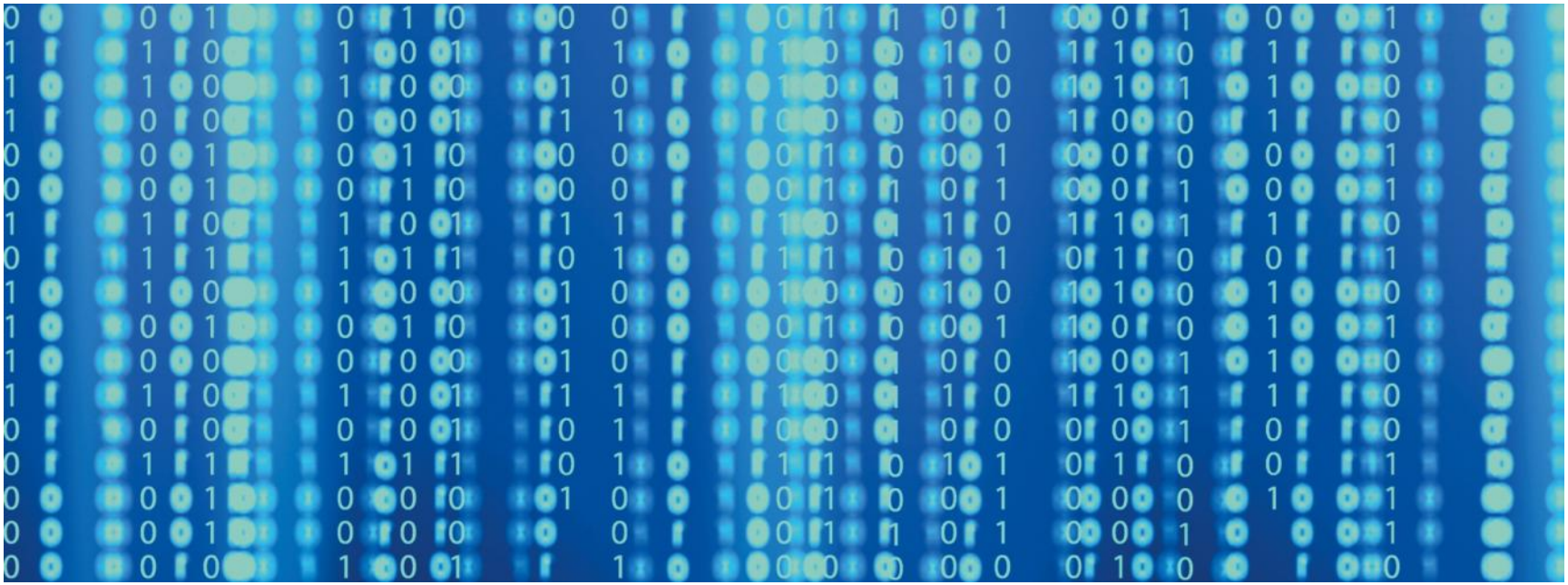
ASSESSING BIG DATA
EXPERIENCES FROM GERMANY

2nd PACITA Conference, Berlin 25th of February 2015
Section: Governance of Big Data

Timo Leimbach/Daniel Bachlechner



AARHUS UNIVERSITY

Fraunhofer ISI

Overview

- Introduction
- Methodological aspects
- Overview on study results and outcomes
- Challenges related to the study
- Reflections on assessing big data
- Impacts on the governance of big data

Introduction

Context of the study

- Title “Big data in the cloud”
- Project emerged from initial wishes by the parliament related to cloud in 2009
- Inclusion of big data for different reasons
- Commissioned to Fraunhofer ISI
- Big data as major focus, but with special attention to cloud based offers
- Timing: Summer 2013 – Spring 2014

Goals of the study

- “Feasibility” study
- Overview on the topics
- Exploration of challenges and opportunities
- Identification of research questions and needs

Methodological aspects

Used sources and methods

- Based on literature review
  - vast number of publications of all kinds
  - explosive growth during the project
  - amount required assessment and selection
- Interviews
  - expert interviews in autumn 2013
  - semi-structured
- Mini scenarios
  - five scenarios along the “Bedarfsfelder” of the German High Tech Strategy
  - departure in existing or envisioned usages of big data
  - aimed at “shaping” big data

Overview on study results and outcomes
Potentials

- Identification and assessment of potentials and challenges
  - based on an introduction to the technology and markets
  - two-stepped approach to relate potentials and challenges

- Assessing potentials
  - potentials split up into two groups
    - general socio-economical potentials (for organisations, science, consumer/citizen, society/ economy as a whole)
    - application potentials in climate/energy, food/health, transport, security, communication (based on mini scenarios)
  - Idea behind application potentials
    - concretize abstract potentials
    - relate potentials and challenges

Overview on study results and outcomes
Potentials

- Some examples for potentials
  - general potentials
    - productivity for organizations → OECD study shows 5% more productivity
    - new knowledge/insights in science (f.e. on societal dynamics, diseases etc.)
    - life quality or participation for citizens (f.e. citizen scientist etc.)
    - employment and growth for society/economy as a whole
  - application scenarios
    - Climate/Energy - advances in climate modeling, energy savings through prediction
    - Food/Health – insights in diseases, improved pharmaceutics
    - Transport – savings through smart transport
    - Security – predictive policing for security
    - Communication – fraud detection or predictive maintenance in advanced manufacturing
  - scenarios should go beyond “customer success story brochure”

Overview on study results and outcomes

- Assessing challenges
  - challenges divided into five sections
    - oriented along general potentials
    - includes technological, business, legal, societal, and structural challenges

- Some examples for challenges
  - technological challenges, f. e.
    - interoperability – data formats, transfer of data
    - data management – big data require new approaches and technologies
  - legal challenges, f. e.
    - IPR – who owns innovation based on multiple data sources
    - data protection – pseudonyms and its problems
    - liability – who is responsible in case of errors

Overview on study results and outcomes

Challenges and needs

- business challenges, f. e.
  - profitability – not all use cases are business cases
  - knowledge – what is needed, where can it be found, how it can be analyzed
- societal challenges
  - trust and reliability – uncertainty of big data
  - “filter problems” – problem of data existence and selection/bias
  - autonomy and manipulation – incentives and control
  - discrimination and participation – exclusion through data
- structural challenges, f. e.
  - Infrastructure – infrastructure requirements for big data (transfers)
  - human capital – data literacy

- Based on this analyses six areas with needs for research and/or action were identified
  - areas consist of three to six needs
  - partly interrelated

Overview on study results and outcomes

Needs and further outcomes

- Needs for research or action and examples
  - use and business cases, incl. inclusion of consumer etc.
  - technological infrastructure, incl. research on new approaches
  - economics of data, incl. research on the value of data
  - competencies and human capital, incl. need for information/data literacy
  - legal certainty, incl. f.e. clarification of IPR issues
  - security and data protection, incl. f.e. security research, data protection regulation

- Further outcomes of the study
  - two in-depth studies
    - spatial/location data
    - patient records
  - low reaction on side of the parliament

Challenges related to the study

- Definition/scoping of big data
  - definitions are often vague
  - moving target driven by hype
- Concreteness
  - many big data applications exist only as visions or re-labeling, only a few real big data applications
  - most often very restrictive access to details
- External impacts – the NSA story
  - big data as a term became problematic → evasion into new terms
  - balance of arguments was challenged through heated debates and statements → focus on few aspects
  - interest for aspects beyond the NSA et al. disappeared in public and politics

Reflections on assessing big data

Can we assess big data, and if yes, how?

- Yes we can, because
  - many challenges related to big data are not new for TA
    - uncertain technological developments
    - ambiguity of general purpose technologies
    - heated, unbalanced debating environment
  - broad set of methodologies can be used
    - concrete examples and generalizations
    - anticipative methods
- Yes we can, but
  - intangibility of software and data
    - require “concretization”
    - does not mean that there are no real world impacts
  - “big data as a whole” is more a phenomenon instead of technology → impacts reach far beyond (i.e. human autonomy etc.)
  - tension between both aspects → inherent risks of missing perspectives

Implications for the governance of big data

Some considerations and remarks

- Governing the intangibility
  - value of concretization, but also the need for understanding the foundations
    - concrete examples help, but does not cover all aspects of general purpose technology
    - algorithms are hard to explain → lack of interests in details
  - Globalization
    - data travels light and fast
    - data can be easily copied

- Governance of fast-emerging technologies
  - regulation as an answer?
    - balance between too early or too late/ too much or too less
    - matching of regulation to intangible technologies
  - normative approaches as an answer
    - acceptance in the industry as a challenge
    - global “outreach”

Implications for the governance of big data

Some considerations and remarks

- Addressing the fundamental aspects
  - big data raises issues which concern fundamental aspects like human autonomy
    - others do as well, but here it sneaks into the real life (others even rely on big data)
    - need for ongoing critical discussion and awareness
    - awareness as a challenge and a problem
  - interesting/engaging citizens, but also politicians
    - how to interest? How to engage? → current research very ambiguous
    - keeping the interest/engagement alive
  - facilitate a reflective discussion
    - explain the need for an ongoing discussion and reflection
    - ways to keep it relevant (beyond the science section)?

Thank you for the attention

Contact

Timo Leimbach
Aarhus University
DAC – Information studies
Helsingforsgade 14
8200 Aarhus N

Timo.Leimbach@dac.au.dk

Daniel Bachlechner
Fraunhofer ISI
CC Emerging Technologies
Breslauer Str. 48
76139 Karlsruhe

Daniel.Bachlechner@isi.fraunhofer.de