PRISME Technical Meeting

Christian Baber

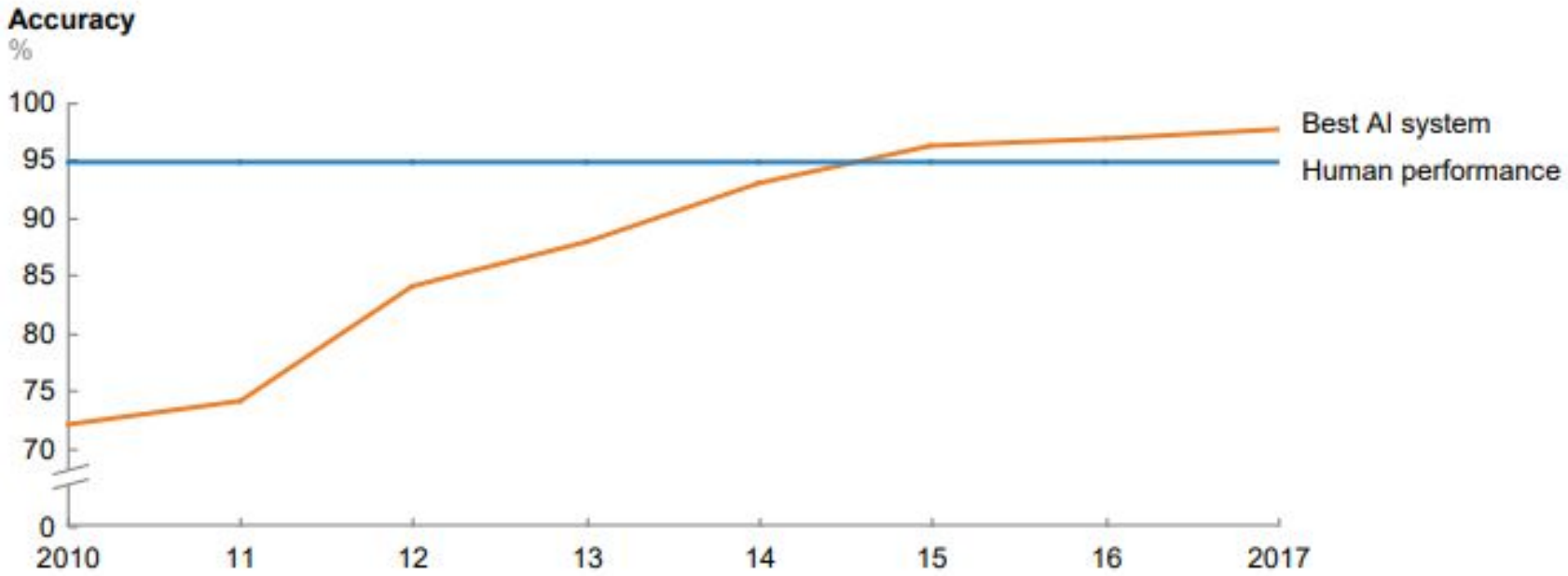
Autumn 2018

Shire

AI Can Exceed Human Performance

Where large amounts of labeled data is available

The ability of AI systems to recognize objects has improved markedly to the point where the best systems now exceed human performance



SOURCE: ImageNet Large Scale Visual Recognition Challenge; McKinsey Global Institute analysis

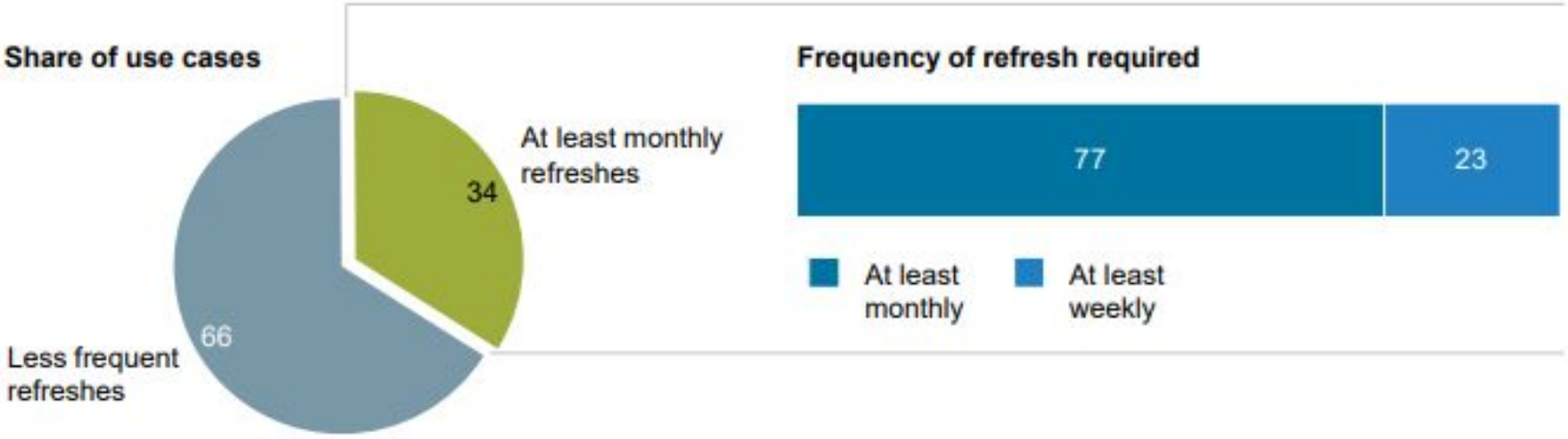
[9] Olga Russakovsky et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, volume 115, issue 3, December 2015.

Shire

But AI Need Regular Retraining

No different from humans but needs available data

For about one-third of use cases, the models require frequent updating: three-quarters of those cases require monthly refreshes, while nearly one-quarter are at least weekly



NOTE: Numbers may not sum due to rounding.

SOURCE: McKinsey Global Institute analysis

Shire

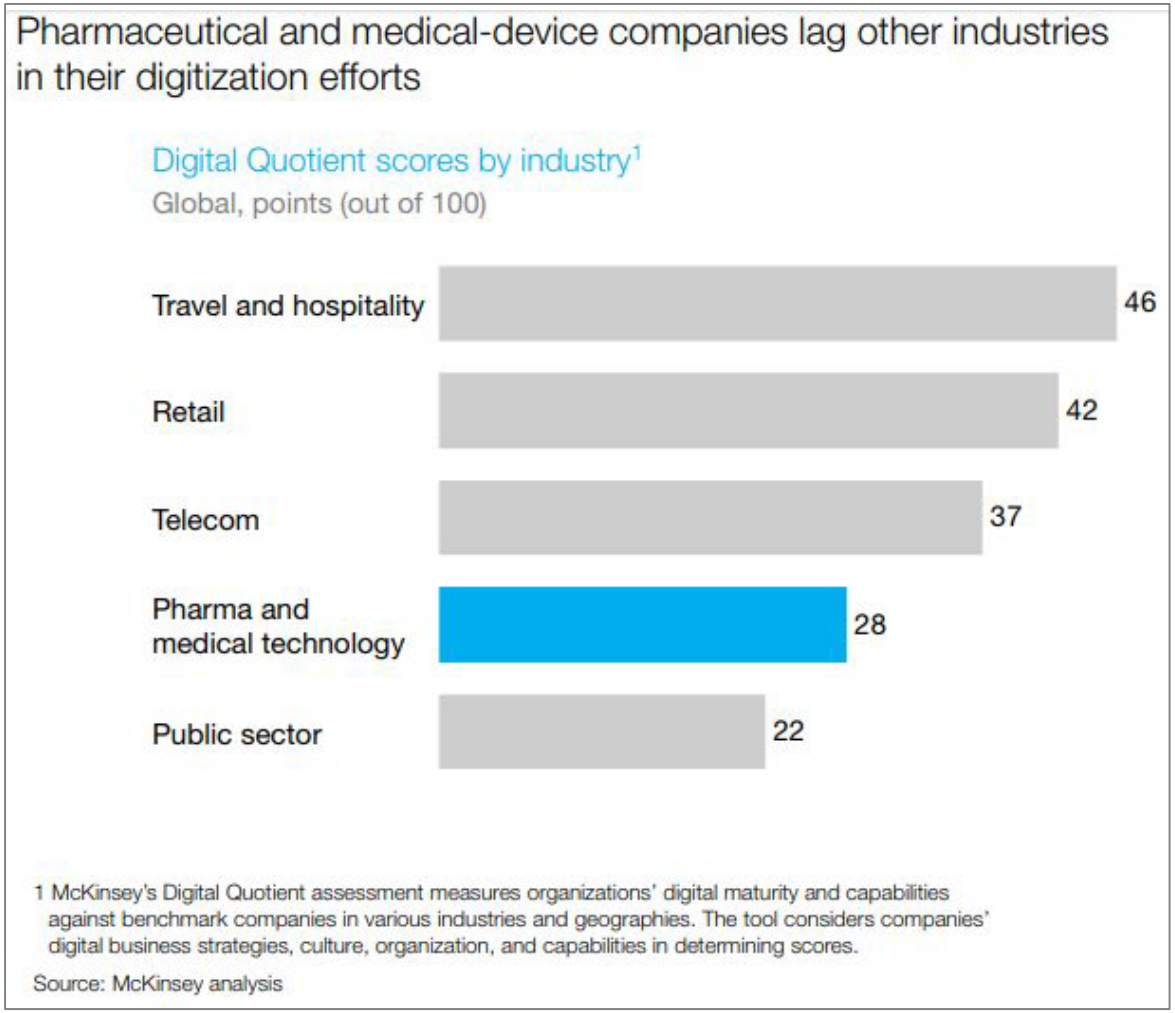
Data as a Fixed Asset
Nothing succeeds like excess!

- **Easier to get right at acquisition than fix up later**
  - Get the metadata right and save time wrangling
- **Gradually depreciates in value**
  - Timeliness and novelty are important
- **Needs to be maintained**
  - Storage, backup, indexing – especially if growing
- **Difficult to move**
  - Data integrity and possibly just volume
- **Only valuable when used**
  - Data holds the *potential* to provide value
  - Value comes from the ability to drive decisions
- **Can be used multiple times**
  - No degradation may even increase as more data is accumulated

Shire

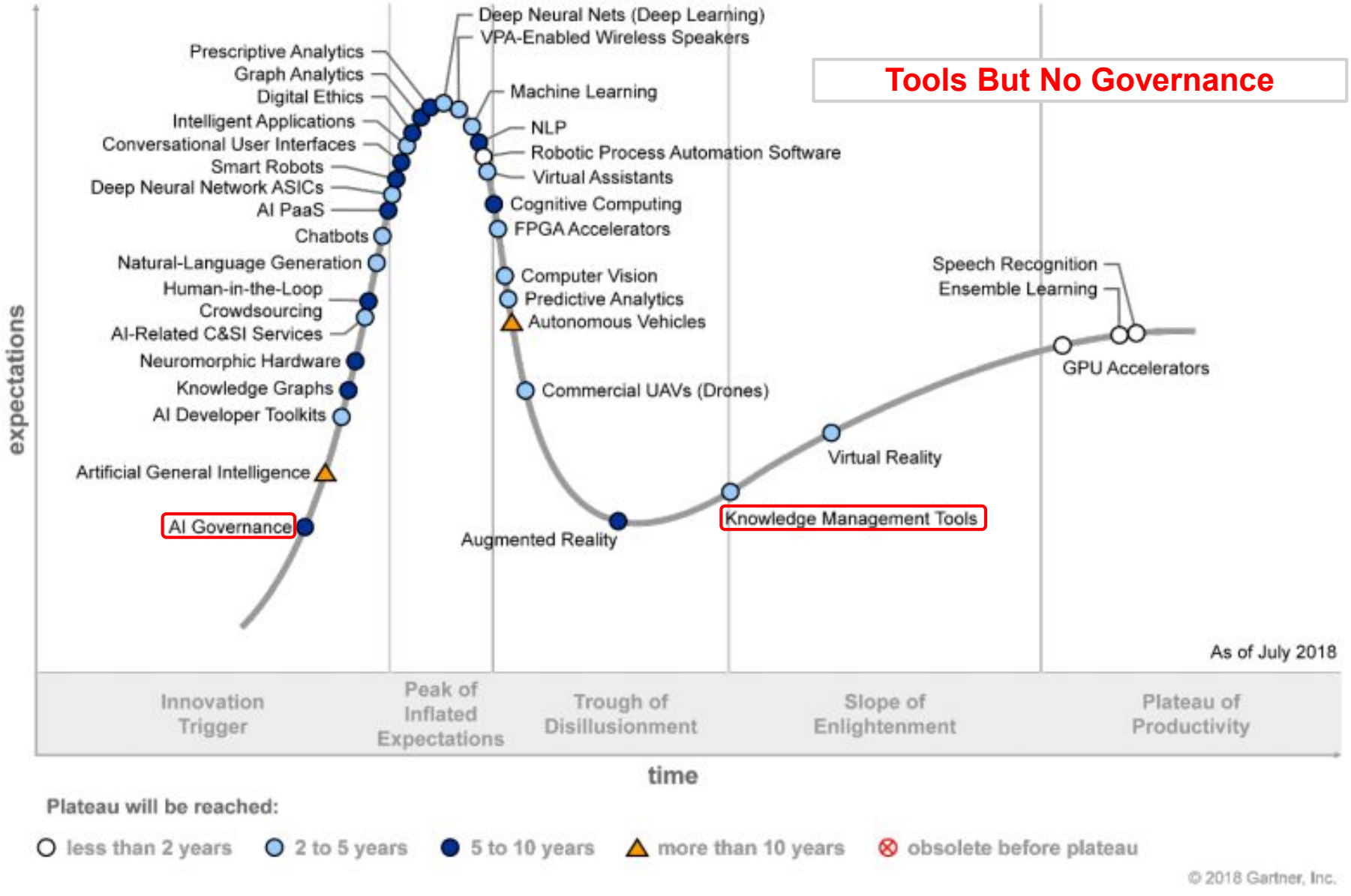
Opportunity to Learn from Other Sectors

The grass is greener…

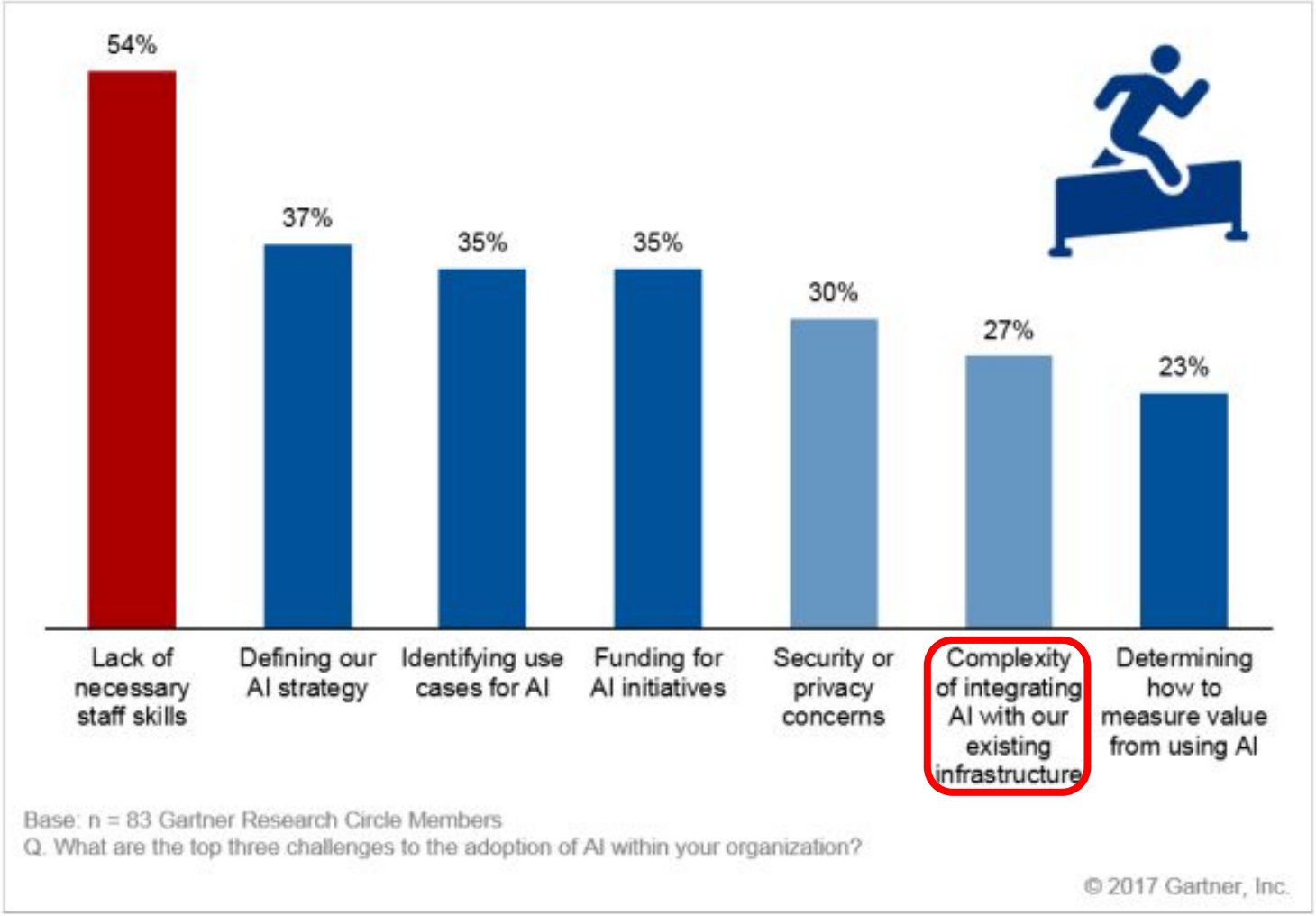


Companies with richly labeled data are ahead

Do we have the Technology?
In principal but maybe not in practice



Informatics Implementation Challenges
Unknown unknowns



Where are challenges around data?

Timely Topic
Get the most from our assets

Plenary - Theory

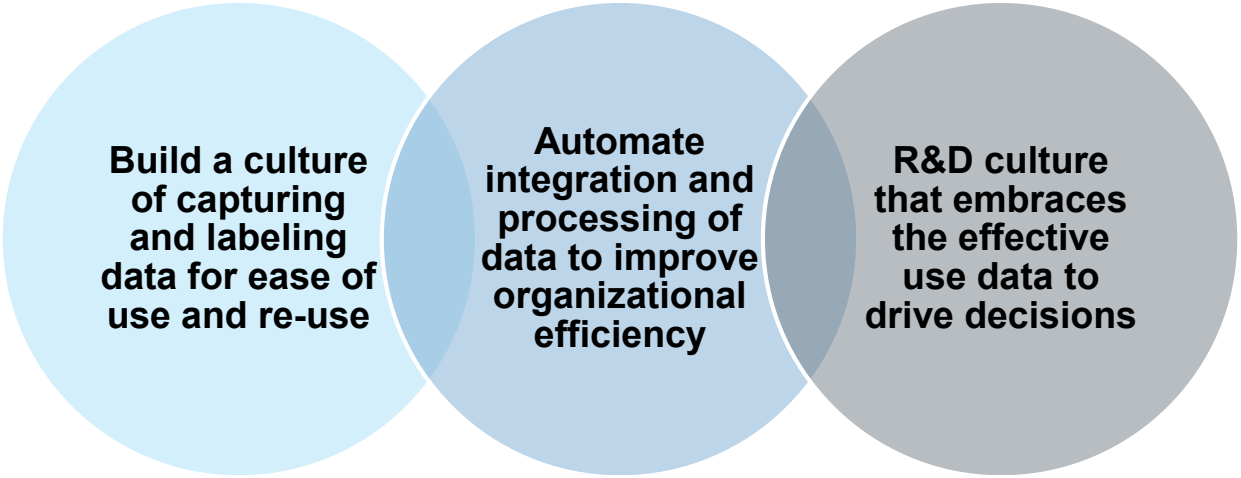
- John Overington (Medicines Discovery Catapult) – Data for Drug Discovery
- Martin Romacker (Roche) - Data First : FAIR Information Procurement

Startups & Posters
David Christie & Dan Chapman

- 5 outstanding startups from a much larger field
- Posters from ontologies to data collection to drug discovery
- Something for everyone

Plenary - Practice

- Andrew Carroll (Google) – Exploiting the data you have
- Mark Musen (Stanford) – Semantic Technology to manage knowledge



<u>North stars</u> to think of data as R&D‘s lifebood with value coming from use

Advisory Committee

Standing on the shoulders of giants

- **Nick Brown**, Head of Technology Incubation Lab, *AstraZeneca*
- **Dan Chapman**, Head of IT New Med. Information Management, *UCB*
- **David Christie**, Vice President, Enterprise Applications Group, *CSL Behring*
- **Lars Greiffenberg**, Director – R&D IT and Translational Informatics, *Abbvie*
- **Carol Rohl**, Executive Director, Scientific Information Management, *Merck*
- **Martin Romacker**, Principal Scientist – Data and Information Architecture, *Roche*
- **Jason Tetrault**, Global Head Data Engineering and Emerging Technologies, Takeda
- **Jianchao (JC) Yao**, Associate Principal Scientist, *Merck*

John Wise – Program Coordinator
Nico Stanculescu – Logistics Coordinator

Shire