Science 2.0 & Big Data
Science 2.0 Conference, Hamburg, March 25, 2015

Prof. Dr. Stefanie Lindstaedt
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
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

Data

Public – Private

Structured – Unstructured

Text

Sensor Data

Social Data

Activity Traces

Linked Data

Uncertainty – Changes

Recommendations – Analyses – Visualizations

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 knowledge

[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

Overview of Educational Technology (what's this?)

Digital Natives

Personal Learning Environment

Online Learning and Technology Adoption

Mobile Learning

Community of Practice

Meta Analysis

Computer-supported Collaborative Learning

Game-based Learning

The Future of Learning

Design-based Research

Cognitive Models

Instructional Design

Technological Pedagogical Content Knowledge

[Kraker, 2013]
Cross-Journal Recommendation based on Click Streams

[Bollen et al., 2009]
Extract facts from research papers

- 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. Objective: crowd-sourced triplification for statistical data

2. 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. 

Displaying 10 of 199 results
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

A Flea Market for Open Data

Manage

Discuss

Discover

Donate

41 resources available
6 questions available
41 resources available
5 users

42-data is maintained by the MICS and DIMIS chairs of the University of Passau, Germany, 2013.
Embedding data/resources in questions and answers
Answering with Data and Insights

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?

Endpoint: http://dbpedia.org/sparql
**42-Data**

**Bookmark your resources data**

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

*European Data Forum 2014*  
*Athens, Greece*  
*March 19th – 20th, 2014*
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)

Management Processes
- Research Strategy
- Human Resource Management
- Knowledge Management
- Financial Management
- Resources Management

Core Processes
- Design
- Development
- Implementation
- Evaluation
- Publication

Support Processes
- Communication
- Collaboration
- Networking
- Creative Thinking
- Reflection
- Project Management
- Information Management
- Dissemination
- Teaching
- Thesis Supervision
- Peer Review
- Editing
- Event/Workshop Organization
- Marketing/Self-Portrayal
- Career Planning

[Kraker & Lindstaedt, 2011]
Research Processes (in TEL)

**TEL Research Process Map**

**Management Processes**
- Research Strategy
- Human Resource Management
- Knowledge Management
- Financial Management
- Resources Management

**Core Processes**
- Design
- Development
- Implementation
- Evaluation
- Publication

**Support Processes**
- Communication
- Collaboration
- Networking
- Creative Thinking
- Reflection
- Project Management
- Information Management
- Dissemination
- Teaching
- Thesis Supervision
- Peer Review
- Editing
- Event/Workshop Organization
- Marketing/Self-Portrayal
- Career Planning

**Tools and Methods**
- Wikis for documentation
- Collaborative writing in wikis and on Google Docs
- Collaborative knowledge construction with concept maps
- Reflection through blogging
- Task management with online tools
- Literature search on the web
- Literature management with collaborative online tools
- Awareness through search, aggregation, and social networks
- Data collection with online survey tools
- Data sharing with trusted peers
- Keeping in contact with colleagues
- Finding information for career planning
- Coordination of research groups
- Finding speakers for conferences
Open Science Data Value Chain?
SAVE THE DATE: i-KNOW Conference

Special Track on Science 2.0
October 21-23, 2015, Graz, Austria