Big Data Software Engineering

Andreas Metzger

PICASSO Expert Group Meeting – Washington, D.C. – May 2016

paluno
The Ruhr Institute for Software Technology

UNIVERSITÄT
DUISBURG
ESSEN

1. About me…

- Senior Academic Councillor (~ tenured assistant prof.), paluno (Ruhr Institute for **Software Technology**), University of Duisburg-Essen
- PhD in Computer Science (**Software Engineering**), University of Kaiserslautern
- Involved in 6 EU and 7 German research projects in **Software Engineering, Services, Cloud**
- Vice Chair of European Technology Platform NESSI: **Software, Services, Cloud & Data**
- Secretary General of European **Big Data** Value Association (BDVA)
- Initiator and co-organizer of ICSE Int'l Workshop on **Big Data Software Engineering**

2. Big Data for Software Engineering

Big Data for software analytics

- ***Data sources:*** forums, forges, blogs, Q&A sites, social networks, …

→ *Mining user needs and experiences to identify **new requirements** and improve customer understanding*

→ *Analyse operational software execution for **failure** patterns, **fault** localization and **root causes***

→ *…*

2. Big Data for Software Engineering

Big Data for self-adaptive software systems

$$M, W \models R$$

Adaptive software systems
modify themselves at run time

- *M* = machine (software) → self-healing / corrective
- *W* = world (context) → context-aware
- *R* = requirements → enhansive

- ***Data sources:*** services execution, things (sensors), cloud infrastructures, users (social networks), … = high volume, velocity data about system context and users

→ *Big Data analytics to better and faster detect and predict* ***changes in R, W, M***

2. Software Engineering for Big Data

Big data engineering

- **Engineering methodologies** for Big Data systems (combination of techniques and tools that achieve goals for known constraints/situations)
- Systematically integrate diverse, **multi-disciplinary** set of aspects for system development, quality assurance and operations

→ *How to accommodate **data scientists as new stakeholders** during software engineering?*

→ *How to integrate concepts, techniques and tools **along software life-cycle**?*

→ *Role of **requirements engineering** if big data insights not known in advance?*

2. Software Engineering for Big Data

Quality assurance for big data software

(≠ data quality!)

UNDETECTED HOLE IN THE OZONE LAYER

The hole in the ozone layer over Antarctica remained undetected for a long period of time because the data analysis software used by NASA in its project to map the ozone layer had been designed to ignore values that deviated greatly from expected measurements.

Bugs especially critical if decisions / software behaviour driven by data!

→ *How to **test** data-intensive systems?*

- Volume: how to generate sufficient / representative test cases (if at all)?
- Velocity: how to monitor and assure quality at run time?
- Veracity: how to check for quality in the presence of “uncertainty” of data quality?

3. Research Landscape in EU

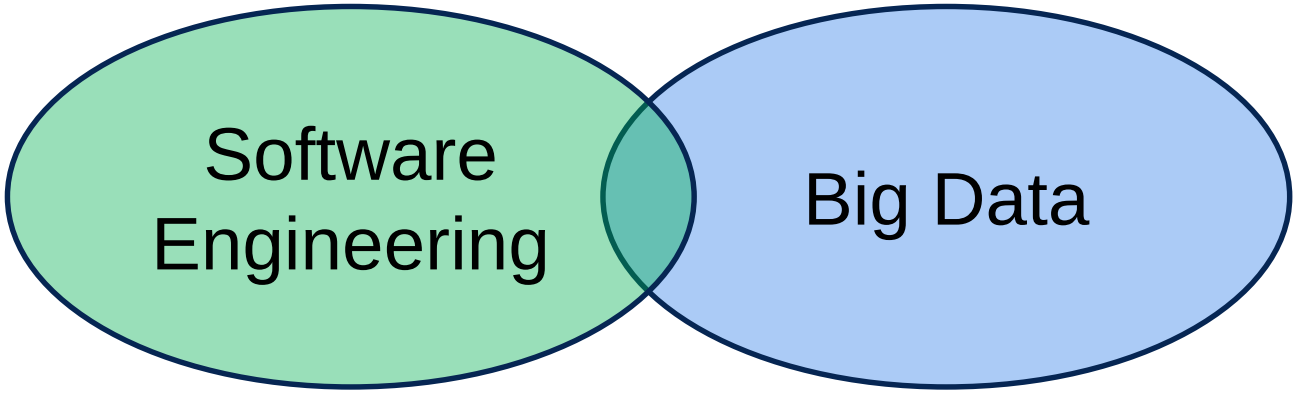
- Research activities on the intersection between Big Data and Software Engineering **(slowly) start gaining momentum**
  - Programme: # **projects**
  - FP7 – ICT: **4** (out of 27)
  - H2020 – ICT: **6** (out of 16)
  - *((25 @ NSF data base))*

- **Growing research interest in EU** on Big Data and Software Engineering
  - Workshop: # **papers**
  - BIGDSE 15: **1** (out of 7 = US)
  - BIGDSE 16: **4** (out of 9 = US)

Workshop on Big Data Software Engineering
@ Int'l Conference on Software Engineering

- **(current) EU funding opportunities**
  - **Cloud and Software Engineering Unit**: increasingly appreciated
  - **Data Value Chain Unit**: not yet appreciated (e.g., software engineering as interesting application domain for big data, beyond e.g., transport & logistics)

4. Opportunity for Joint Research Funding



Research

- Total publications 7,200 **20** 3,300
  (last 5 years; source: Microsoft Academic)

- **Highly Multi-disciplinary**

Impact

- Key **commercial Big Data players** from EU / US
- Complementary „constraints“ and „backgrounds“: e.g., **privacy**, **EU Digital Single Market**, …

Thanks!

- A. Metzger (Ed.), "**Software engineering: Key enabler for innovation**," White Paper, European Technology Platform for Software, Services, Cloud and Data (NESSI), July, 2014
- S. Zillner, E. Curry, A. Berre, A. Metzger, C. Upstill (Eds.), "**Big data value strategic research and innovation agenda** (SRIA)," Version 2.0, European Big Data Value Association (BDVA), February, 2016.

- L. Baresi, T. Menzies, A. Metzger, and T. Zimmermann, "1st International Workshop on Big Data Software Engineering (**BIGDSE 2015**)," in Proceedings 37th International Conference on Software Engineering, ICSE 2015, May 23, 2015, Firenze, Italy
- L. Baresi, T. Menzies, A. Metzger, and T. Zimmermann, "2nd International Workshop on Big Data Software Engineering (**BIGDSE 2016**)," in Proceedings 38th International Conference on Software Engineering, ICSE 2016, May 16, 2016, Austin, Texas, USA