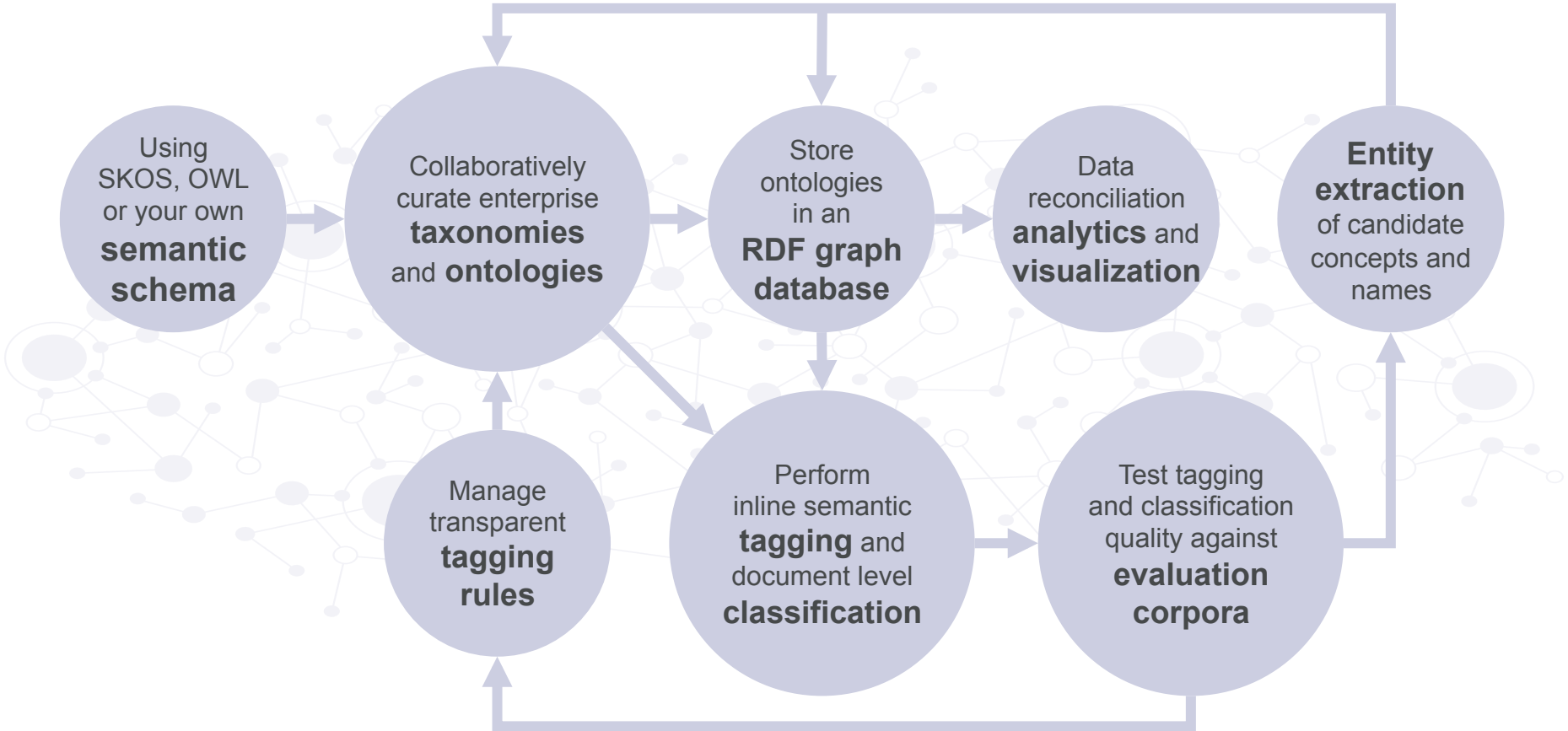
Both are human and machine readable for semantic contextualization

Dave Clarke, Synaptica
Patrick Lambe, Straits Knowledge

International Society for Knowledge Organization
Singapore Chapter

ISKO Singapore
July 15, 2022

Synaptica example of an Info-Science / Data Science feedback loop



Knowledge Management

Not = information management 1

Processes for creating, using, sharing 2

Knowledge management

From Wikipedia, the free encyclopedia

Not to be confused with Content management or Information management.

Knowledge management (**KM**) is the collection of methods relating to creating, sharing, using and managing the knowledge and information of an organization.[1] It refers to a multidisciplinary approach to achieve organisational objectives by making the best use of knowledge.[2]

An established discipline since 1991,[3] KM includes courses taught in the fields of business administration, information systems, management, library, and information science.[3][4] Other fields may contribute to KM research, including information and media, computer science, public health and public policy.[5] Several universities offer dedicated master's degrees in knowledge management.

Many large companies, public institutions and non-profit organisations have resources dedicated to internal KM efforts, often as a part of their business strategy, IT, or human resource management departments.[6] Several consulting companies provide advice regarding KM to these organizations.[6]

Knowledge management efforts typically focus on organisational objectives such as improved performance, competitive advantage, innovation, the sharing of lessons learned, integration and continuous improvement of the organisation.[7] These efforts overlap with organisational learning and may be distinguished from that by a greater focus on the management of knowledge as a strategic asset and on encouraging the sharing of knowledge.[2][8] KM is an enabler of organizational learning.[9][10]

https://en.wikipedia.org/wiki/Knowledge_management

3 Aligned with organizational objectives

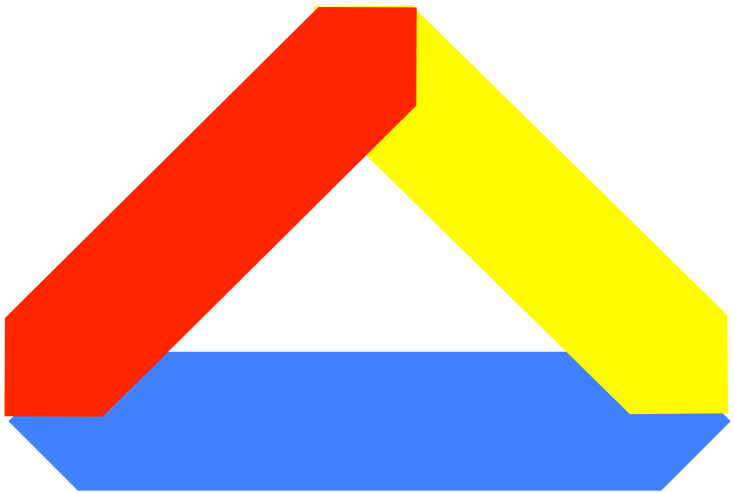
4 Interdisciplinary – should orchestrate IS, IM and DS

5 Enables org learning

Dave Clarke, Synaptica
Patrick Lambe, Straits Knowledge

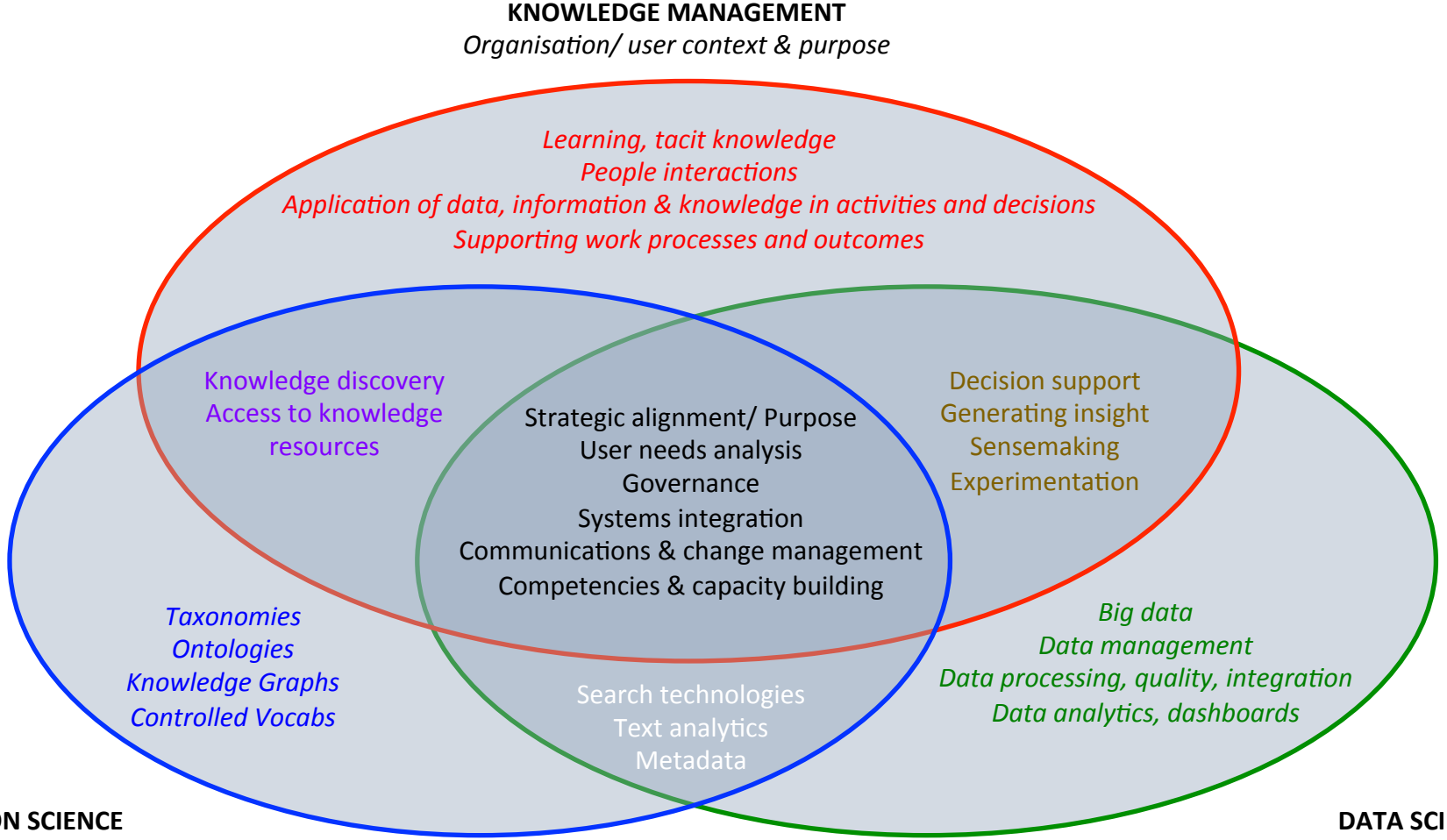
The Big Picture

KNOWLEDGE MANAGEMENT
Organisation/ user context & purpose



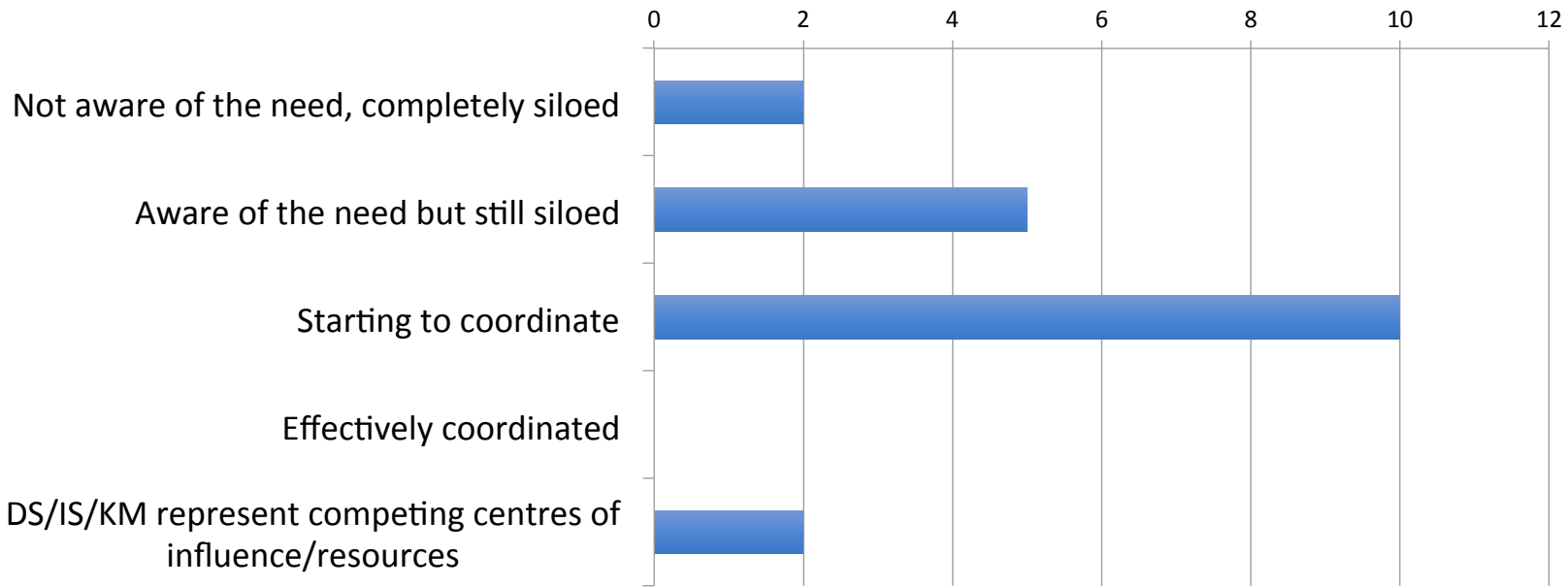
INFORMATION SCIENCE
Information structure, standards, rules

DATA SCIENCE
Machine learning, data processing & analytics



Dave Clarke, Synaptica
Patrick Lambe, Straits Knowledge

Poll 2: Where is your organization at in relation to managing DS/IS/KM orchestration?



Discussion: “Mind the Gap” – Orchestrating but Respecting Differences

- Some horror stories
- Where have you seen orchestration work well?
- Let’s talk about weaknesses:
  - KM – Tends to open up too many fronts for change – progress is slow
  - IS – Tends to be overly conceptual – high resource to implement and govern well
  - DS – Tends to rush to technology, not always easy to validate accuracy of insights

Additional Resources

Critique of the DIKW pyramid for being too simplistic

- Williams, D. (2014). Models, metaphors and symbols for information and knowledge systems. Journal of Entrepreneurship, Management and Innovation, 10(1), 79–107.
- https://jemi.edu.pl/vol-10-issue-1-2014/models-metaphors-and-symbols-for-information-and-knowledge-systems

Explainability of AI/ML/ data science outputs

- Klein, G. (2020). AIQ: Artificial Intelligence Quotient. Helping people get smarter about smart machines. Psychology Today, July 1st.
- https://www.psychologytoday.com/ie/blog/seeing-what-others-dont/202007/aiq-artificial-intelligence-quotient

Contact

Dave Clarke

- www.synaptica.com
- dave.clarke@synaptica.com

Patrick Lambe

- www.straitsknowledge.com
- plambe@straitsknowledge.com