KNOW Center

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

Science 2.0 & Big Data

Science 2.0 Conference, Hamburg, March 25, 2015

Prof. Dr. Stefanie Lindstaedt

www.know-center.at

© Know-Center GmbH

Know-Center

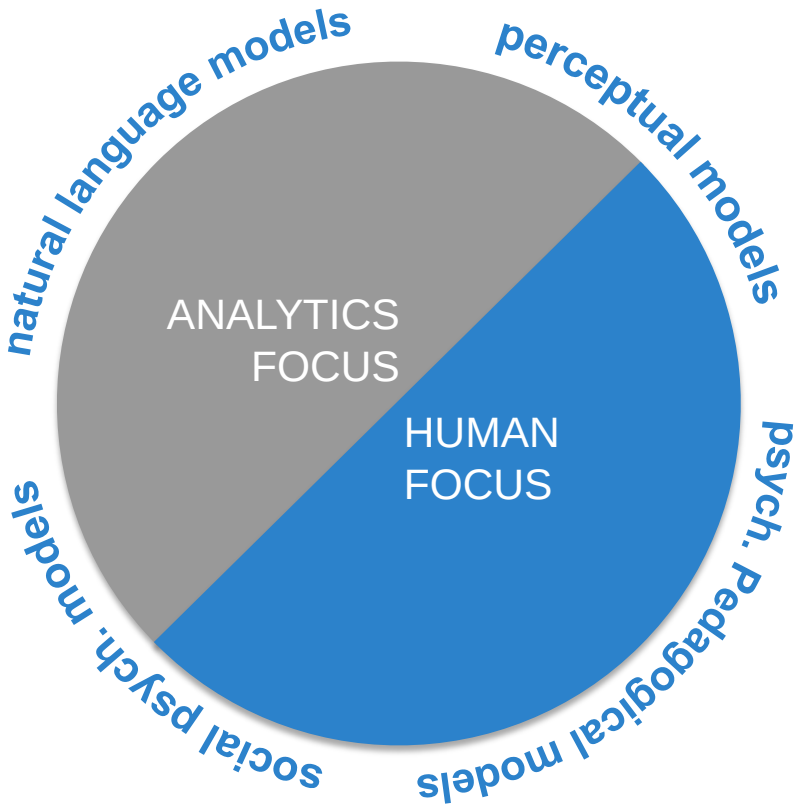
Austria‘s Research Center for Data-driven Business and Big Data Analytics

- founded 2001
- >70 researchers
- >450 applied research projects with company partners
- >25 European funded research projects
- **just received > 20 Mio € funding for next 4 years**

Data-driven Business as a Cognitive Computing Challenge

Knowledge Discovery

Knowledge Visualization



Social Computing

Ubiquitous Personal Computing

Cognitive Computing Systems interact naturally with humans, learn from their experiences, generate and evaluate evidence-based hypotheses

Big Data

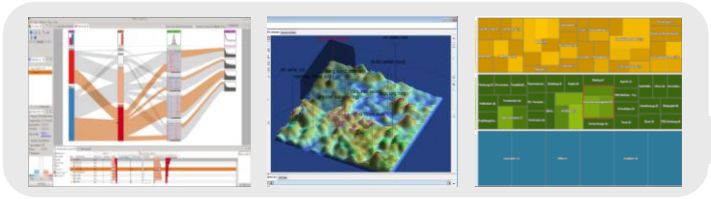
"Data unprecedented in its scale and scope
in relation to a given phenomenon
which allows for the generation of new knowledge."

[Oxford Internet Institute, 2014]

➔ **Increased potential to gain insights**

Big Data Analytics

Recommendations – Analyses – Visualizations



Human

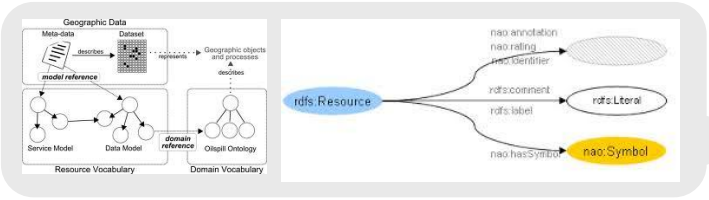
Intelligent Algorithms

Analytics

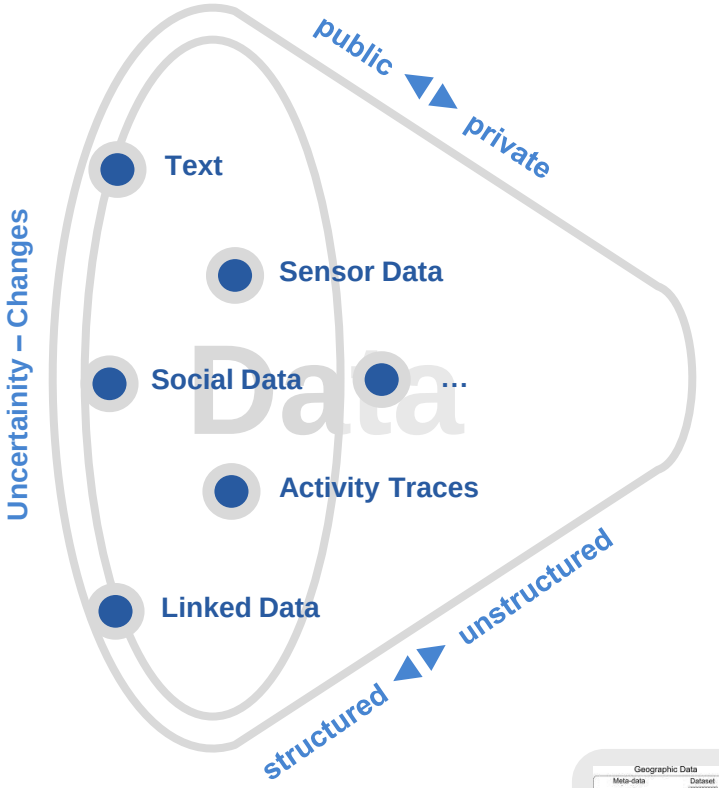
Scientist in the Loop

Actionable Knowledge

Computer



Models – Interfaces – Representations



Data-driven Business

- Business processes which are based on the automatic generation, interpretation, and exploitation of large amounts of information and data.

- Four central steps

  (1) Provide appropriate data & IT infrastructure

  (2) Democratize data within the company

  (3) Enable experimentation with data

  (4) Support data-driven culture

[LinkedIn, 2013]

Data-driven Science

- **Research practices** and processes which are based on automatic generation, interpretation, and exploitation of large amounts of information and data.

- Four central steps

  (1) Provide appropriate data & IT infrastructure

  (2) Democratize data across the community

  (3) Enable experimentation with data

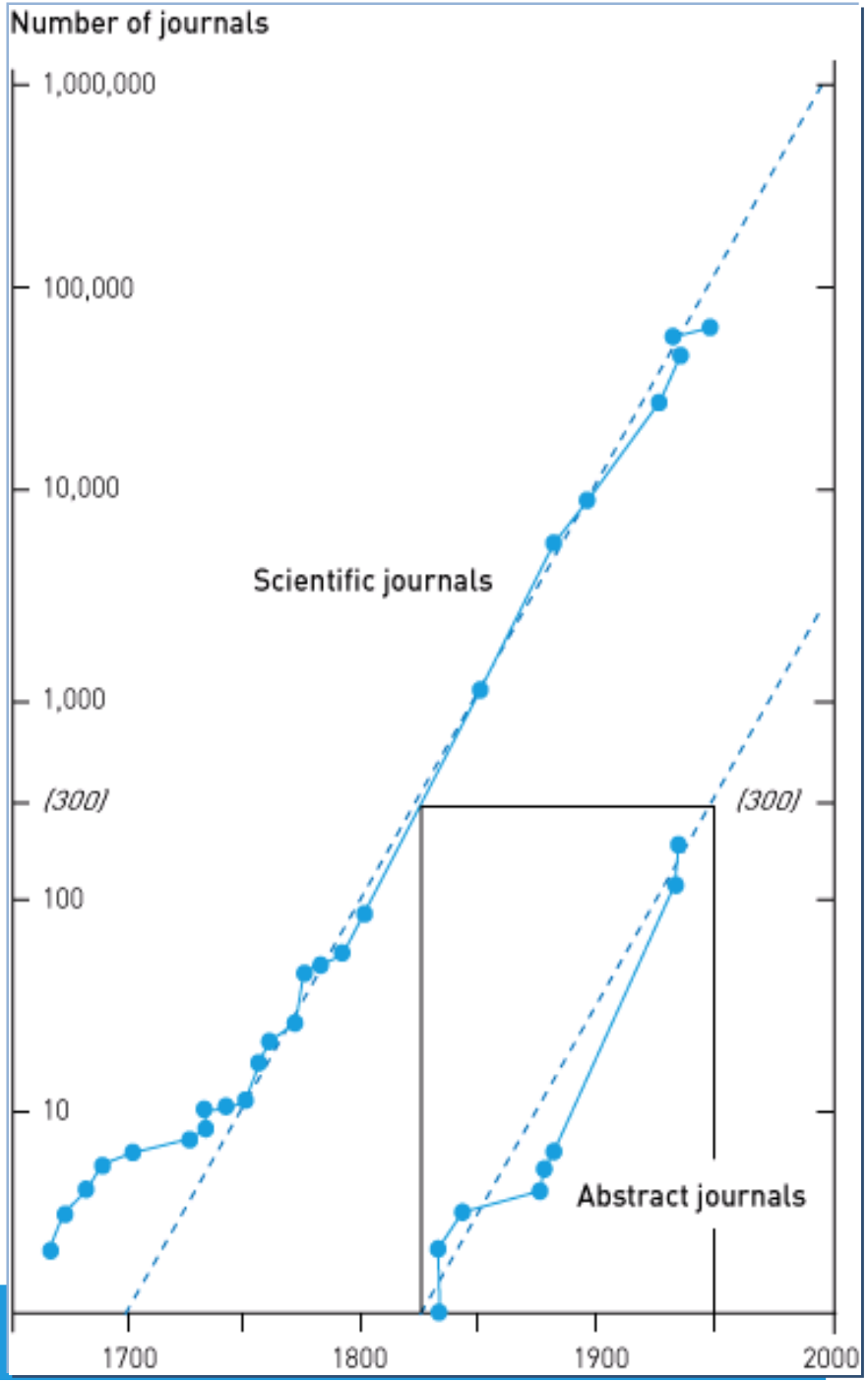
  (4) Support data-driven culture

PUBLICATIONS AS (BIG) DATA

**“One of the diseases of this age is the multiplicity of books;
they doth so overcharge the world that it is not able to digest the abundance of idle matter that is every day hatched and brought forth into the world.“**

Attributed to Barnaby Rich in 1613

Publications as Big Data



[Price, 1963]

The Vision …

- Research builds on past insights
- We share knowledge to create new knowled

[Isaac Newton, 1676]

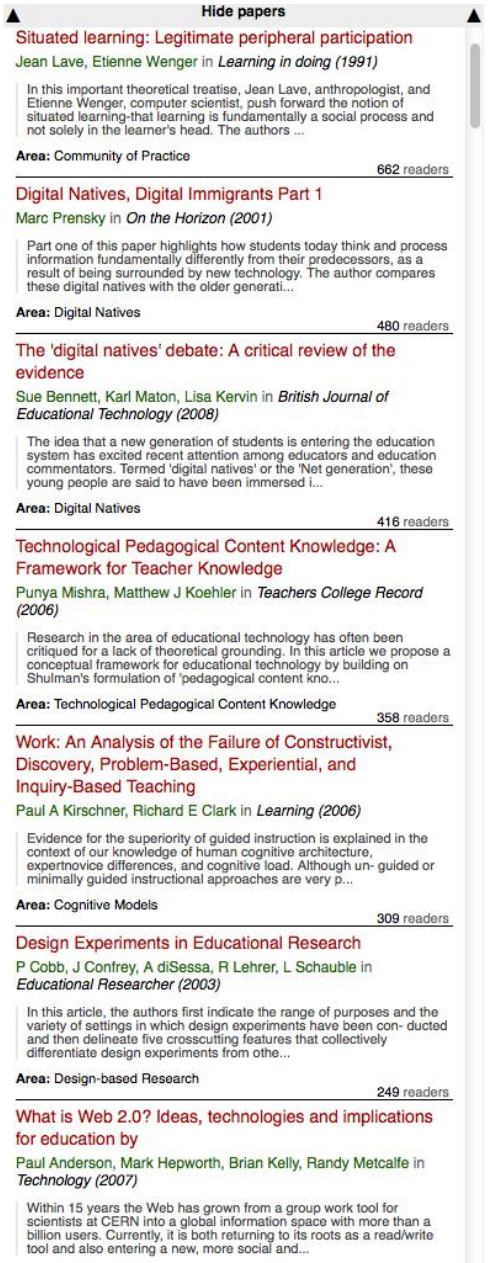
We are dwarfs standing on the shouldes of giants

The Reality ….

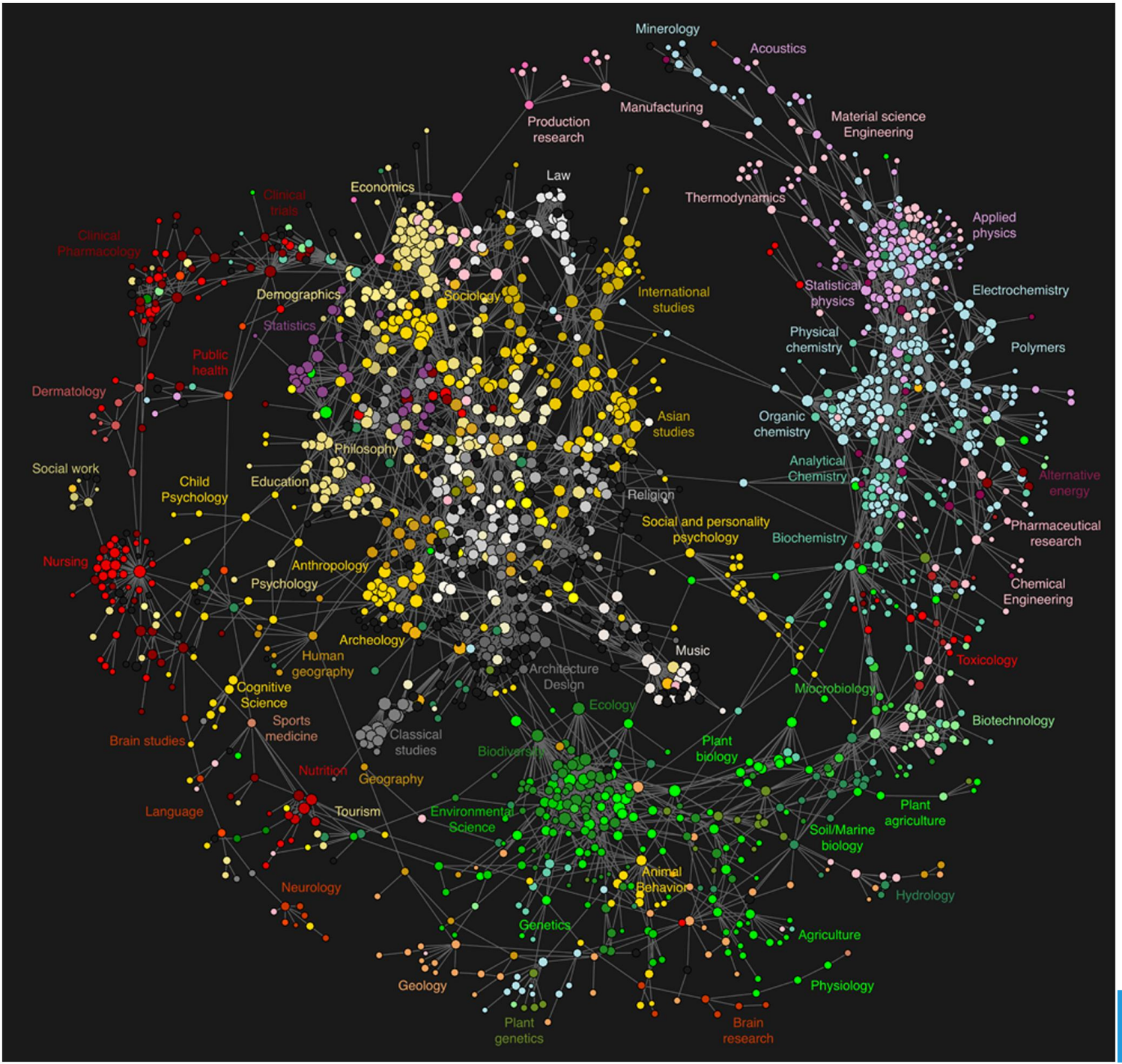
Covered under a pile of paper

- .. with varying quality
- .. with contradicting facts
- .. with missing data

Overview of a Research Domain based on Usage Data



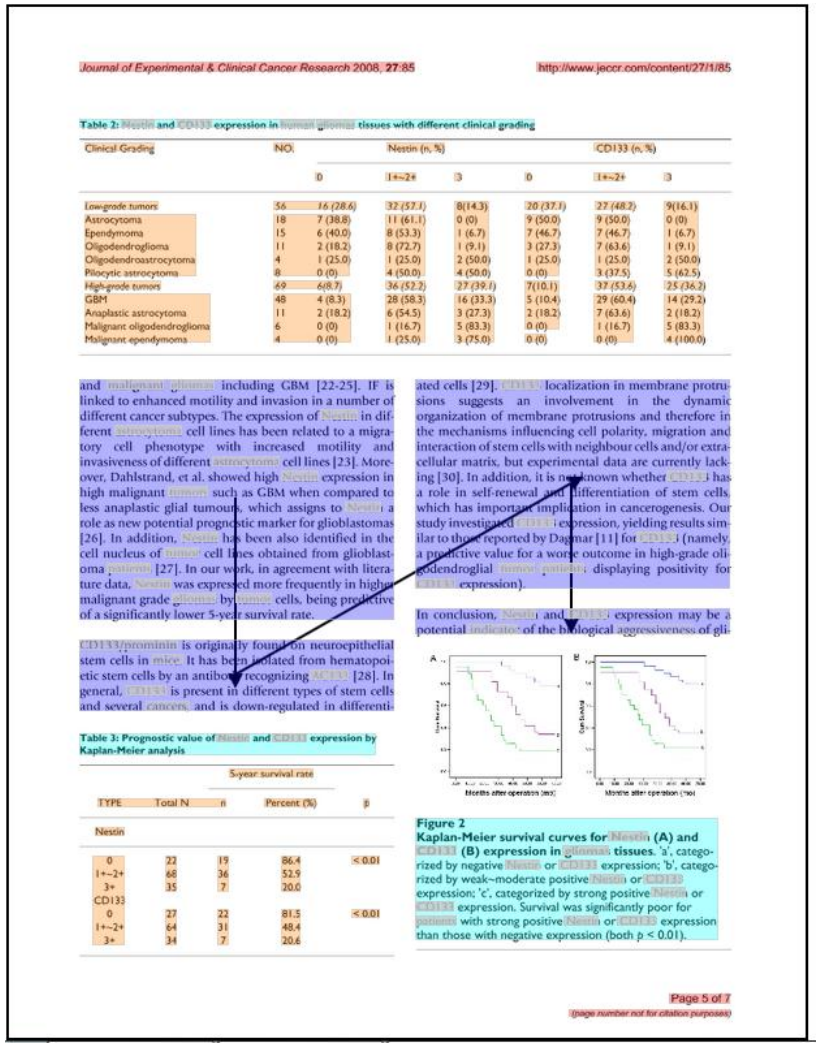
[Kraker, 2013]



Cross-Journal Recommen-dation based on Click Streams

[Bollen et al., 2009]

Extract facts from research papers



Journal of Experimental & Clinical Cancer Research 2008, 27:85 http://www.jeccr.com/content/27/1/85

Table 2: Nestin and CD133 expression in human glioma tissues with different clinical grading

Clinical Grading	NO.	Nestin (n, %)			CD133 (n, %)		
		0	1+~2+	3	0	1+~2+	3
Low-grade tumors	56	16 (28.6)	32 (57.1)	8(14.3)	20 (37.1)	27 (48.2)	9(16.1)
Astrocytoma	18	7 (38.8)	11 (61.1)	0 (0)	9 (50.0)	9 (50.0)	0 (0)
Ependymoma	15	6 (40.0)	8 (53.3)	1 (6.7)	7 (46.7)	7 (46.7)	1 (6.7)
Oligodendroglioma	11	2 (18.2)	8 (72.7)	1 (9.1)	3 (27.3)	7 (63.6)	1 (9.1)
Oligodendroastrocytoma	4	1 (25.0)	1 (25.0)	2 (50.0)	1 (25.0)	1 (25.0)	2 (50.0)
Pilocytic astrocytoma	8	0 (0)	4 (50.0)	4 (50.0)	0 (0)	3 (37.5)	5 (62.5)
High-grade tumors	69	6(8.7)	36 (52.2)	27 (39.1)	7(10.1)	37 (53.6)	25 (36.2)
GBM	48	4 (8.3)	28 (58.3)	16 (33.3)	5 (10.4)	29 (60.4)	14 (29.2)
Anaplastic astrocytoma	11	2 (18.2)	6 (54.5)	3 (27.3)	2 (18.2)	7 (63.6)	2 (18.2)
Malignant oligodendroglioma	6	0 (0)	1 (16.7)	5 (83.3)	0 (0)	1 (16.7)	5 (83.3)
Malignant ependymoma	4	0 (0)	1 (25.0)	3 (75.0)	0 (0)	0 (0)	4 (100.0)

and malignant glioma including GBM [22-25]. IF is linked to enhanced motility and invasion in a number of different cancer subtypes. The expression of Nestin in different astrocytoma cell lines has been related to a migratory cell phenotype with increased motility and invasiveness of different astrocytoma cell lines [23]. Moreover, Dahlstrand, et al. showed high Nestin expression in high malignant tumors such as GBM when compared to less anaplastic glial tumours, which assigns to Nestin a role as new potential prognostic marker for glioblastomas [26]. In addition, Nestin has been also identified in the cell nucleus of tumor cell lines obtained from glioblastoma patients [27]. In our work, in agreement with literature data, Nestin was expressed more frequently in higher malignant grade gliomas by tumor cells, being predictive of a significantly lower 5-year survival rate.

CD133/prominin is originally found on neuroepithelial stem cells in mice. It has been isolated from hematopoietic stem cells by an antibody recognizing AC133 [28]. In general, CD133 is present in different types of stem cells and several cancers, and is down-regulated in differentiated cells [29]. CD133 localization in membrane protrusions suggests an involvement in the dynamic organization of membrane protrusions and therefore in the mechanisms influencing cell polarity, migration and interaction of stem cells with neighbour cells and/or extracellular matrix, but experimental data are currently lacking [30]. In addition, it is not known whether CD133 has a role in self-renewal and differentiation of stem cells, which has important implication in cancerogenesis. Our study investigated CD133 expression, yielding results similar to those reported by Dagmar [11] for CD133 (namely, a predictive value for a worse outcome in high-grade oligodendroglial tumor patients displaying positivity for CD133 expression).

In conclusion, Nestin and CD133 expression may be a potential indicator of the biological aggressiveness of gli-

Table 3: Prognostic value of Nestin and CD133 expression by Kaplan-Meier analysis

TYPE	Total N	n	Percent (%)	p
	5-year survival rate			
Nestin				
0	22	19	86.4	< 0.01
1+~2+	68	36	52.9	
3+	35	7	20.0	
CD133				
0	27	22	81.5	< 0.01
1+~2+	64	31	48.4	
3+	34	7	20.6	

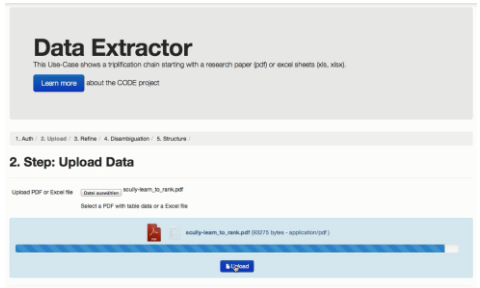
Figure 2
Kaplan-Meier survival curves for Nestin (A) and CD133 (B) expression in glioma tissues. 'a', categorized by negative Nestin or CD133 expression; 'b', categorized by weak~moderate positive Nestin or CD133 expression; 'c', categorized by strong positive Nestin or CD133 expression. Survival was significantly poor for patients with strong positive Nestin or CD133 expression than those with negative expression (both $p < 0.01$).

Page 5 of 7
(page number not for citation purposes)

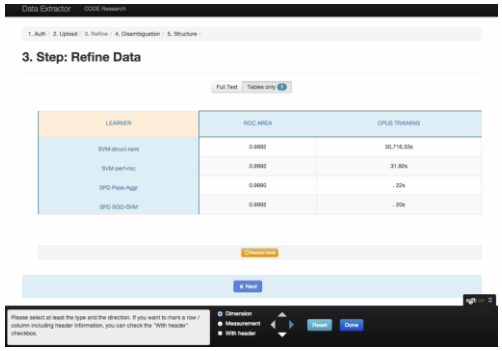
- Link research papers and the facts therein to LOD
- Extract information from PDFs
  - Tables, figures, structure, references, named entities
- Integration of LOD concepts into papers

Integrate facts with existing knowledge

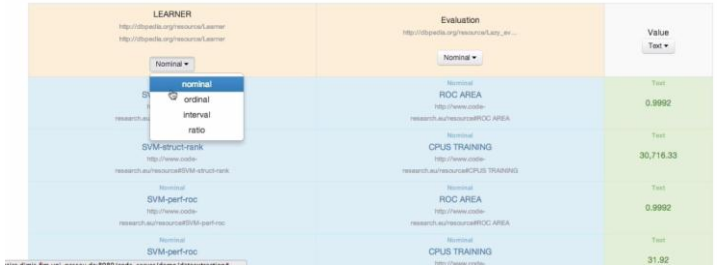
1.

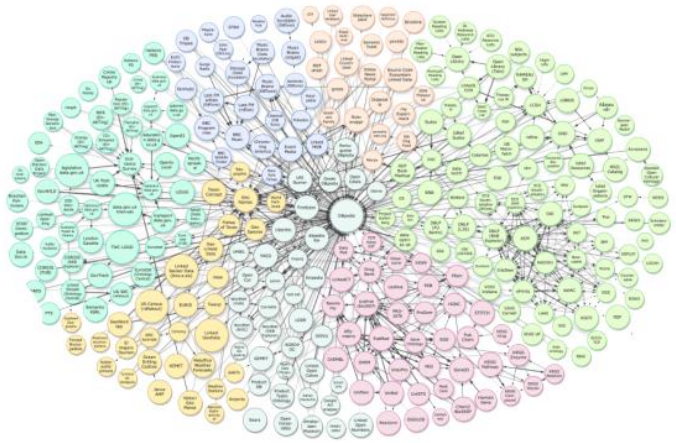


2.



3.



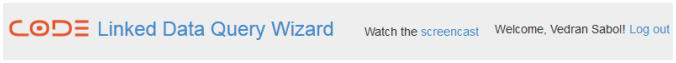


- Objective: crowd-sourced triplification for statistical data
- Semi-automatic creation, storing and merging of statistical data

Make facts available for visual analysis

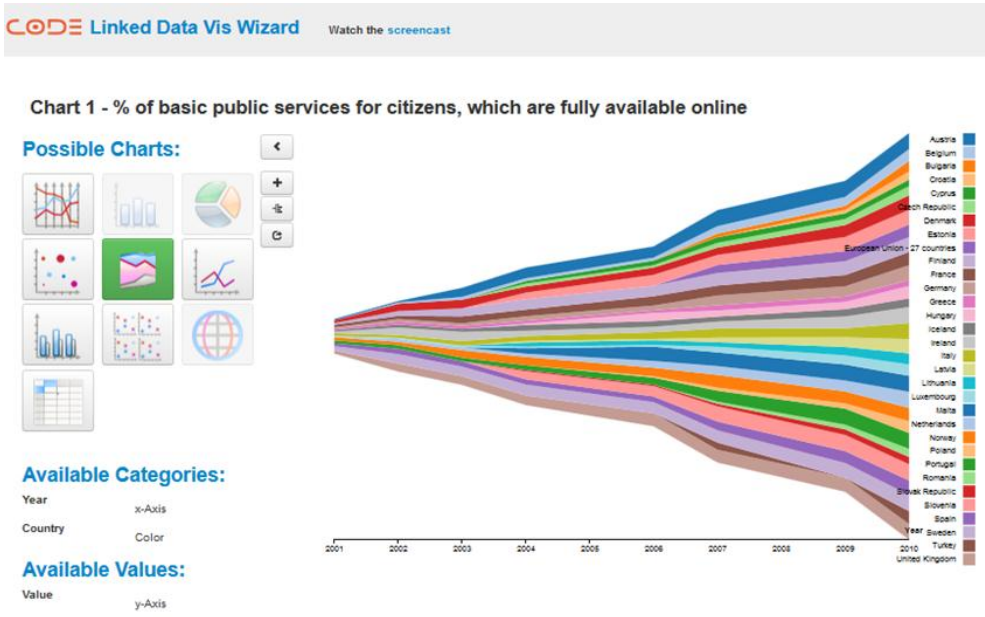
Query Wizard and Vis Wizard: designed for IT-laymen

1.



2.

3.



Summary:
Turning Publications into Scientific (Big) Data

- Based on usage data and click streams we can
  - Generate an overview of a given research domain
  - Provide cross-journal recommendations of relevant articles

- Given textually encoded scientific knowledge, we can
  - Extract facts from research papers
  - Integrate those facts with existing knowledge
  - Make it available for visual analysis

SUPPORT DATA-DRIVEN CULTURE

Socialising Research Data

- Observation 1: Open Data Platforms today
  - Domain specific provider (e.g. EU digital scoreboard, Eurostats)
  - Data centred (e.g. data set list/management, Datahub/CKAN)

➔ **Great data (but unsocial)**

- Observation 2: success of the Social Web
  - Successful web platforms are social: blog, discuss, share, bookmark
  - People engage with the digital item making it more valuable

➔ **Socialising resources as success factor**

Socializing Research Data

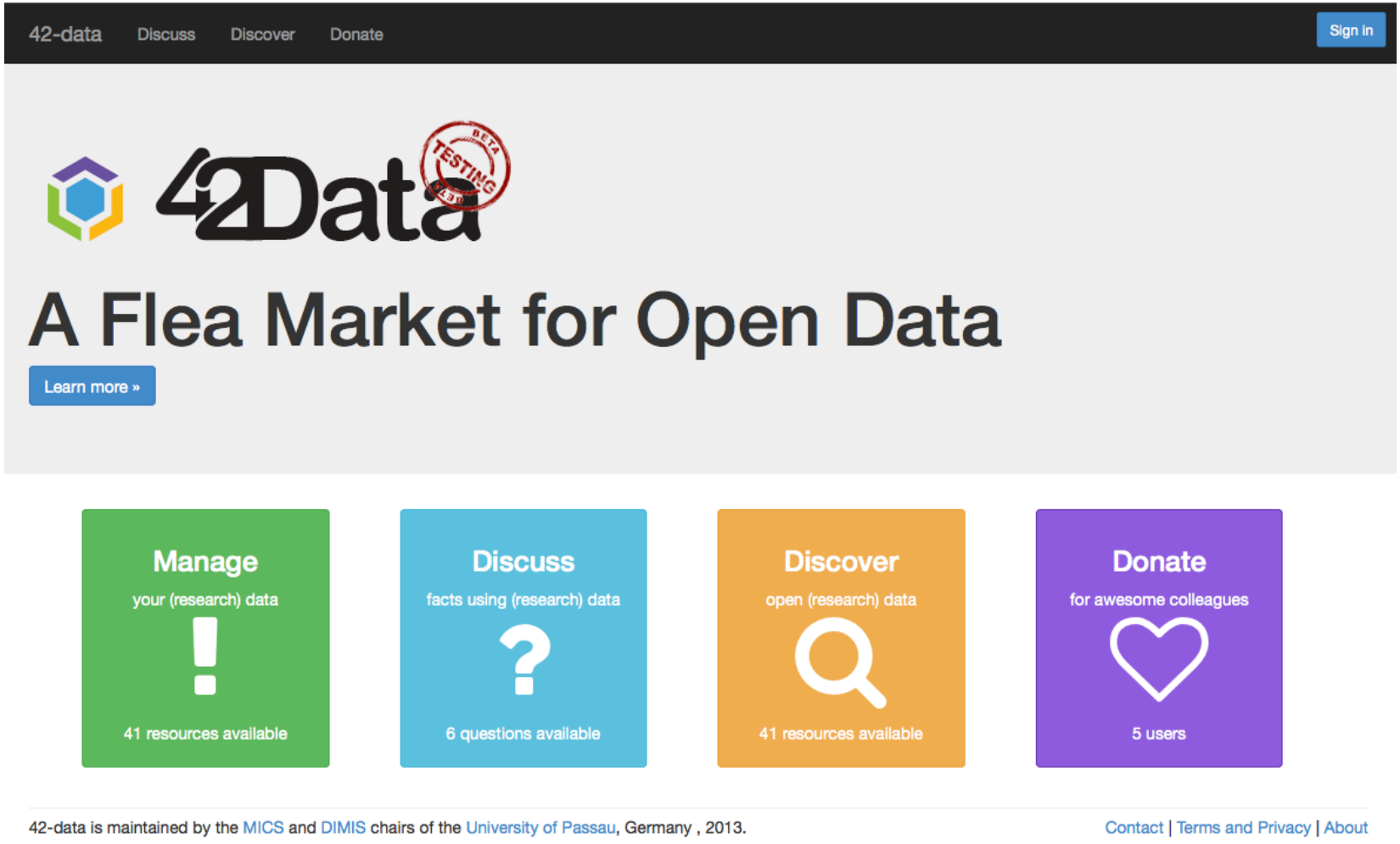
- Observation 3: Open Data are not easy to consume
  - No mass market (such as for images or videos)
  - Socialising around data has to focus on a special group of users and needs (interpretation)

➔ **Specialised services for socialising data are necessary**

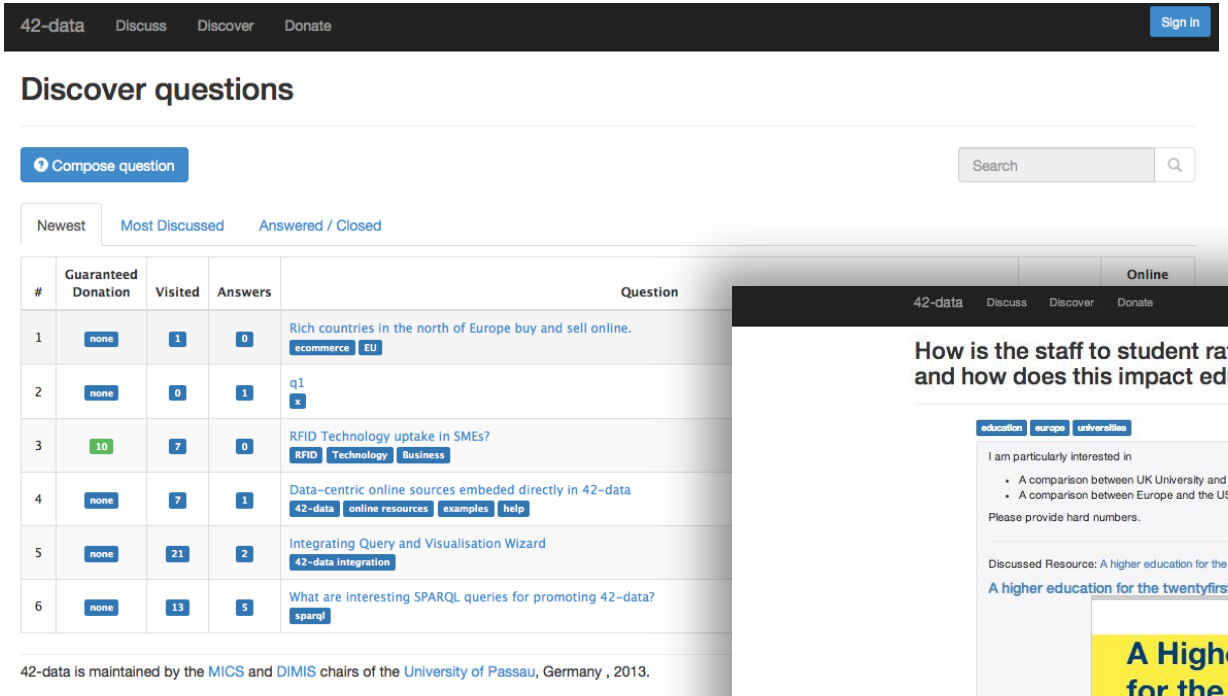
- Observation 4: Research Data Properties
  - Raw data is less of importance for sharing knowledge
  - Needs interpretation from different points of view (socialisation)
  - Currently sharing but no socialising in LOD

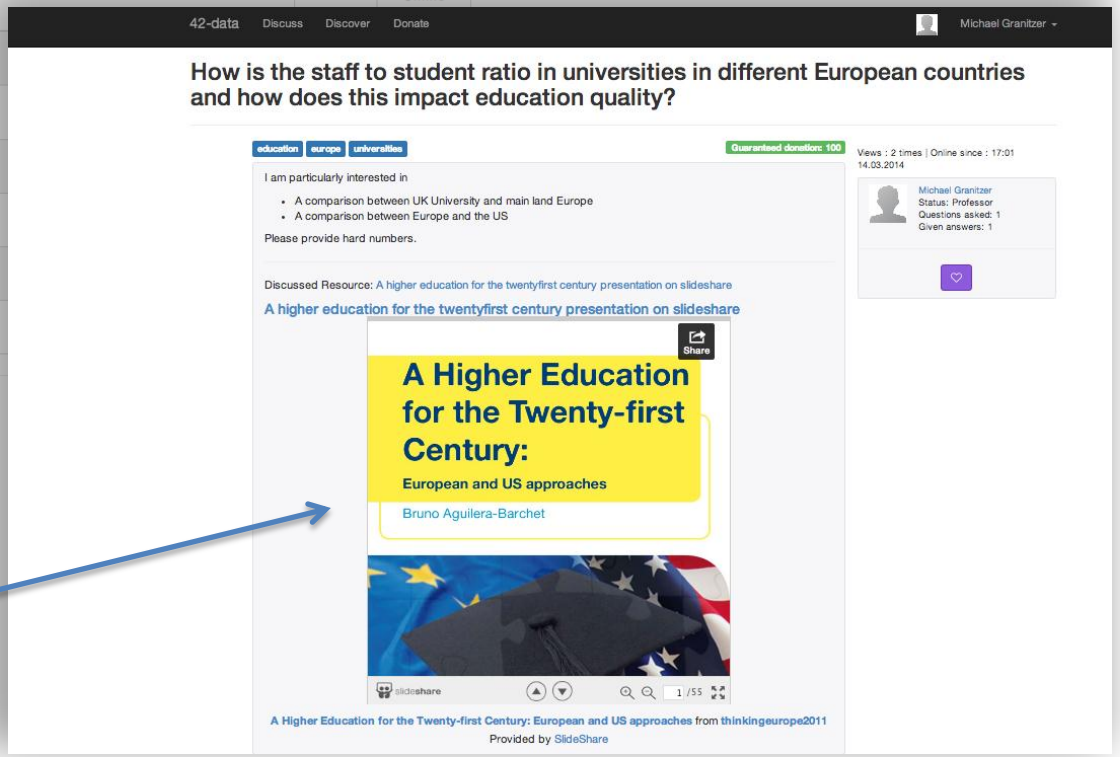
➔ **Socializing research data for interpretation and generating insights**

42-Data


42-data
Discuss
Discover
Donate
Sign in
42Data
BETA TESTING
A Flea Market for Open Data
Learn more »
Manage
your (research) data
41 resources available
Discuss
facts using (research) data
6 questions available
Discover
open (research) data
41 resources available
Donate
for awesome colleagues
5 users
42-data is maintained by the MICS and DIMIS chairs of the University of Passau, Germany , 2013.
Contact | Terms and Privacy | About

42-Data
Create Date Centric Questions

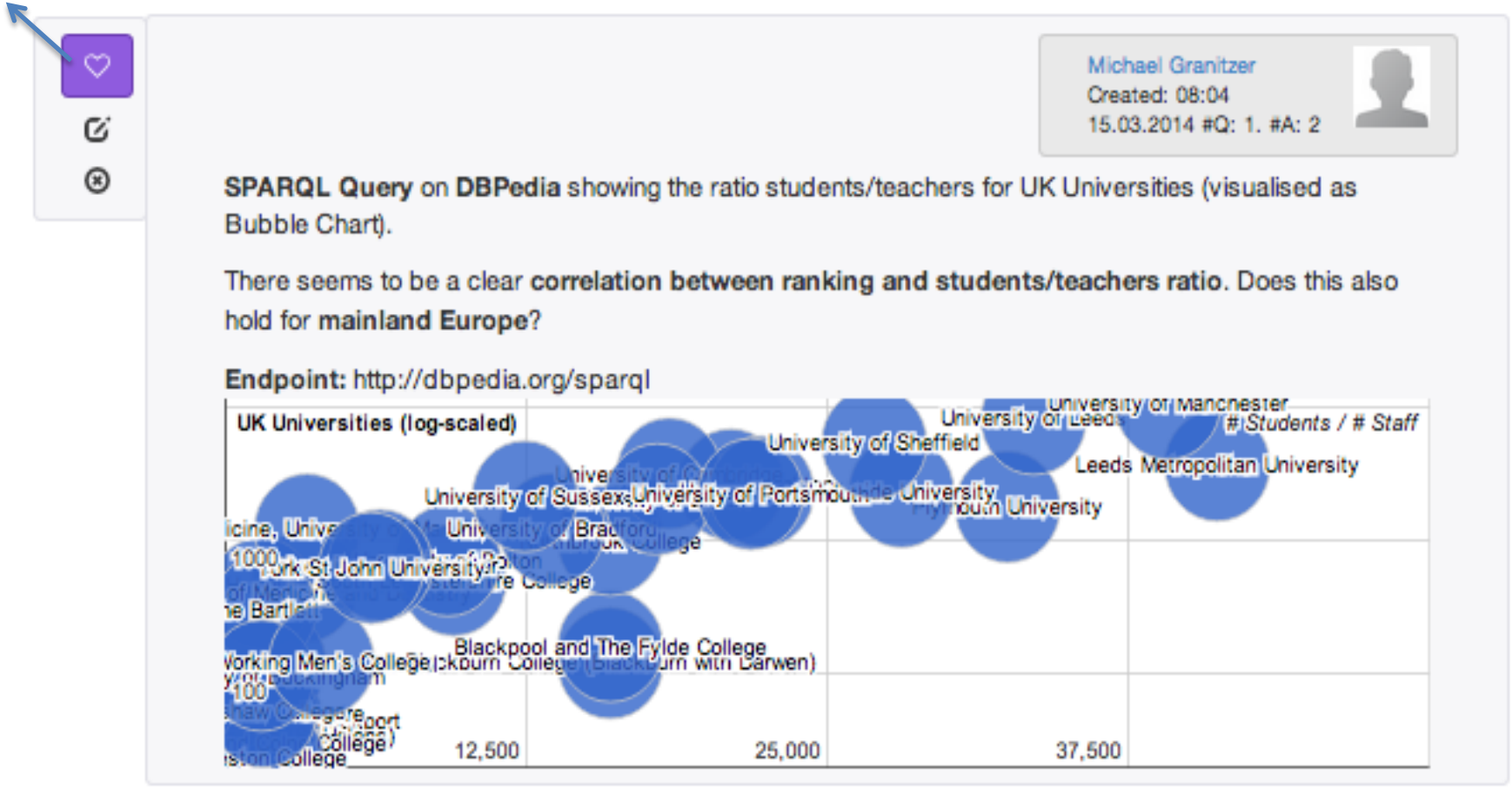




Embedding data/resources
in questions and answers

42-Data
Answering with Data and Insights

Donate



42-Data
Bookmark your resources data

- -

Your questions | Your Anwers | Your Data cubes | Your Resources | Your Publications

Search:

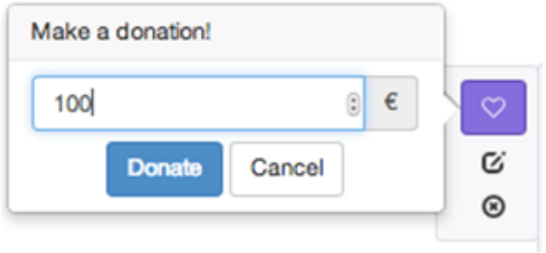
Type	Label	Description
	UK University Ratio	**SPARQL Query** on **LOD Cloud Cache** showing the ratio students/teachers for UK Universities (visualised as Bubble Chart).
	UK University Ratio	**SPARQL Query** on **LOD Cloud Cache** showing the ratio students/teachers for UK Universities (visualised as Bubble Chart).
		SPARQL Query on **DBPedia** showing the ratio students/teachers for UK Universities (visualised as Bubble Chart). There seems to be a clear **correlation between ranking and students/teachers ratio**. Does this also hold for **mainland Europe**?
	UK University Ratio	**SPARQL Query** on **LOD Cloud Cache** showing the ratio students/teachers for UK Universities (visualised as Bubble Chart). There seems to be a clear correlation between ranking and students/teachers ratio. Does this also hold for mainland Europe?
	UK University Student/Staff Ratio	**SPARQL Query** on **LOD Cloud Cache** showing the ratio students/teachers for UK Universities (visualised as Bubble Chart). There seems to be a clear correlation between ranking and students/teachers ratio. Does this also hold for mainland Europe? **Note:** This answer shows the SPARQL Query for creating the resource/answer.
	A higher education for the twentyfirst century presentation on slideshare	
	How is the staff to student ratio in universities in different European countries?	I am particularly interested in * A comparison between UK University and main land Europe * A comparison between Europe and the US Please provide hard numbers.

Showing 1 to 10 of 11 entries

← Previous | 1 | 2 | Next →

42-Data
Economic Sustainability

- Customer-to-customer situation
- Long tail of niche topics and transactions
  - Transactions are low value on average
  - Need a high number of transactions to be sustainable
- Valuable micro-transactions
  - Answers, questions, data sources, data sets, derived insights
- Donations (no legal/contract issues)
- Non-monetary incentives: social reputation models



Summary: Support Data-driven Culture & Processes

- Data-centric discussions
  - Re-use many existing data sets
  - Combine data, visualisations
- Data (relationship) discovery
  - Discovery of data and relationships in the LOD: Query Wizard
  - Upload/manage own (aggregated) datasets

Data-driven Science

- **Research** practices and **processes** which are based on automatic generation, interpretation, and exploitation of large amounts of information and data.

- Four central steps

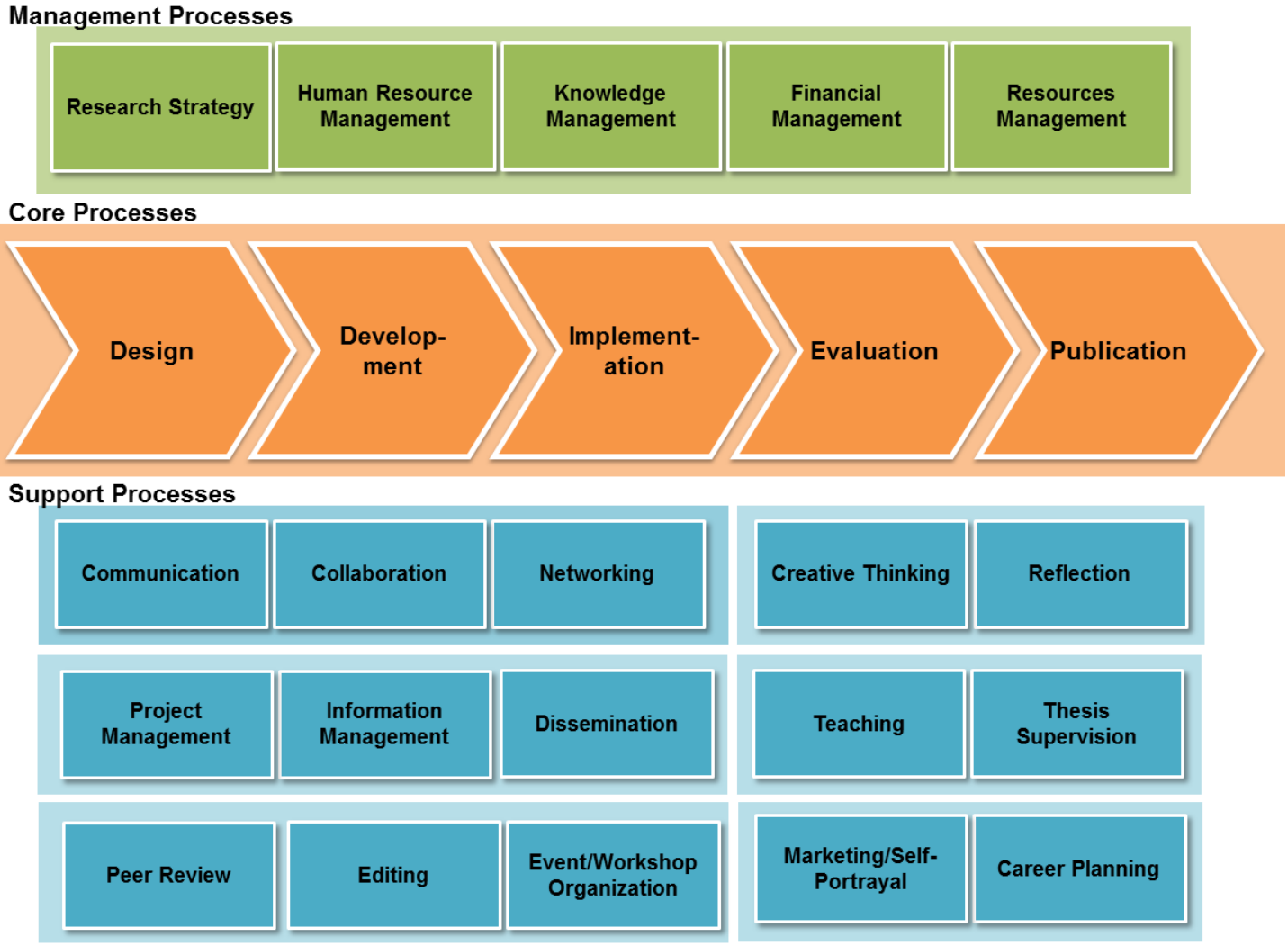
  (1) Provide appropriate data & IT infrastructure

  (2) Democratize data across the community

  (3) Enable experimentation with data

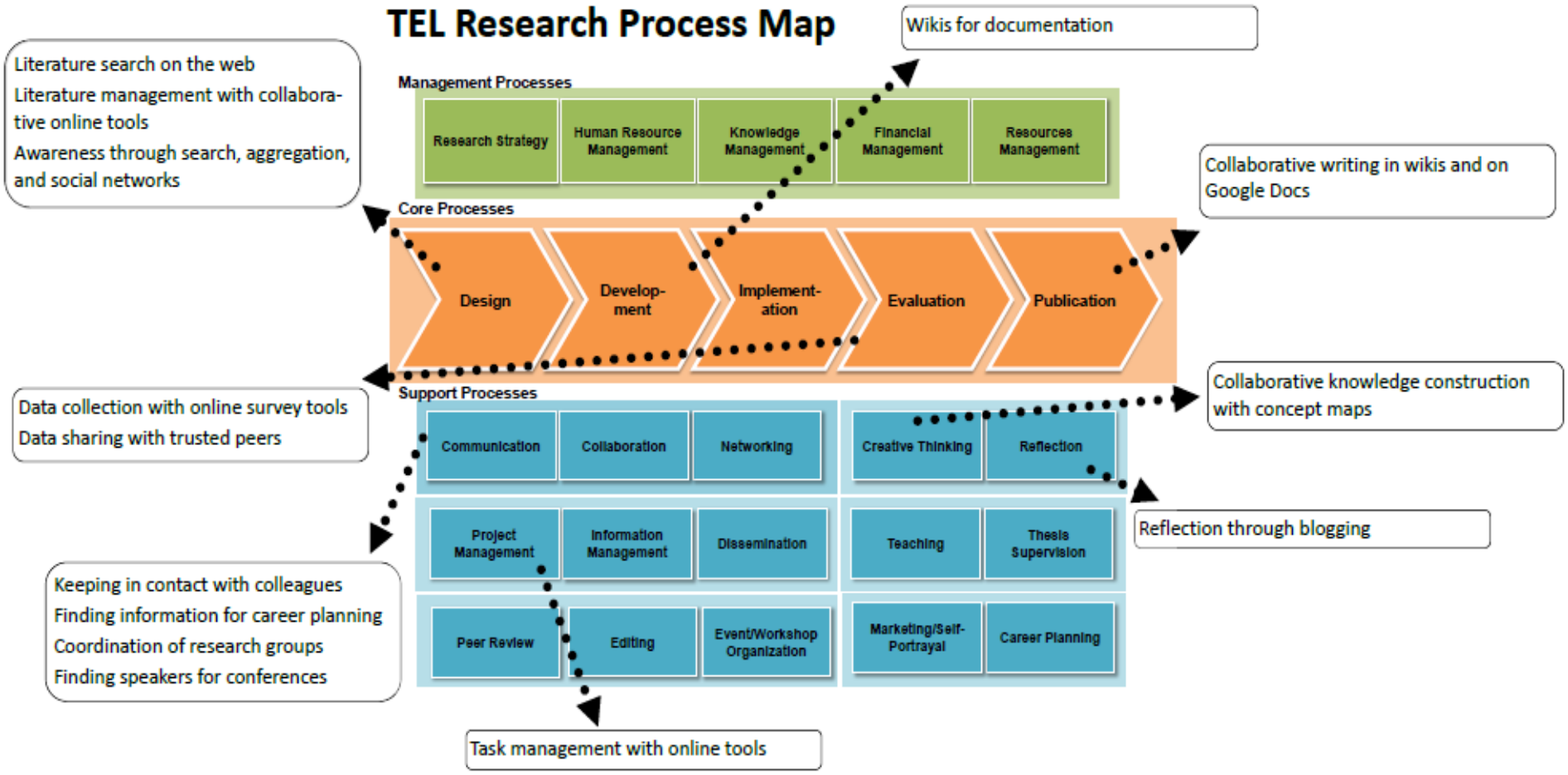
  (4) Support data-driven culture

Research Processes (in TEL)

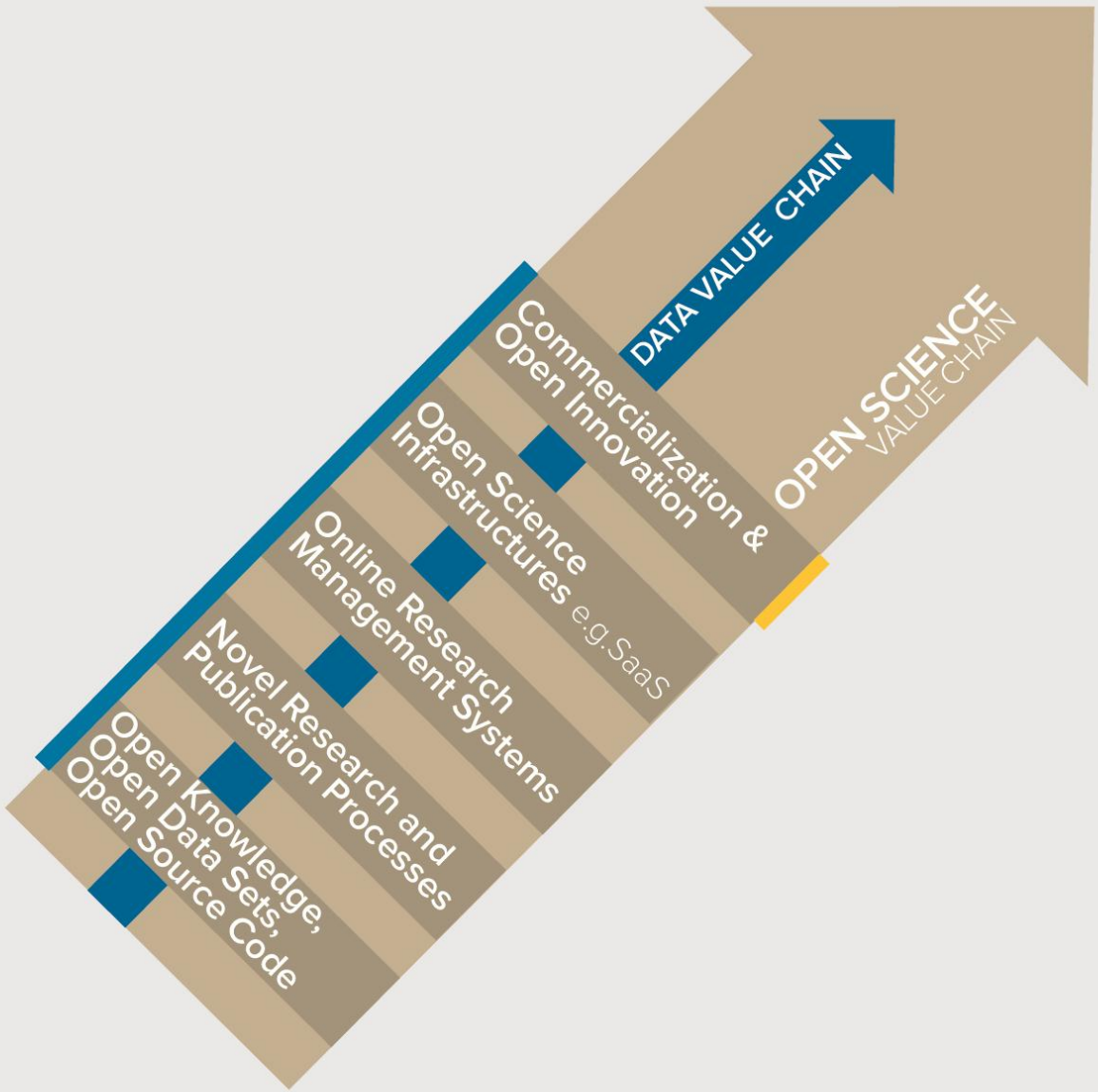


[Kraker & Lindstaedt, 2011]

Research Processes (in TEL)



Open Science Data Value Chain?



KNOW Center

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

SAVE THE DATE: i-KNOW Conference

Special Track on Science 2.0

October 21-23, 2015, Graz, Austria

COMET bmvit bmwfw FFG Das Land Steiermark SFG STEIRISCHE WIRTSCHAFTSFÖRDERUNG

© Know-Center GmbH