# Research Data Management: Pushing the Frontiers of Good Research Practice


***Yaşar Tonta***

**Hacettepe University**

**Department of Information Management**

**yasartonta@gmail.com**

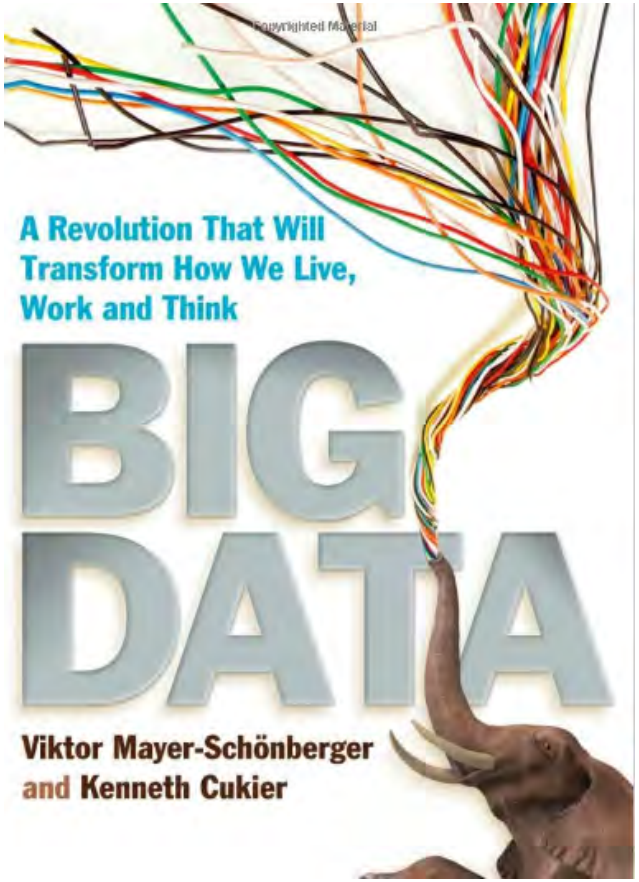
**@yasartonta**

**yunus.hacettepe.edu.tr/~tonta/**

# Plan

- Big data
- Text and Data Mining (TDM)
- Research data management
- Data management policies
- Conclusion

# Big data

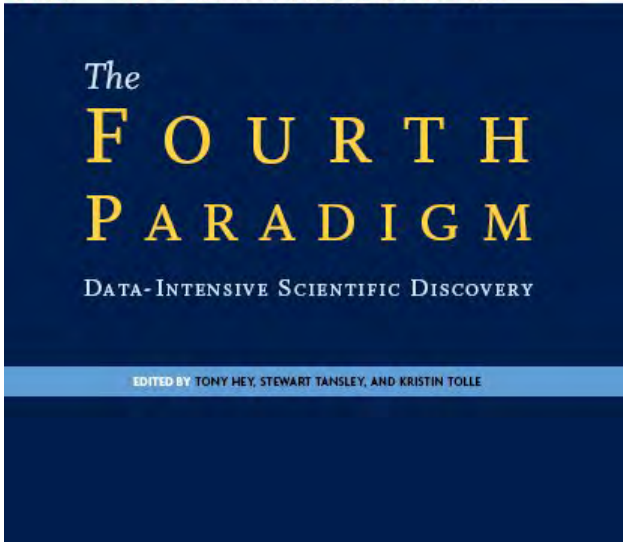



- 2,5 quintillion bytes of data (2Exabyte) created per day
- 90% created in the last 2 years
- Data doubling every 3 years
- Google processes 24 petabyte of data every day
- Google's prediction of the spread of winter flu based on data

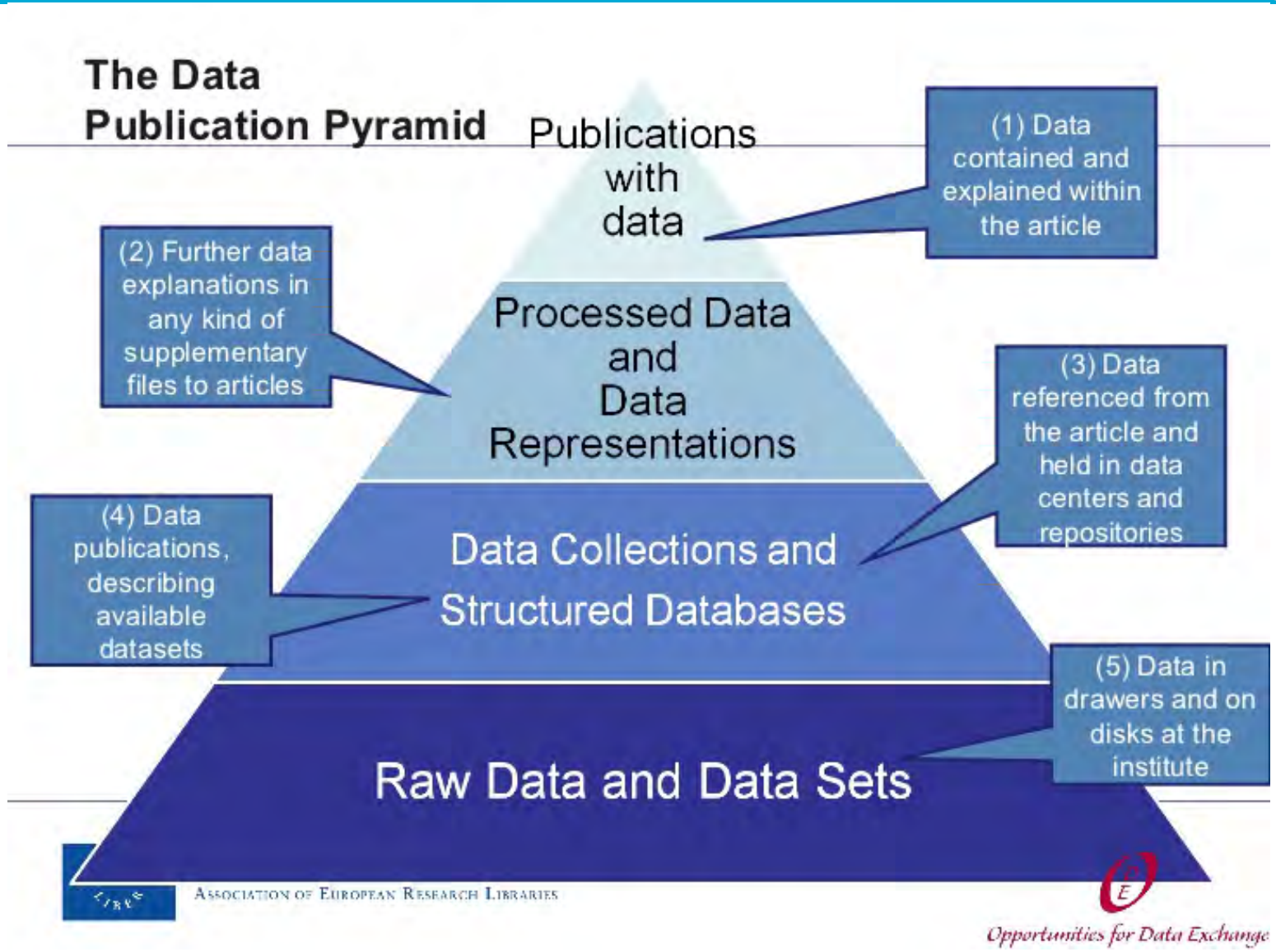
# The Fourth Paradigm

- Thousand years ago:
  science was **empirical**
  describing natural phenomena
- Last few hundred years:
  **theoretical** branch
  using models, generalizations
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today:
  **data exploration** (eScience)
  synthesizing theory, experiment and computation with advanced data management and statistics

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

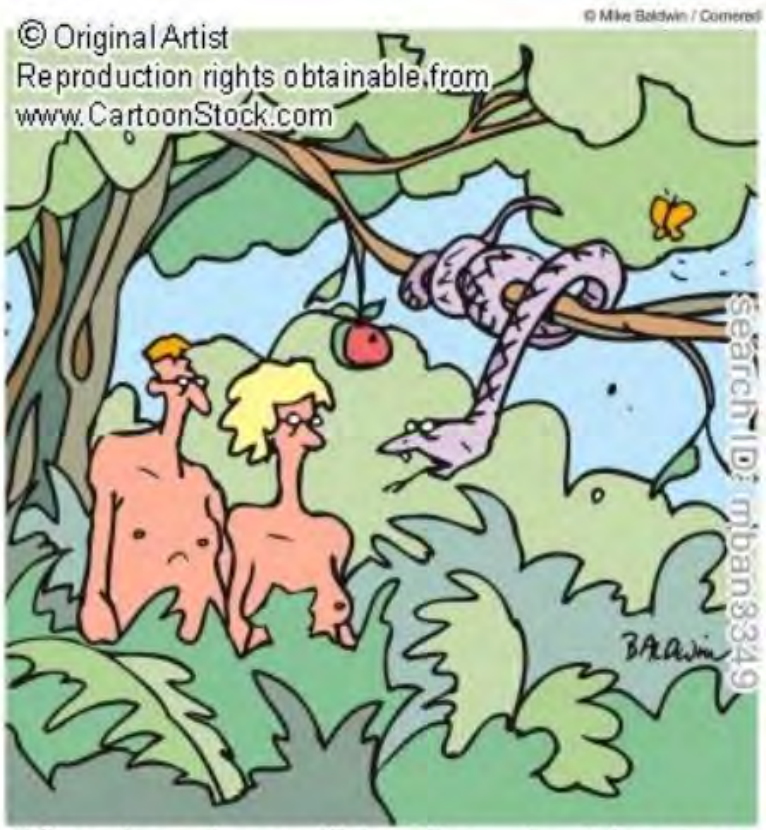


# All scientific papers & data online




http://www.slideshare.net/libereurope/research-data-sharing-leru

# Mine, all mine!




- Text and data mining (TDM)
- Right to read = right to mine?
- TDM adoption & uptake
- TDM & IPR
  - Exception in copyright law
  - Licensing
  - Commercial vs. noncommercial use

# Why share the data?


- By unlocking the potential “Big data” European economies could save between €150 billion and €300 billion annually
- Direct or indirect impact across the EU at 140 billion annually

Source: Open data access policies and strategies in the European Research Area and beyond. August 2013.

# Why manage research data?

- Research expenditures (60 bn USD in per year in USA, circa 80 bn EUR in Horizon 2020)
- Data growth rate is 30% pa (the human genome produces 30 TB data, CERN produces 30PB data every year)
- Boost scientific discoveries
- Verify findings / avoid «bad science»
- Return on investment

Source: Managing research data, p. 2;
For citation advantage: https://peerj.com/preprints/1.pdf

# Data management policies

- USA: requires 2 pages of data management plan for NSF projet proposals (2011)
- EU Horizon 2020 requires a DMP within 6 months of project's start (if included in Open Data Pilot)
- UK: Access to many output of the research it funds, along with data management, long-term preservation. But there is no sanction yet.
- Who is taking care of our research data? Academic research community?

Source: Managing Research Data, s. 5

# Research data infrastructure

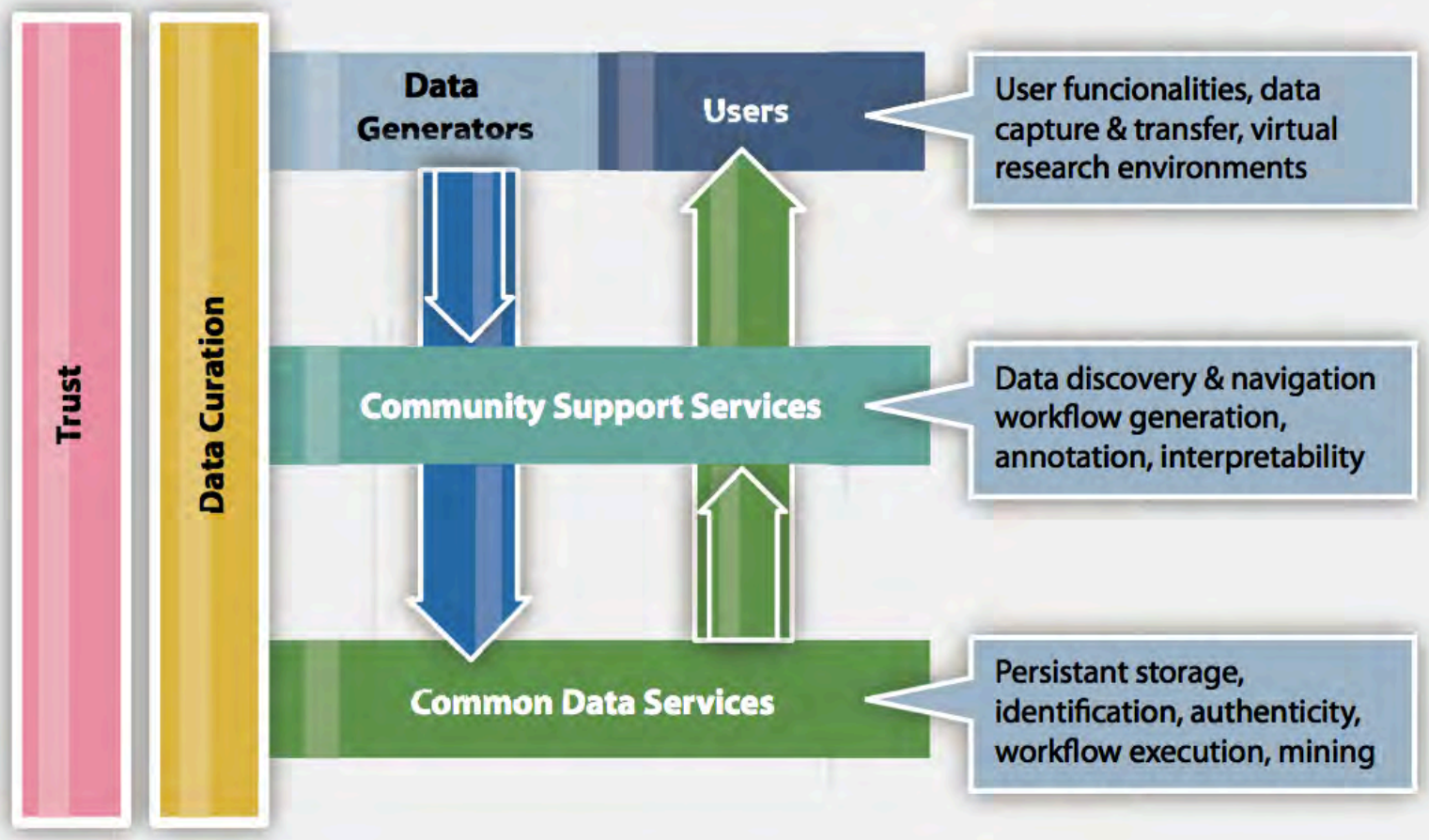
"Without the infrastructure that helps scientists manage their data in a convenient and efficient way, no culture of data sharing will evolve."

Stefan Winkler-Nees
(German Research Foundation, DFG)

http://www.slideshare.net/libereurope/research-data-sharing-leru

Opportunities for Data Exchange

# Collobarative Data Infrastructure – a framework for the future




Source: http://www.slideshare.net/libereurope/research-data-sharing-leru


OPEN YOUR
DATA
AND
KEEP
CALM

# Research Data Management: Pushing the Frontiers of Good Research Practice


***Yaşar Tonta***

**Hacettepe University**

**Department of Information Management**

**yasartonta@gmail.com**

**@yasartonta**

**yunus.hacettepe.edu.tr/~tonta/**