

Big Data Research and Innovation

A Gap Analysis from the Perspective of the European Big Data Value PPP

Andreas Metzger

(paluno @ U Duisburg-Essen, BDVA)

PICASSO 2017, Minneapolis

Background: Big Data Value PPP and BDVA

BDV PPP = Public Private Partnership

- Public side = European Commission – EC
- Private side = BDVA
- Develop new **Big Data technologies, products and services** to give European industry leading position
- **Leverage public funding** with additional investments
- Based on **roadmaps for research and innovation**
- Implemented through **calls for actions** under H2020



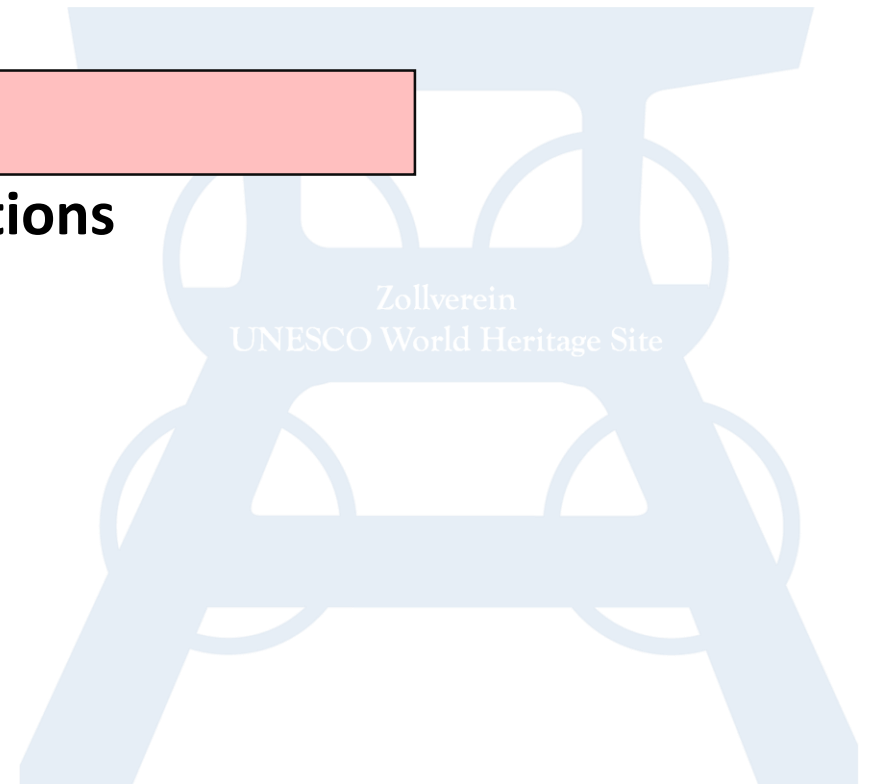
BDVA = Big Data Value Association

- **169 members** from 28 different countries
 - 55% industry/other
 - 45% academia/research



Agenda

- **BDVA Research Priorities**
- **Gap Analysis of ongoing R&I actions**
- **Conclusion**



BDV PPP Roadmap for research and innovation

BDVA Strategic Research and Innovation Agenda (SRIA)



Technical Priorities:

- Data Management
- Data Processing Architectures
- Data Analytics
- Data Visualisation and User Interaction
- Data Protection

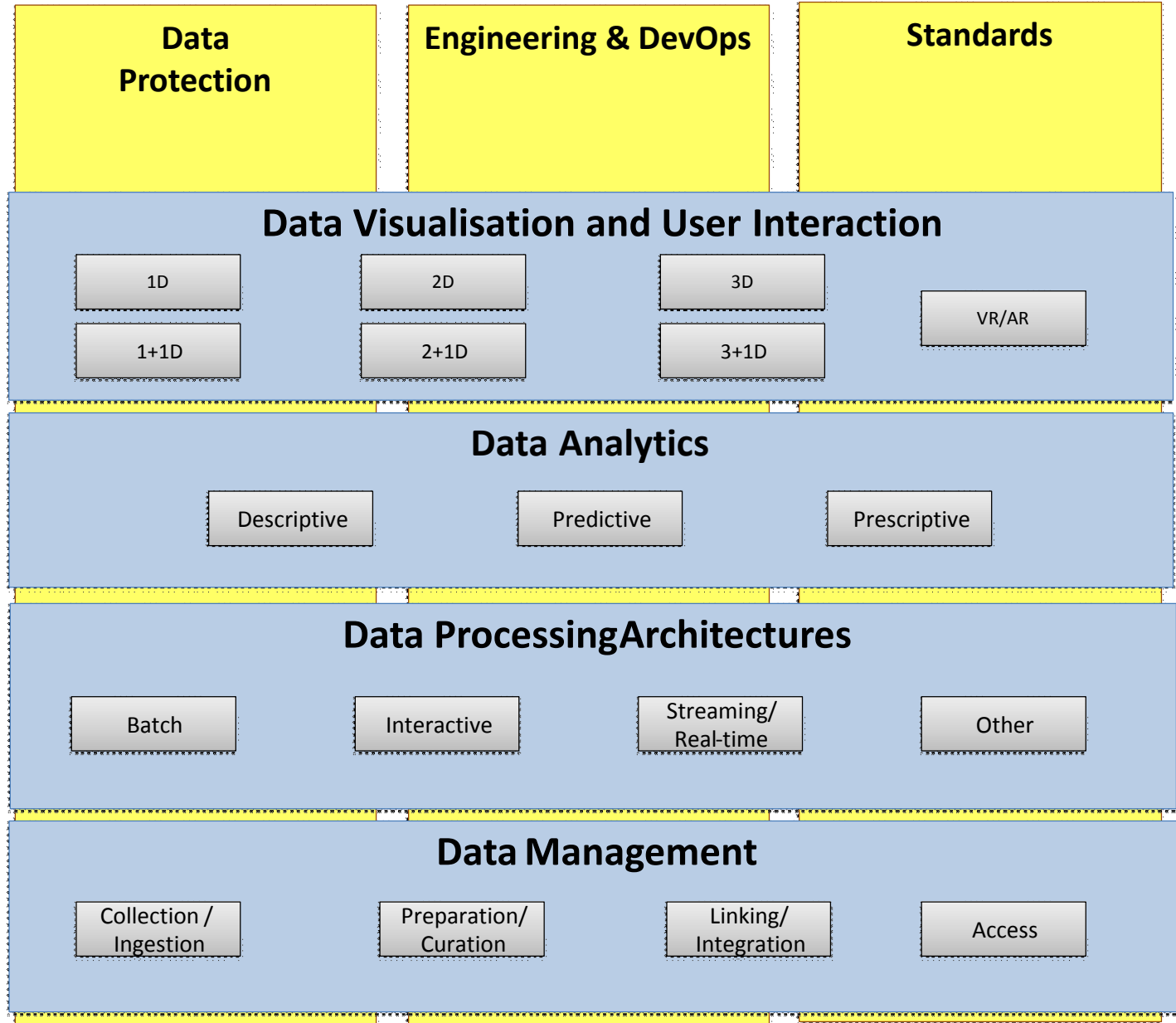
Non-technical Priorities:

- Skills development
- Ecosystems and Business Models
- Policy, Regulation and Standardisation
- Social perceptions and societal implication

Available from
<http://www.bdva.eu/>

BDV Reference Model

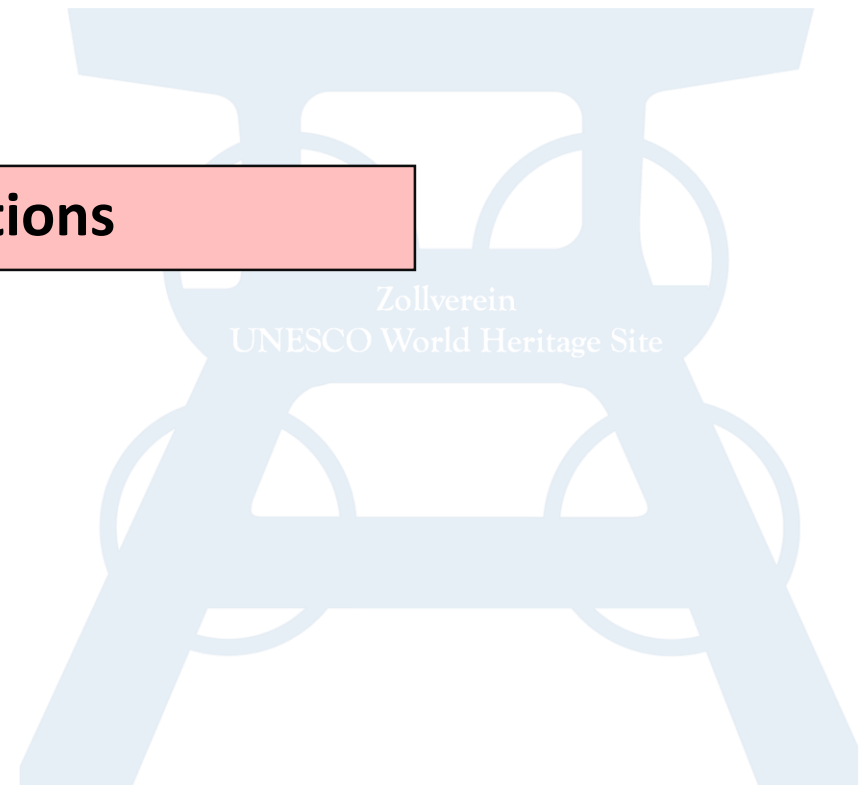
Aim:
“Structuring”
the Priorities



Note:
*Horizontals do
not imply
layered architecture*

Agenda

- BDVA Research Priorities
- Gap Analysis of ongoing R&I actions
- Conclusion

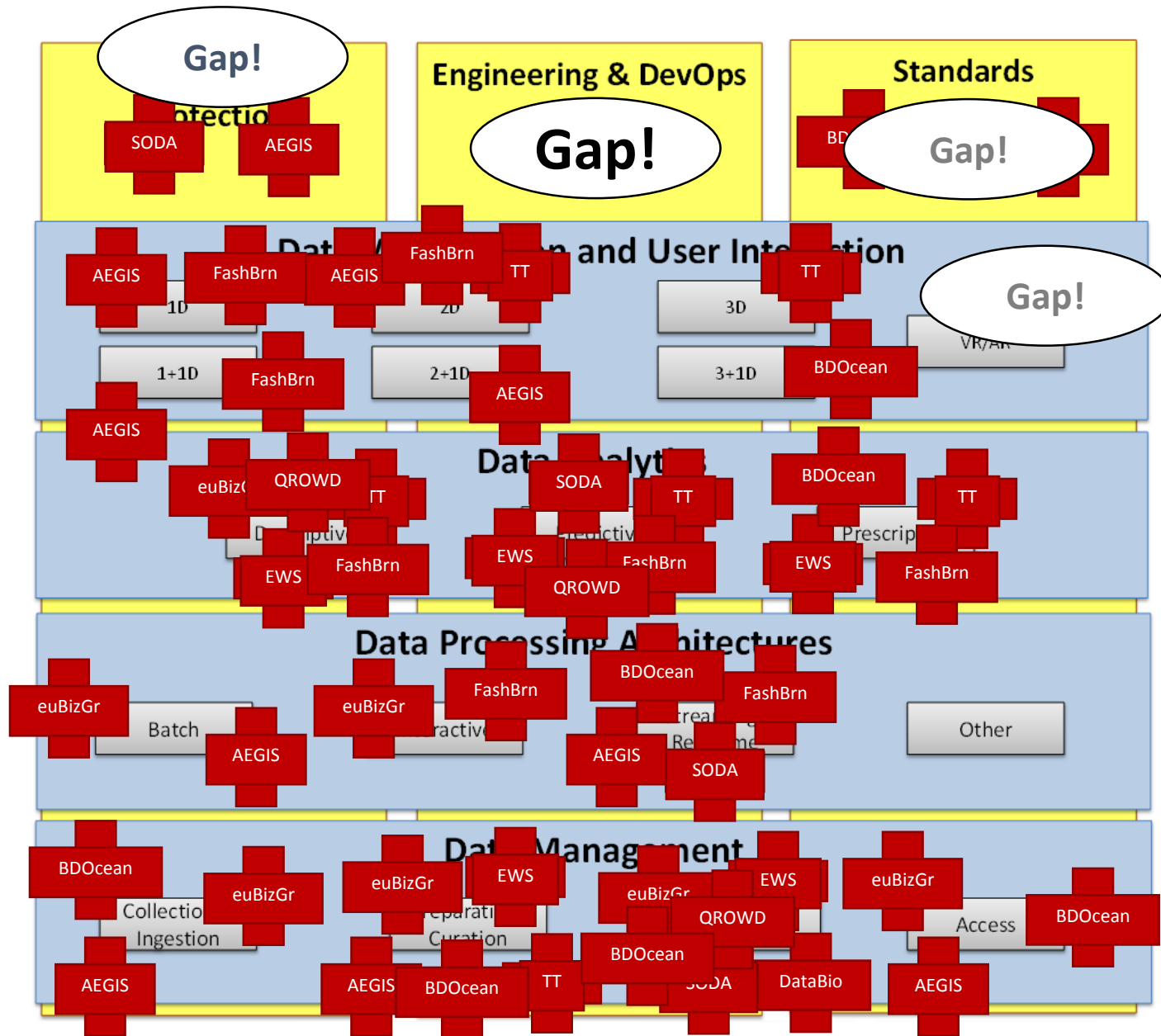


Technology Map of ongoing BDV PPP actions

Technical projects of ongoing PPP actions analysed



= project develops new technology



Drilldown: „Engineering & DevOps“

- Many Big Data technologies (“building blocks”) and “platforms” available (open source and commercial)
- **But: Lack of...**
 - ...proven and empirically sound **development and engineering paradigms and methodologies** for building next generation Big Data systems
 - ...**continuous development and operations (DevOps) approaches and techniques** for Big Data Value systems (considering different life-cycles of data and software)
 - ...end-to-end development, deployment and operations **tool chains**

Drilldown: „Engineering & DevOps“

(1) Engineering methodologies and tooling

- Systematic integration of diverse, **multi-disciplinary** aspects of data analytics, system development, quality assurance and operations
 - i.e., beyond CRISP-DM, OLAP, ...
 - *E.g., support for business analyst to express analytics questions and then seamlessly move to implementation*
- Helps addressing **skills / experience differences**



(Some) Challenges:

- How to accommodate the diverse **“data professional”** roles (e.g., data scientist, business/data analyst, ...) **as new stakeholders** during systems engineering processes?
- How to integrate different concepts, techniques and tools **along software/system life-cycle?**
- How to reconceptualise role of **requirements engineering** if big data insights not known in advance?

Drilldown: „Engineering & DevOps“

(2) Quality Assurance

- Techniques and tools for delivering trustworthy and reliable Big Data systems
- Testing, verification, monitoring, self-diagnosis, ...



(Some) Challenges:

- How to test data-intensive systems?
 - **Volume:** how to generate sufficient / representative test cases (if at all)?
 - **Velocity:** how to monitor / self-diagnose at run time?
 - **Veracity:** how to assure system quality if “uncertain” about data quality?
- How does **testing** data-intensive systems relate to **(cross-)validation** done in data analytics?
- How to provide quality guarantees, e.g., via **formal verification**, to give confidence in the systems’ quality?

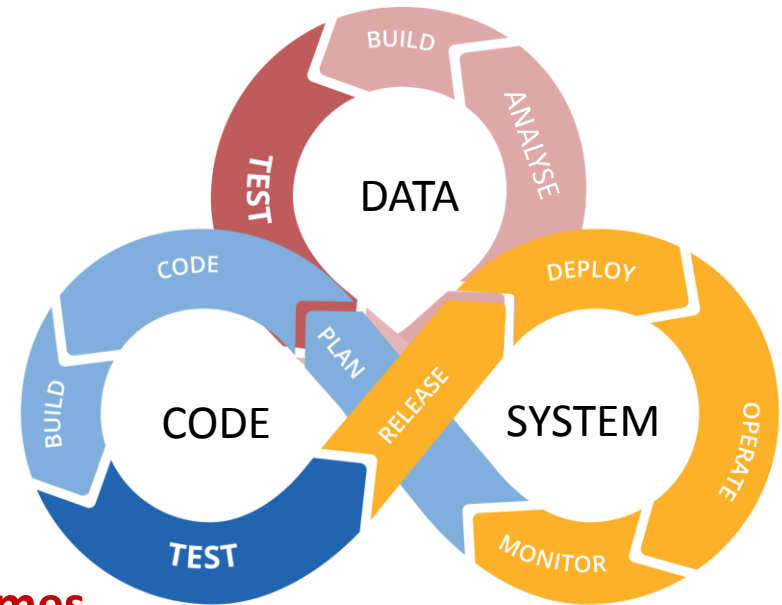
Drilldown: „Engineering & DevOps“

(3) DevOps++

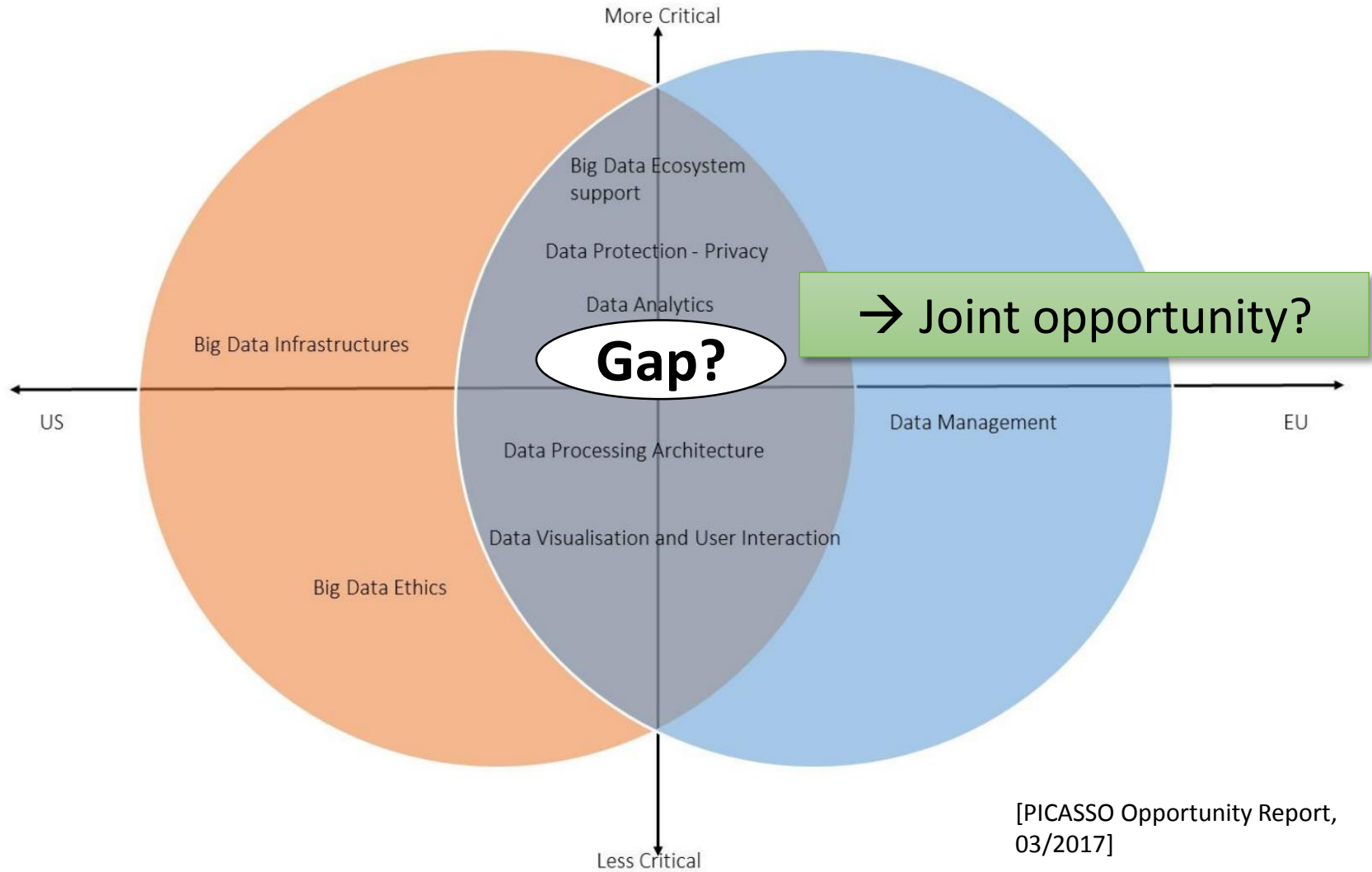
- Facilitate continuous release and improvement of big data systems
- Tight integration of system development, data analytics and cloud/fog deployment/operations

(Some) Challenges:

- How to **reconcile the different cycle times** of data life-cycle vs. system life-cycle?
- How to incorporate and align **dynamic self-adaptation** (e.g. dynamic cloud resource management vs. business process adaptation based on analytics)?
- How to **ensure data protection** along various life-cycles and dynamic adaptations?



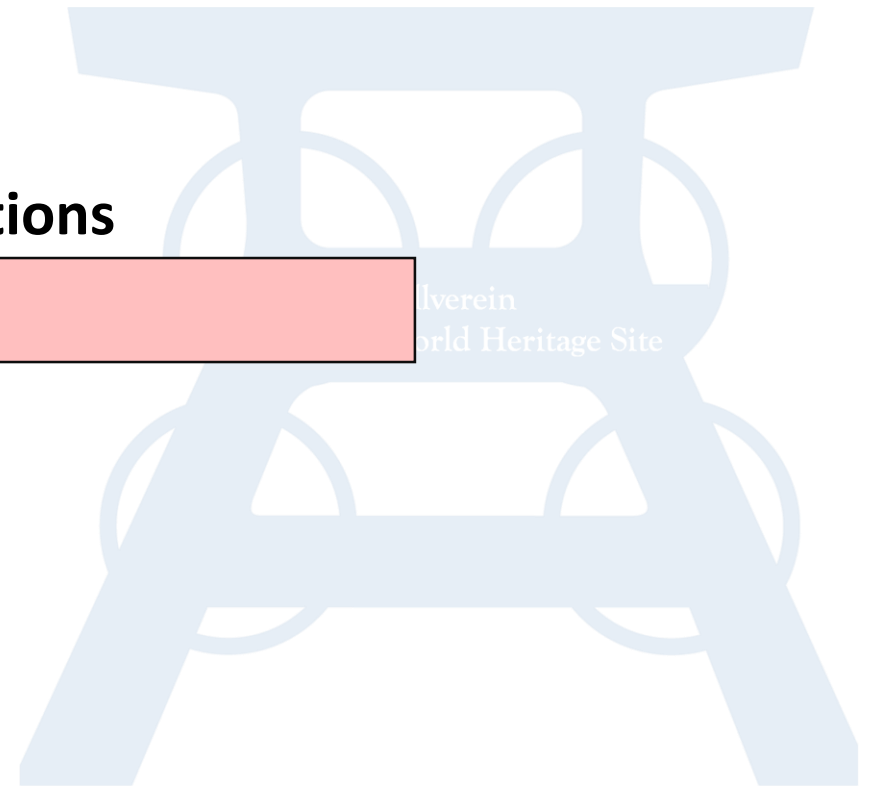
„Engineering & DevOps“ Gap also in the US?



[PICASSO Opportunity Report, 03/2017]

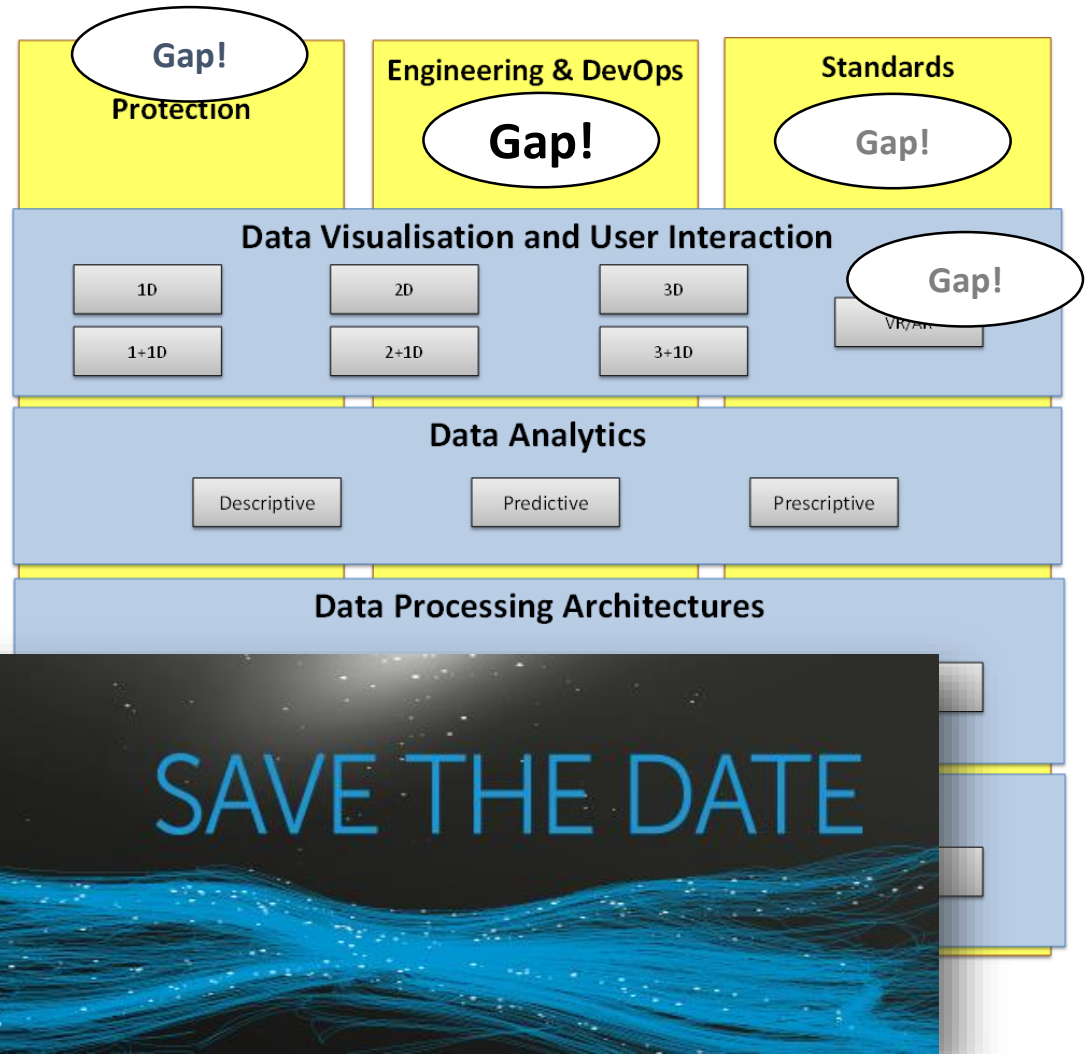
Agenda

- **BDVA Research Priorities**
- **Gap Analysis of ongoing R&I actions**
- **Conclusion**



Conclusion

R&I Opportunities:



Another Opportunity 😊