



Workshop on Knowledge Science

EAISI
EINDHOVEN
AI SYSTEMS
INSTITUTE

TU/e

George Fletcher¹, Paul Groth², Juan Sequeda³

¹Eindhoven University of Technology

²University of Amsterdam

³data.world

Finding digital truth—that is, **identifying and combining data** that accurately represent reality—is becoming more difficult and more important.

More difficult because data and their sources are multiplying.

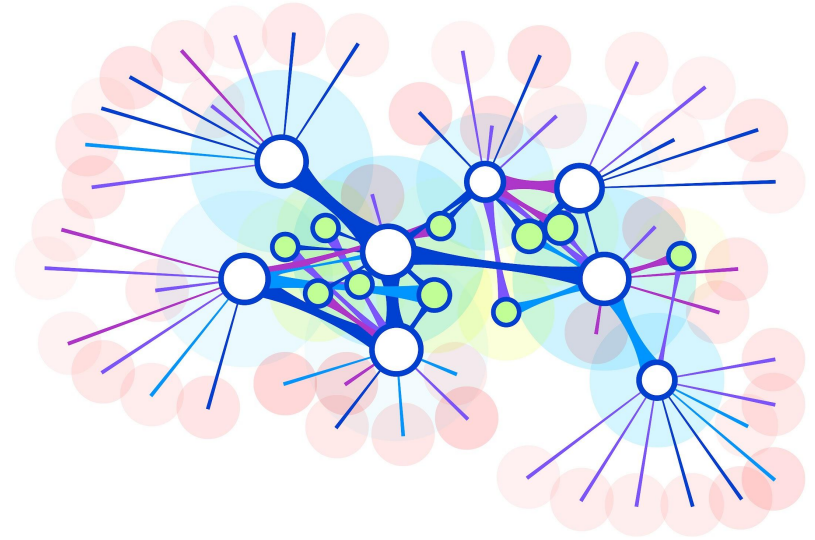
And more important because *firms need to get their data house in order to benefit from AI*, which they must to stay competitive.

-- *The Economist*, February 2020



Data interoperability and quality, as well as their structure, authenticity and integrity are key for the exploitation of the data value, especially in the context of AI deployment

-- European Commission, “A European strategy for data”, February 2020



(andrio/Shutterstock)

Analytics And Data Science

Why Is It So Hard to Become a Data-Driven Company?

by Randy Bean

February 05, 2021

“Less than half of companies noted success in these key metrics of progress: Driving innovation with data, Competing on data and analytics, Managing data as a business asset, Forging a data culture, Creating a data-driven organization.”

Big Data and AI Executive Survey 2021.

The next big wave is about KNOWLEDGE

Thesis. Big Data and Data Science have been the rave for the past years, with tremendous successes. However, organizations need to next focus on *data and knowledge culture*.

This requires new roles (and their technologies) who understand, manage and are responsible for the meaning of the data in a data team!

Towards a paradigm shift to realize the full potential of *data- and knowledge-driven* science, industry, and society

Recent data history

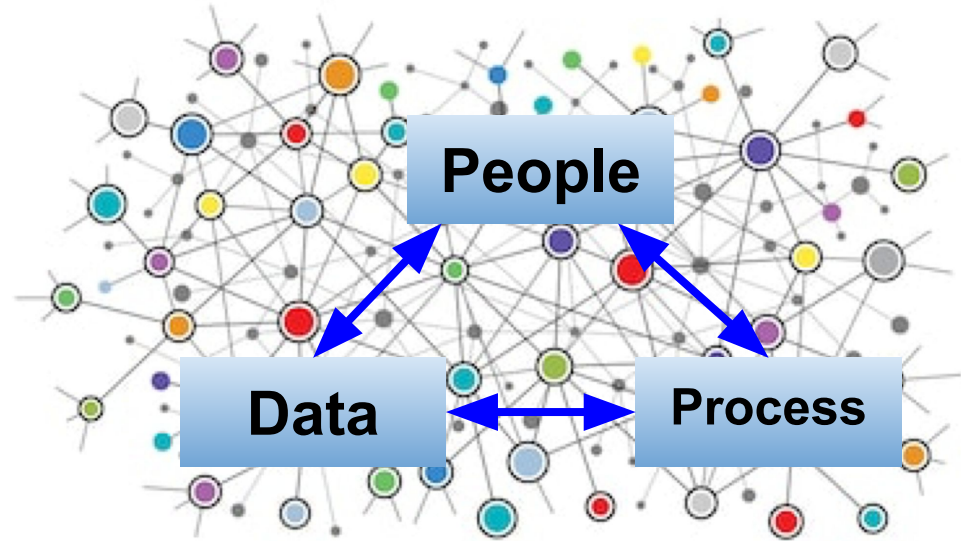
Context	Trend	Organizational Need	Technology	Role
Web + Moore's Law	Big Data	Harness and collect data	Commodity distributed computing platforms (e.g. Hadoop)	Data Engineer
Big Data + GPU Compute	AI Revolution	Draw value from data	Commodity machine learning (e.g., TensorFlow, SciPy)	Data Scientist
AI Revolution + Cloud Computing	Data-Driven Organization, Digital Transformation	<i>Trustworthy data</i>	Clean, meaningful, beautiful data technologies (e.g. knowledge graphs, data wrangling systems, data catalog platforms)	

Towards trustworthy data

Data wrangling (up to 80% of a data scientist's work) is not just grunt work.

The **work of reliable data** is about

- *understanding the ecosystem between people, data and tasks in an organization* and
- *communicating, documenting, and maintaining that knowledge.*



This is vital work **at the heart** of any data-driven organization

Recent data history

Context	Trend	Organizational Need	Technology	Role
Web + Moore's Law	Big Data	Harness and collect data	Commodity distributed computing platforms (e.g. Hadoop)	Data Engineer
Big Data + GPU Compute	AI Revolution	Draw value from data	Commodity machine learning (e.g., TensorFlow, SciPy)	Data Scientist
AI Revolution + Cloud Computing	Data-Driven Organization, Digital Transformation	Rely on data	Clean, meaningful, beautiful data technologies (e.g. knowledge graphs, data wrangling systems, data catalog platforms)	<i>Knowledge Scientist</i>

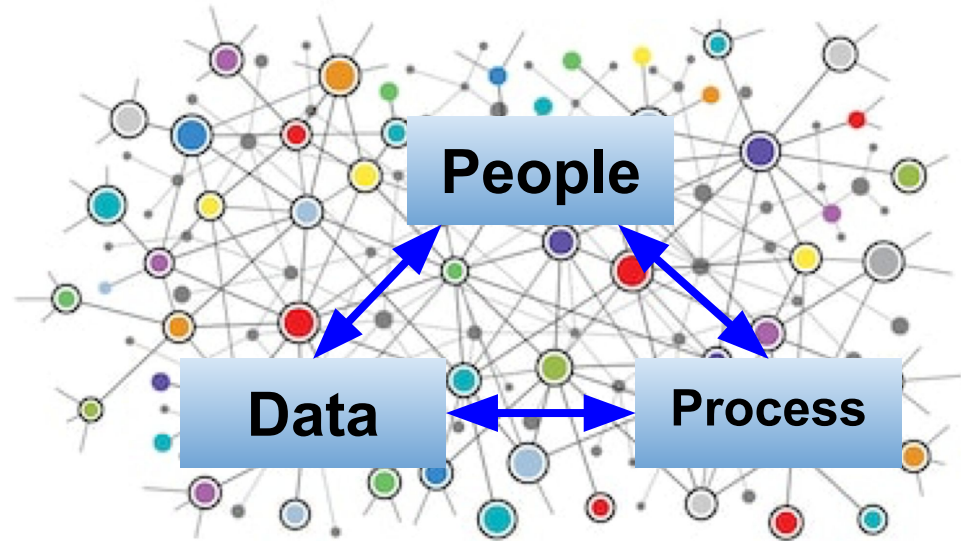
The Knowledge Scientist is responsible for reliable trustworthy data

Who is the Knowledge Scientist?

Knowledge Scientist: the bridge builder between data, people, and their organizational requirements, questions, and needs.

Role: understand and document knowledge from *Users, Data Scientists, Data Engineers*, and *their environment*.

Goal: reliable data for the data-driven organization.



Who is the Knowledge Scientist?

Knowledge Science is people work:

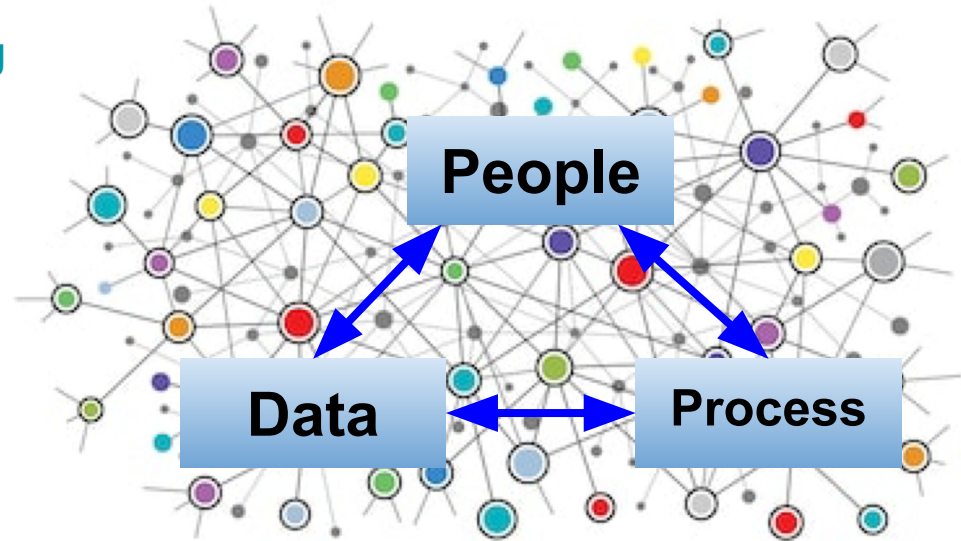
data in context, grounded in shared meaning

- Working together with data consumers and producers for explainable meaningful data

Knowledge Science is technical work:

maintained, accessible, clean data

- Safeguarding actionable, trustworthy, and timely data



Some Implications

For students: Where are there opportunities to be a leader and make impact?

For researchers: New critical research challenges in knowledge at extreme scale in data-driven organizations, in human context

For universities: Are we preparing our students for the next data wave?

For organizations: Are you treating data as a product? Will you be able to grow as an organization without the best data?



insperity

Today's program



Time	Session	Speaker
3:00 - 3:10 PM	Opening	
3:10 - 3:40 PM	University session Chaired by Paul Groth	Elena Simperl
3:40 - 4:40 PM	Knowledge Science in action Chaired by George Fletcher	Pieter Colpaert <i>Knowledge Science for mobility</i> Catia Pesquita <i>Knowledge Science for AI-based biomedical and clinical applications</i> Ilaria Tiddi <i>Knowledge-aware robots</i>
4:40 - 5:20 PM	Industry session Chaired by Juan Sequeda	Mohammed Aaser Tim Gasper
5:20 - 5:30 PM	Closing remarks	

Today's goals

- International experts share their views on knowledge in academia, research, and industry
- Start conversations around implications and future of knowledge science

Takeaways

- Bring software engineering practices to data and knowledge
- Language/ML Models are new forms of knowledge representation. How are these maintained and evaluated?
- Regulatory compliance of data and knowledge
- What is reliable? What is trustworthy?
- Courses on Semantic Web technologies and complemented with how to stop using 80% of your time on finding and cleaning data
- Bias and explainability, in ML models... work for knowledge science?
- We need context. Need to understand users
- Knowledge Science part of CS curriculum, but also in any general Science curriculum to train the next generation of scientist
- Complex/Long ML/data pipelines emphasize the need for knowledge science
- Knowledge Discovery Process 1996, what has really changed? More of everything, but ...do we have *more interpretation*?
- Reality gap! Knowledge fills the gap

Industry Takeaways

- the knowledge world and modern data world are on a collision course
 - analytics engineering --> semantic layer <-- knowledge engineering/science
- importance of data has increased to the executive level, data is a competitive advantage, build a view of customers, suppliers, employees.
- no more silos, bring together to unlock insights
- Data as a Service makes it even more necessary to have roles to curate/manage data supply?
- they haven't invested in common models. If you want to scale, you need to invest in those efforts.
what are the most common requests that we get from our biz stakeholders. how do we standardize that.
- Role or a Skill that other roles should have?
- Knowledge Scientist or a Data Product Manager or Analytics Engineer?
- Skills: Data Modeling, Business (Accounting, Marketing), SQL, Knowledge Graph
- Knowledge science is truly interdisciplinary: CS, Data Sci, Industrial Design, Economics,
- How do you scale data usage in your organization? Get ontology/models right, you can save 50%-80% of effort
- (Knowledge) Graph technology is still not well known. Need to make it easier

The great barrier to AI adoption in healthcare and biomedical research is lack of trust.

Knowledge Science is the answer.

<https://www.knowledgescientist.org/>

Continue the conversation on Slack
<https://knowledgescientist.slack.com/>