The Value of Big Data
From Data-Driven Enterprises to a Data-driven Economy

Prof. Dr. Stefan Wrobel

Fraunhofer-Institute for Intelligent Analysis and Information Systems IAIS
Fraunhofer Alliance Big Data

www.iais.fraunhofer.de
bigdata.fraunhofer.de

This presentation contains copyrighted material and may not be reproduced without permission © Fraunhofer, 2014
Fraunhofer IAIS: Intelligent Analysis and Information Systems

Do more with data

200 people, at the campus Birlinghoven castle close to Bonn

Research areas
- Data Science and Big Data
- Data Linking, Open Data
- Machine Learning, Data Mining, Multimedia Pattern Recognition, Sensor Analytics
- Visual Analytics
- Big data in business processes
Fraunhofer Alliance Big Data

Joint competences in a »Big Data Factory« for Germany

Strategies, Solutions and Successes

24 Fraunhofer institutes – one central coordination point

Synchronized and broad competence portfolio with many years of expertise in big data in different sectors

Best of class Big Data solutions for individual projects, consulting and qualification of personnel

bigdata.fraunhofer.de

Contact: Prof. Dr. Stefan Wrobel (Chairman)
Die Age of Big Data

GOOD with numbers? Fascinated by data? The sound you hear is opportunity knocking.

Der ungebobene Schatz

Insbesondere große Unternehmen versuchen sich Wettbewerbsvorteile von einer gezielten Analyse ihrer Datenmengen zu verschaffen. Doch auch kleinere Unternehmen profitieren von dem Trend.

Prof. Stefan Wrobel (Computervissenschaftler) Als Direktor des Fraunhofer-Instituts für Intelligente Analyse- und Informationssysteme IAIS beschäftigt er sich mit den technischen Möglichkeiten von Big Data und deren Folgen. Stefan Wrobel warnt vor Kulturpessimismus – die Umwälzungen seien eine große Chance für die Entwicklung der Weltwirtschaft.

Süddeutsche Zeitung

Nach unserer Zukunft

Ein Besuch bei den Fraunhofer-Forschern für „Intelligente Analyse- und Informationssysteme“, Deutschlands avanciertesten Big-Data-Propheten.
Big Data Trends

Convergence

Ubiquitous Intelligent Systems

User Content

Open Data

© Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Big Data Trends

40 Zettabyte by 2020 [IDC 2012]
Big Data

A definition attempt

Big Data in general refers to

- The trend towards availability of ever more detail than ever closer to real-time data
- The switch from a model-driven to a model- and data-driven approach
- The economic potentials that result from the analysis and use of big data when properly integrated into company processes

Big Data currently focuses technically on the following aspects

- Volume, Variety, Velocity
- In-memory computing, Hadoop etc.
- Real-time analysis and effects of scale

Big Data must take implications to society into account
Big Data

The view of BITKOM, The German IT Association

Volume
- Number of records and files
  - Yottabytes
  - Zettabytes
  - Exabytes
  - Petabytes
  - Terabytes

Velocity
- High speed data generation
  - Constant transmission of generated data in realtime
    - Milliseconds
    - Seconds | minutes | hours

Variety
- External data (web open data, etc.)
- Company data
  - Unstructured, semistructured, structured data
  - Presentations | text | video | images | tweets | blogs
  - Machine to machine communication

Analytics
- Discovery of relationships, patterns, meaning
  - Prediction models
- Data Mining
- Text Mining
- Image Analytics | Visualization | Realtime

Source: BITKOM Big Data Leitfaden, 2012. BITKOM AK Big Data
Innovation study Big Data

Desk research (current state of affairs)
- Detailed overview of the national and international Big Data landscape
- More than 50 systematic Big Data Business Cases

In-depth workshops for industry sectors (qualitative study)
- Expert workshops
- Finance, Telecom, Market research, E-Comm., Insurance

Online survey (quantitative study)
- 1.10.2012 to 30.11.2012
- 82 high-ranking executives from small and large companies
Big Data Competitive Edge

Percentage believing that business analytics creates a competitive advantage in their organization

<table>
<thead>
<tr>
<th>Year</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>37%</td>
<td>58%</td>
<td>67%</td>
</tr>
</tbody>
</table>

Sources: © MIT Sloan Management Review 2012, 400 companies
Big Data Uptake Worldwide

U.S. Ahead, Europe Coming

Percentage of Companies (by Region) With Big Data Initiatives in 2012

- Total: 53% with initiatives, 47% without
- U.S.: 68% with initiatives, 32% without
- Europe: 45% with initiatives, 55% without
- Asia-Pacific: 39% with initiatives, 61% without
- Latin America: 51% with initiatives, 49% without

[Sources/©: TCS 2013 Trend Study Big Data, 643 companies]
Big Data ROI

Very positive ROIs across all regions

Mean Percentage of Expected Return in 2012 on Big Data Investments by Region

- Asia-Pacific: 70.6%
- Latin America: 62.2%
- Average: 45.5%
- Europe: 42.5%
- North America: 36.5%

Percentage ROI = \( \frac{\text{Gain from investment} - \text{Cost of investment}}{\text{Cost of investment}} \) * 100

[Sources/©: TCS 2013 Trend Study Big Data, 643 companies]
Big Data Efforts Will Increase

Comparison of Actual Volume 2012 with Projected Volume 2015

[Sources/©: TCS 2013 Trend Study Big Data, 643 companies]
Use Of Big Data in Company Functions

From controlling to research

- Controlling: 24%
- Marketing: 19%
- Sales: 18%
- IT: 18%
- Production: 17%
- Research and development: 14%
- Supply Chain: 7%
- Other: 6%

[Source/© BARC Big Data Survey Europe 2013, 274 Europ. Companies]
Published Success Stories Across all Sectors

Prof. Dr. Stefan Wrobel

© Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Fraunhofer alliance projects across all sectors

Highlights

- Business & Finance
- Energy & Environment
- Life Sciences & Health Care
- Logistics & Mobility
- Production & Industry 4.0
- Security
Big Data Comprehensive Strategies Rare

Only few companies have comprehensive strategies

- No: 63%
- Planned: 23%
- Yes: 14%

[BARC Big Data Survey Europe ©2013, 274 Companies]

Ist Big Data für Ihr Unternehmen relevant?

- Ja: 71%
- Nein: 29%

- 9% Wir setzen Big Data bereits ein
- 31% Wir planen den Einsatz von Big Data konkret
- 28% Wir überlegen noch
- 33% Wir haben uns noch nicht damit beschäftigt

[BITKOM Big Data Survey Germany ©2014, 507 Companies]
Barriers to Big Data in Companies

- Too few big data specialists: 70%
- Technical IT Security Requirements: 61%
- Insufficient Budget: 59%
- Privacy Regulations are a barrier: 48%
- Big Data Tools/Solutions are not sufficiently mature yet: 42%
- I know of too few supplier of big data solutions: 35%
- We don't have enough data: 22%
- I don't know of sufficiently many usage areas: 22%
- Our data are of insufficient quality: 9%

[BITKOM Big Data Survey Germany ©2014, 507 Companies, transl. SW]
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways

Technical solutions must be designed-to-fit

Further innovation needs beyond off-the-shelf software

Data Linking and brokering need open standards

Security and Privacy are demanded by business and society alike – „by design“

Enormous education and training needs

SMEs and startups face special challenges and need a supportive ecosystem
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways

Technical solutions must be designed-to-fit

Further innovation needs beyond off-the-shelf software

Data Linking and brokering need open standards

Security and Privacy are demanded by business and society alike – “by design“

Enormous education and training needs

SMEs and startups face special challenges and need a supportive ecosystem

Prof. Dr. Stefan Wrobel
Big Data and the Digital Company

Intensity and Leadership!

- Revenue
  - Revenue/Employee
  - Fixed Assets Turnover

- Profitability
  - EBIT Margin
  - Net Profit Margin

- Market Valuation
  - Tobin’s Q Ratio
  - Price/Book Ratio

[MIT Sloan Management Review ©2012]
OE is a complex interplay of multiple factors

- Who do we want to be?
- What are we offering?
- How are we organized?
- What is our style of working?
- What is our common understanding?
- How do we optimally use our means?
- ... 

Whenever too few factors are being considered, a lot of potential is lost
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways.

Technical solutions must be designed-to-fit

Further innovation needs beyond off-the-shelf software

Data Linking and brokering need open standards

Security and Privacy are demanded by business and society alike – „by design“

Enormous education and training needs

SMEs and startups face special challenges and need a supportive ecosystem
Suppliers and technologies in the context of Big Data (Selection)
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways.

Technical solutions must be designed-to-fit.

Further innovation needs beyond off-the-shelf software.

Data Linking and brokering need open standards.

Security and Privacy are demanded by business and society alike – „by design“.

Enormous education and training needs.

SMEs and startups face special challenges and need a supportive ecosystem.
Multimedia dominates
Types of Data in Companies

Structured, Semistructured, Unstructured

Mean Estimated Percentage of Structured, Unstructured and Semi-Structured Data, Across all of the Company’s Big Data Initiatives

- North America: 55% Structured, 27% Unstructured, 19% Semi-Structured
- Latin America: 53% Structured, 29% Unstructured, 18% Semi-Structured
- Europe: 50% Structured, 25% Unstructured, 25% Semi-Structured
- Asia-Pacific: 45% Structured, 34% Unstructured, 21% Semi-Structured

[TCS ©2013 Trend Study Big Data, 643 companies]
The data iceberg

- Database tables
- Excel spreadsheets
- Other data with fixed structure

- Email, Notes
- Word documents
- PDF, Power Point
- Other text
- Images
- Video, audio

20% vs. 80%
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways
Technical solutions must be designed-to-fit
Further innovation needs beyond off-the-shelf software

Data Linking and brokering need open standards

Security and Privacy are demanded by business and society alike – „by design“

Enormous education and training needs

SMEs and startups face special challenges and need a supportive ecosystem
Data sources and origin in companies

Internal vs External

Mean Estimated Percentage of Data that Comes from Internal or External Sources, Across All of the Company’s Big Data Initiatives

North America: 71% Internal, 29% External
Latin America: 70% Internal, 30% External
Europe: 68% Internal, 32% External
Asia-Pacific: 62% Internal, 38% External

[TCS ©2013 Trend Study Big Data, 643 companies]
The Linked Open Data Universe

Est. 50 billion facts 2013
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways

Technical solutions must be designed-to-fit

Further innovation needs beyond off-the-shelf software

Data Linking and brokering need open standards

Security and Privacy are demanded by business and society alike – „by design“

Enormous education and training needs

SMEs and startups face special challenges and need a supportive ecosystem
## Case study: privacy attitudes in Germany

<table>
<thead>
<tr>
<th>Percent</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>62</td>
<td>Percent want better privacy protection</td>
</tr>
<tr>
<td>87</td>
<td>Percent believe that companies are using data beyond what is publicly announced</td>
</tr>
<tr>
<td>95</td>
<td>Percent pay attention to whom they give their data</td>
</tr>
<tr>
<td>50</td>
<td>Percent read the general terms and conditions and privacy rules</td>
</tr>
<tr>
<td>10</td>
<td>Percent ready to give their data in social web for coupons</td>
</tr>
<tr>
<td>10</td>
<td>Percent would give data for personalized recommendations</td>
</tr>
<tr>
<td>75</td>
<td>Percent would provide them for medical good</td>
</tr>
<tr>
<td>80</td>
<td>Percent sold their data for 5 € in a lab setting</td>
</tr>
</tbody>
</table>

[Handelsblatt Research Institute 2013]
Roadblocks seen in survey

- Companies see the main challenges in the following areas:
  - Privacy and security (49%)
  - Budgets and priorities (45%)
  - Technical challenges of data management and analytics (38%)
  - Expertise (36%)
  - Lack of familiarity with big data technologies (35%).

- To address these issues, 95% of companies requested
  - Best Practices, Trainings, Supplier and solution catalogues and better privacy lawas and regulation
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways.

Technical solutions must be designed-to-fit.

Further innovation needs beyond off-the-shelf software.

Data Linking and brokering need open standards.

Security and Privacy are demanded by business and society alike – „by design“.

Enormous education and training needs.

SMEs and startups face special challenges and need a supportive ecosystem.

Prof. Dr. Stefan Wrobel
Data Scientists

From data and analytics to business

Prof. Dr. Stefan Wrobel

© Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Big Data – Challenges towards Data Value

From data-driven companies to a data-driven economy

Big Data is not an isolated IT topic, but must address business value end-to-end in company/sector specific ways.

Technical solutions must be designed-to-fit.

Further innovation needs beyond off-the-shelf software.

Data Linking and brokering need open standards.

Security and Privacy are demanded by business and society alike – „by design“.

Enormous education and training needs.

SMEs and startups face special challenges and need a supportive ecosystem.

Prof. Dr. Stefan Wrobel
Example National Health Service

The benefits of a working open data ecosystem

NHS provides enormous datasets:

- Hospital Episode Statistics (HES) repository: 100 million records per year (outpatient appointments, A&E attendances, hospital admissions)

- Prescription data: 500 Million records per year, increasingly openly available

Success story:

- Analysis of Statin prescriptions 2011-12 (37 Million records) by a team of Mastodon C, Open Health Care UK and BadScience.net

- Annual savings of more than 200 Mio Pounds identified (equally effective medications)

[Guardian, theodi.org, prescribinganalytics.com, wikimedia, 2013]
Returning the value of data

An interesting experiment in the U.S.

- 8 $/month
- 1500 beta users

SELL YOUR DATA

Let go
Get paid every month. You'll get paid like clockwork.

You decide
Decide who buys your data. To sell or to not sell, it's up to you.

Transparency
Sell only what you want. Pick and choose what data you want to sell.

Learn more
Beautiful data visualizations. See your data like never before.

Add
Choose multiple data sets. From social media to your debit or credit card, it's up to you.

Secure
Bank level security. 256 Bit AES Encryption

[www.datacoup.com, March ©2014]
Data as a Product

Is it worth selling?

Q10: Percent of Companies by Industry that Sold their Digital Data in 2012

- Telecommunications: 50% Yes, 46% No, 4% Don't know
- Utilities: 38% Yes, 52% No, 10% Don't know
- High Tech: 34% Yes, 61% No, 5% Don't know
- Media and Entertainment: 32% Yes, 58% No, 11% Don't know
- Travel/Hospitality/Airlines: 32% Yes, 63% No, 5% Don't know
- Banking/Financial Services: 30% Yes, 65% No, 5% Don't know
- Total: 27% Yes, 67% No, 6% Don't know
- Insurance: 23% Yes, 71% No, 6% Don't know
- Retail: 21% Yes, 73% No, 6% Don't know
- Consumer Goods: 20% Yes, 71% No, 9% Don't know
- Life Sciences: 18% Yes, 79% No, 3% Don't know
- Manufacturing: 16% Yes, 78% No, 7% Don't know
- Energy & Resources: 14% Yes, 77% No, 9% Don't know

Q13a: Mean Annual Revenue Per Company by Industry in 2012 from Selling Digital Data

- Insurance: $40 million
- Telecommunications: $29 million
- High Tech: $26 million
- Life Sciences: $22 million
- Total: $22 million
- Utilities: $21 million
- Media and Entertainment: $19 million
- Banking/Financial Services: $19 million
- Manufacturing: $17 million
- Retail: $13 million
- Consumer Goods: $9 million
- Travel/Hospitality/Airlines: $7 million
- Energy & Resources: $2 million

[TCS ©2013 Trend Study Big Data, 643 companies]
Dimensions of the big data ecosystem

Legal Dimension
- Ownership
- Copyright/IPR
- Liability
- Insolvency
- Privacy

Skill Dimension
- Best Practices
- Frameworks
- Tools

Application Dimension
- Data-driven Decision Making
- Risk Management
- Competitive Intelligence
- Digital Humanities
- Verticals
- Industry 4.0

Technology Dimension
- Scalable Data Processing
- Signal Processing
- Statistics/ML
- Linguistics
- HCI/Visualization

Social Dimension
- User Behaviour
- Societal Impact
- Collaboration

Economic Dimension
- Business Models
- Benchmarking
- Open Source
- Deployment Models
- Information Pricing

Big Data Ecosystem

[Cavanillas, Markl, May, Platte, Urban, Wahlster, Wrobel – Big Data Value (Draft), 2014]
Smart Data Innovation Lab

A nationwide industry-research platform for big data value
Creating a European Big Data Ecosystem

A Partnership of multiple stakeholders will be needed

- Big Data Vendors
- Policy Makers
- Public and corporate users
- SMEs and startups
- Researchers

Establish legal and policy framework

Advance technology

Create innovation and growth

Prof. Dr. Stefan Wrobel
© Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
Conclusion

- Big Data is here to stay: significant uptake in companies
- Enormous potential and growth expected
- Significant barriers exist: the big 7 challenges
  - Business value, designed-to-fit, innovation, data linking, privacy, education, SMEs
- Coordinated action by multiple stakeholders at European level needed

[Guardian, theodi.org, prescribinganalytics.com, wikimedia, 2013]