

Sponsored by



Lucerne University of  
Applied Sciences and Arts

**HOCHSCHULE  
LUZERN**

Informatik

## **Big Data Seminar 3**

Big Data bei der Swisscom

**Dr. Michael Kaufmann**

Dozent für Datenbanken

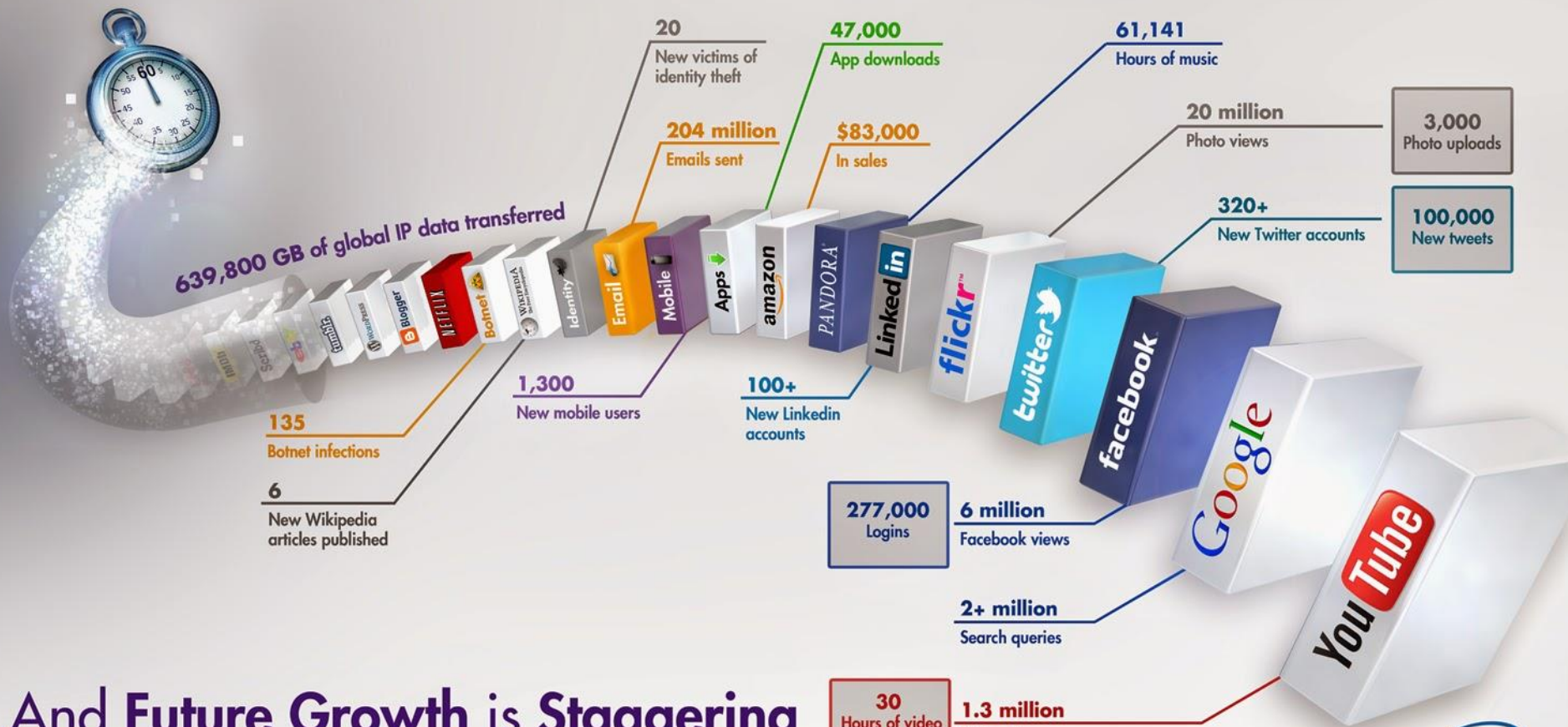
+41 41 757 68 48

m.kaufmann@hslu.ch

*Big Data Seminar der Hochschule Luzern, 27.09.2016*



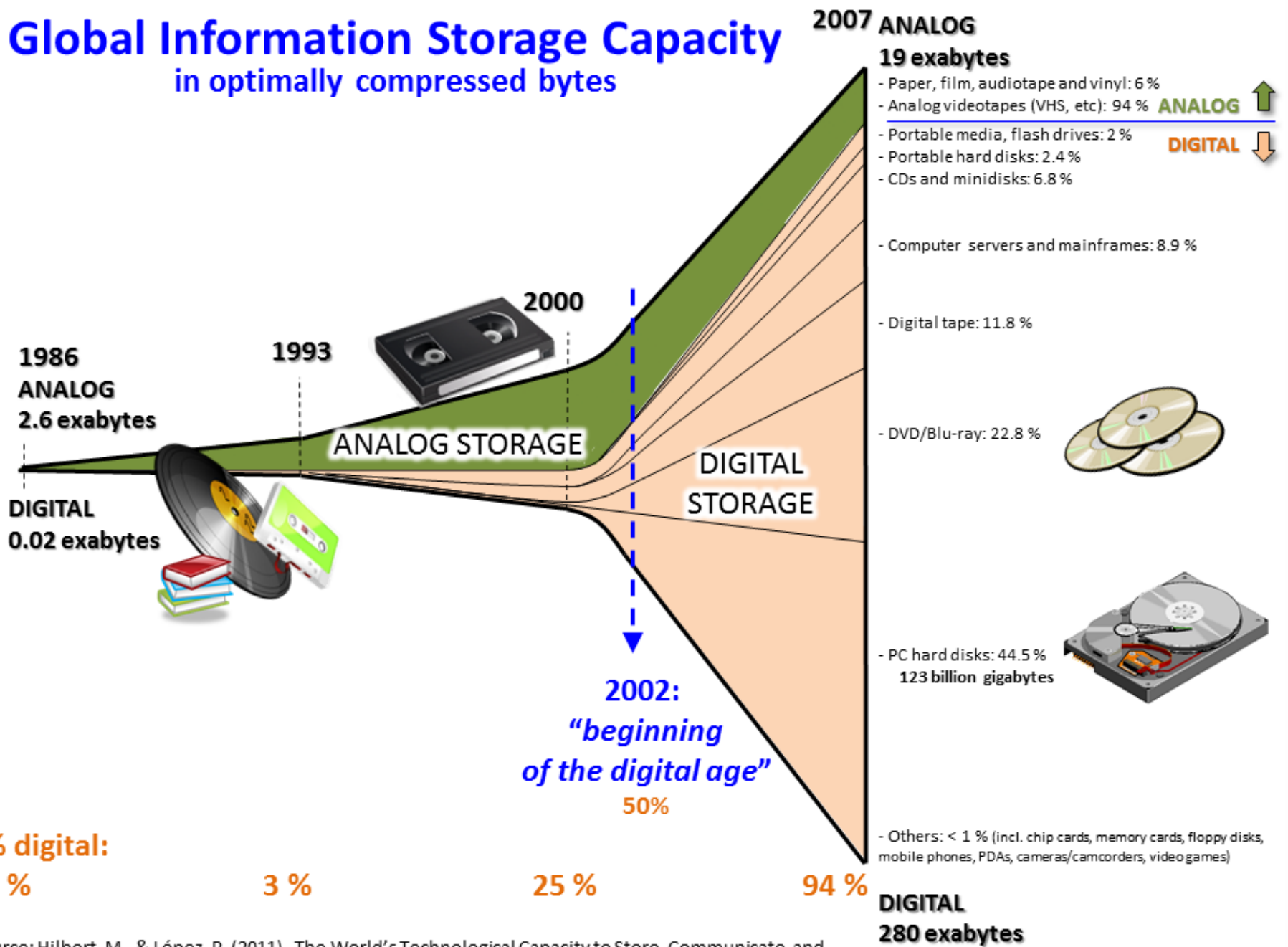
# What Happens in an Internet Minute?



## And Future Growth is Staggering



# Global Information Storage Capacity in optimally compressed bytes



Source: Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), 60–65 <http://science.sciencemag.org/content/332/6025/60>

4, 29.09.2016

(c) 2013 Myworkforwiki (CC-BY-SA 3.0)

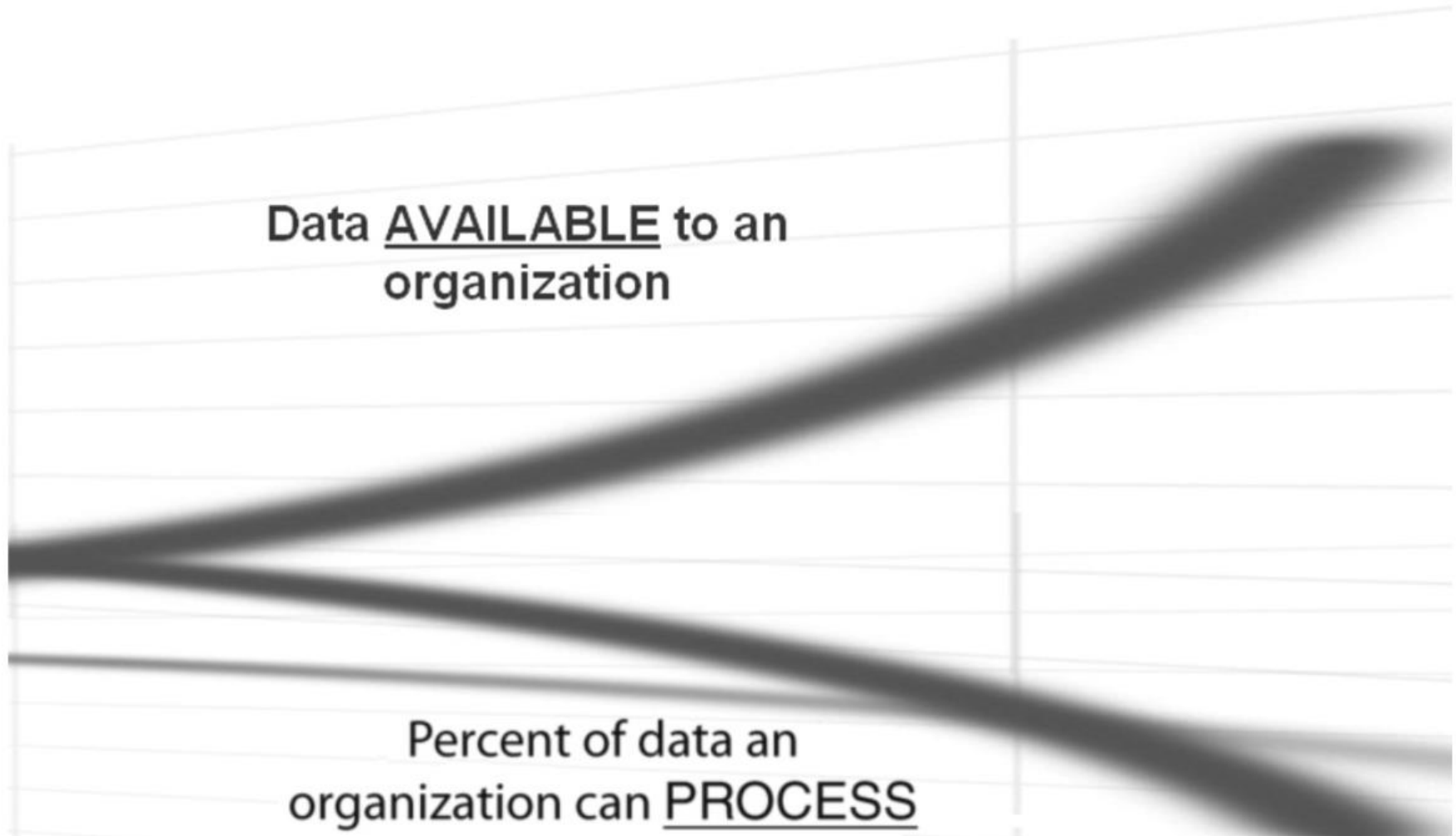


FACEBOOK  
GOOGIE  
MICROSOFT  
NSA/CIA  
STAATSSCHUTZ  
→ ALLES DIE SELBE  
SAUBANDE

---

BIG DATA  
SUCKERS

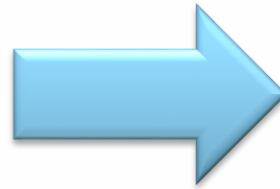
---



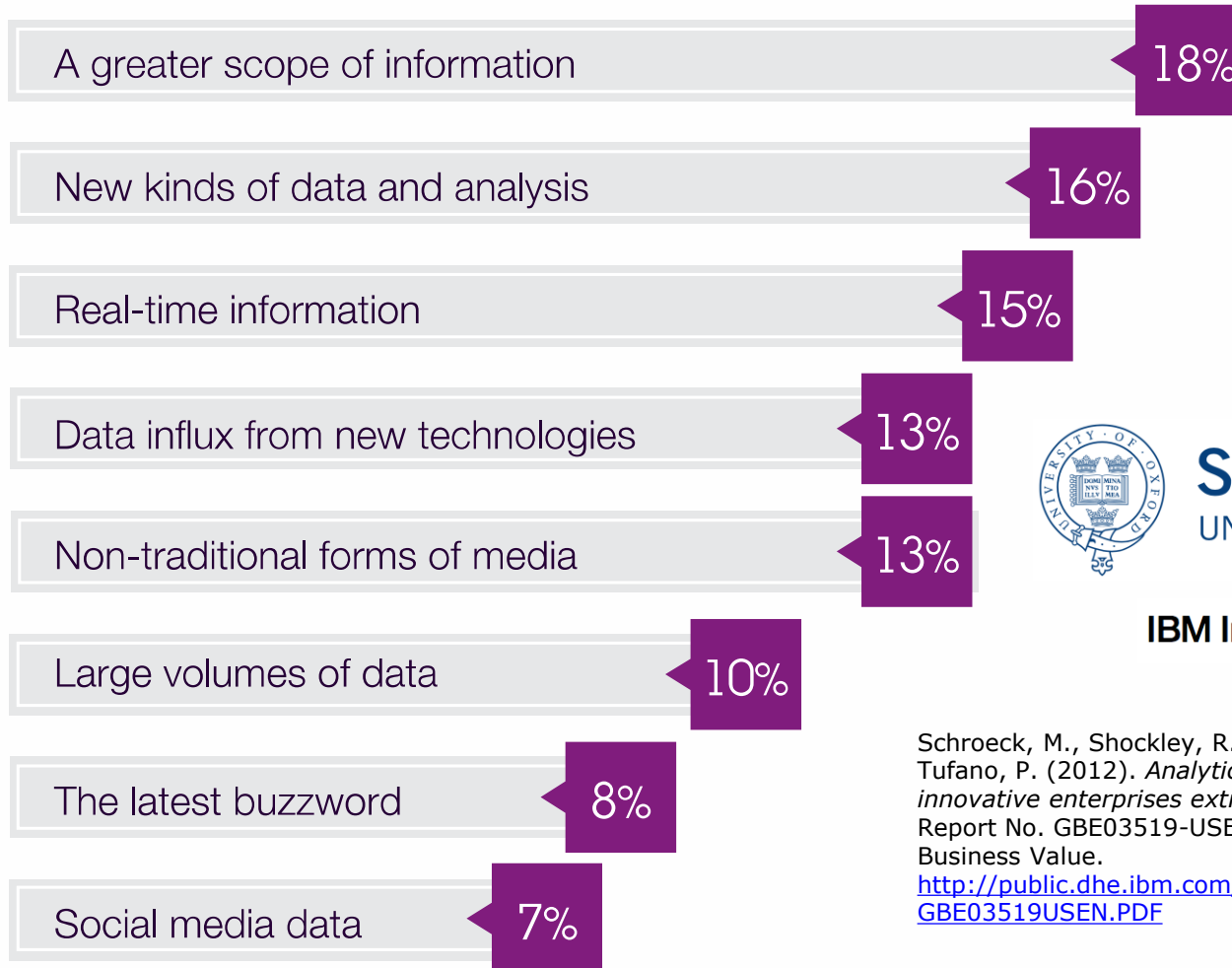
**Figure 1-2** *The volume of data available to organizations today is on the rise, while the percent of data they can analyze is on the decline.*

IBM, Zikopoulos, P., & Eaton, C. (2011). *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data* (1st ed.). McGraw-Hill Osborne Media. Retrieved from [https://www.ibm.com/developerworks/community/blogs/SusanVisser/entry/flashbook\\_understanding\\_big\\_data\\_analytics\\_for\\_enterprise\\_class\\_hadoop\\_and\\_streaming\\_data?lang=en](https://www.ibm.com/developerworks/community/blogs/SusanVisser/entry/flashbook_understanding_big_data_analytics_for_enterprise_class_hadoop_and_streaming_data?lang=en)

## From Big Data to Lean Knowledge



# Defining big data



**Saïd Business School**  
UNIVERSITY OF OXFORD

**IBM Institute for Business Value**

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). *Analytics: The real-world use of big data - How innovative enterprises extract value from uncertain data* (Executive Report No. GBE03519-USEN-00). New York: IBM Institute for Business Value.  
<http://public.dhe.ibm.com/common/ssi/ecm/gb/en/gbe03519usen/GBE03519USEN.PDF>

Respondents were asked to choose up to two descriptions about how their organizations view big data from the choices above. Choices have been abbreviated, and selections have been normalized to equal 100%. Total respondents=1144.



## 4V Big Data Definition von IBM

### Volume



#### Data at scale

Terabytes to  
petabytes of data

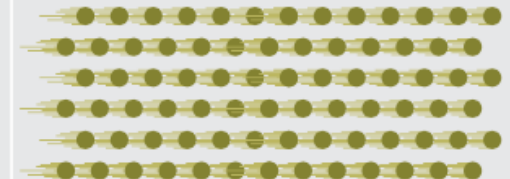
### Variety



#### Data in many forms

Structured, unstructured,  
text, multimedia

### Velocity



#### Data in motion

Analysis of streaming data  
to enable decisions within  
fractions of a second

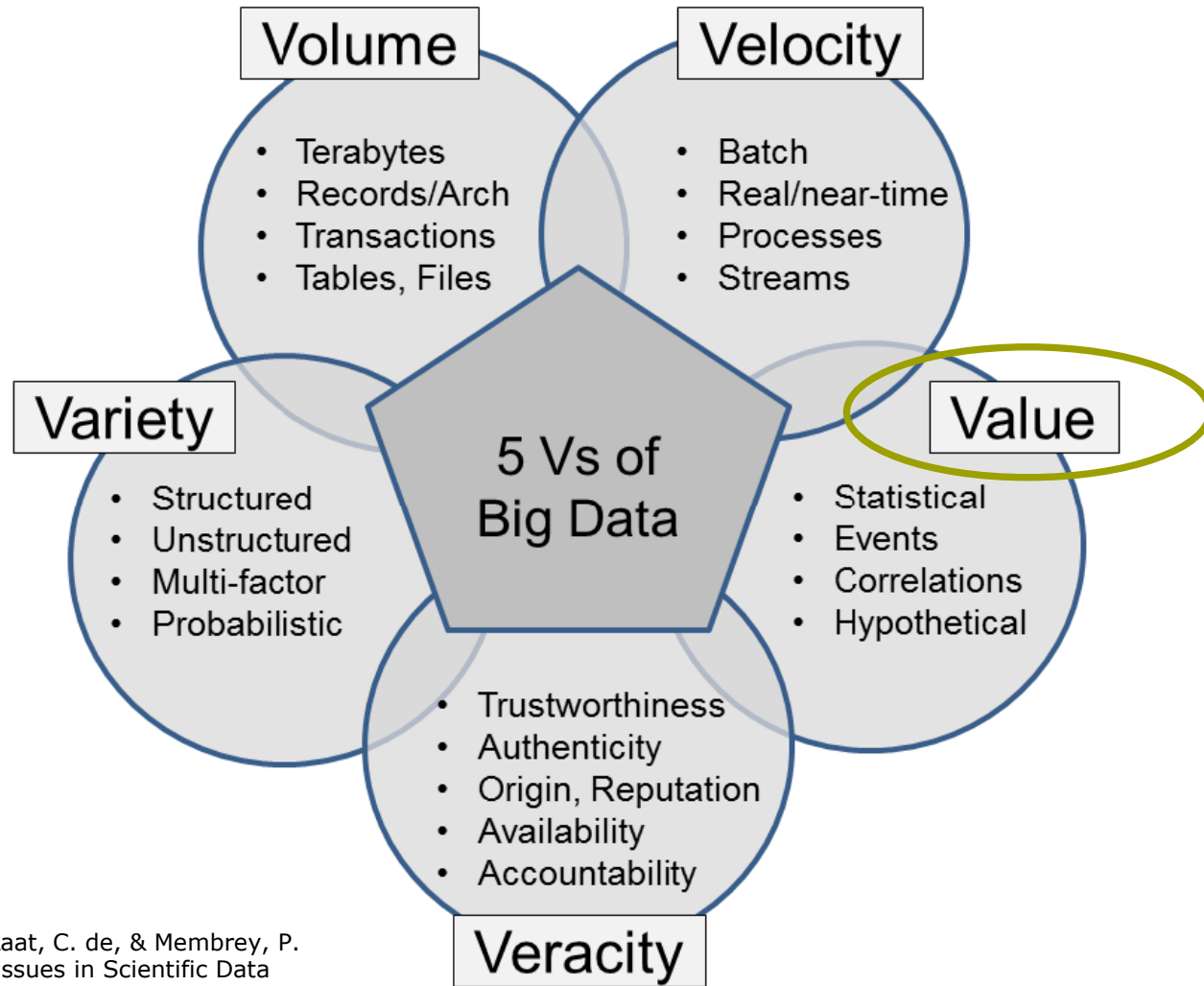
### Veracity



#### Data uncertainty

Managing the reliability and predictability  
of inherently imprecise data types

## 5V-Model of Big Data (Demchenko et al. 2013)



## Die NIST-Standard-Definition des Begriffs „Big Data“: 4V

- US National Institute for Standards and Technology - Big Data Public Working Group:
- “*Big Data* consists of extensive **datasets** primarily in the characteristics of **volume, variety, velocity, and/or variability** that require a scalable architecture for efficient storage, manipulation, and analysis.” (NIST, 2015, p. 5)

big data  $\subset$  data

## What Big Data isn't

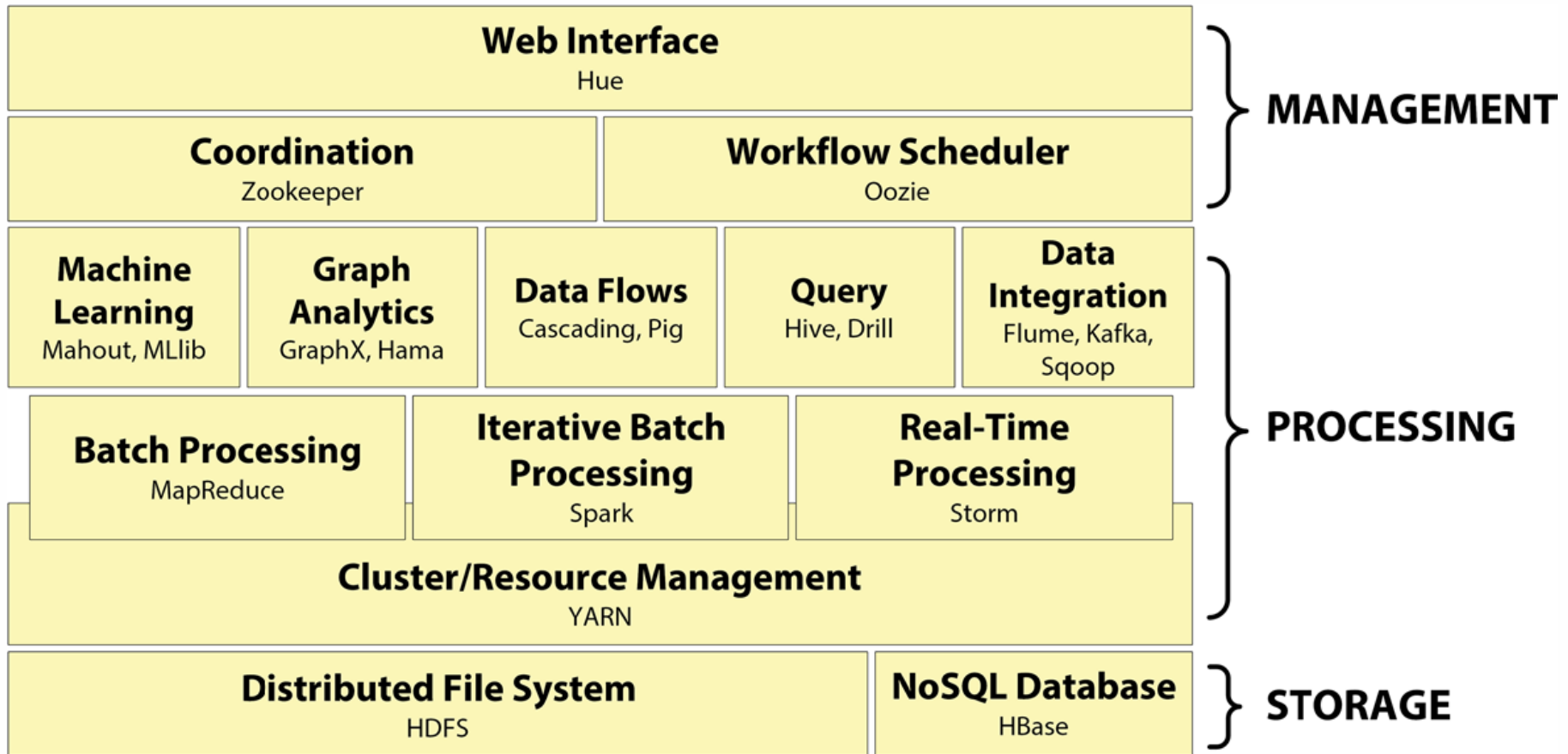
- “**Data science** is the extraction of actionable knowledge directly from data through a process of discovery, or hypothesis formulation and hypothesis testing” (NIST, 2015, p. 7)

big data  $\neq$  data science

- “The Big Data **paradigm** consists of the distribution of data systems across horizontally coupled, independent resources to achieve the scalability needed for the efficient processing of extensive datasets”. (NIST, 2015, p. 5)

big data  $\neq$  Hadoop

# The Hadoop-Ecosystem



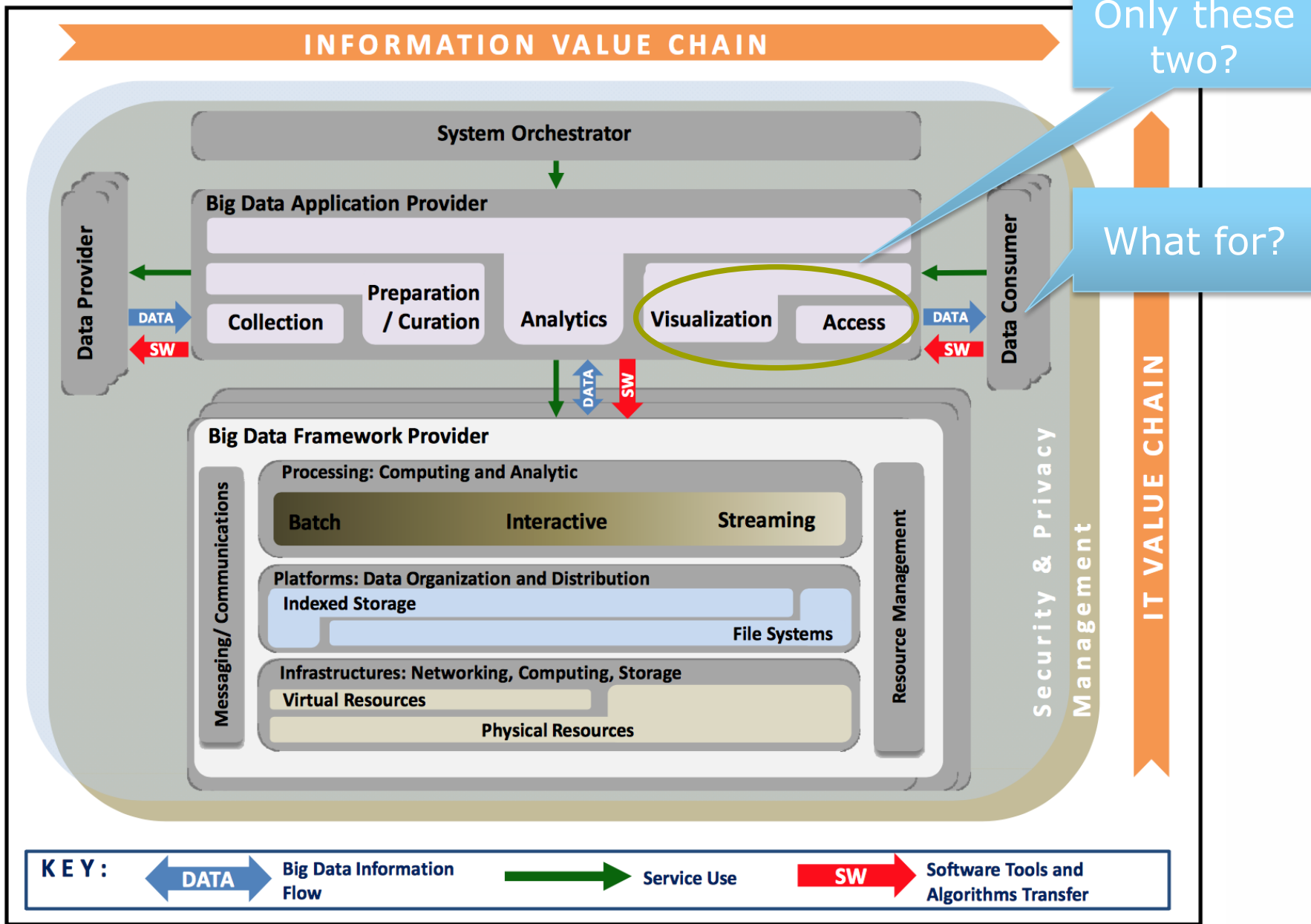
## The NIST Data Lifecycle

The data life cycle consists of the following four stages:

- **Collection:** This stage gathers and stores data in its original form (i.e., raw data).
- **Preparation:** This stage involves the collection of processes that convert raw data into cleansed, organized information.
- **Analysis:** This stage involves the techniques that produce synthesized knowledge from organized information.
- **Action:** This stage involves processes that use the synthesized knowledge to generate **value** for the enterprise.



But... how  
exactly?

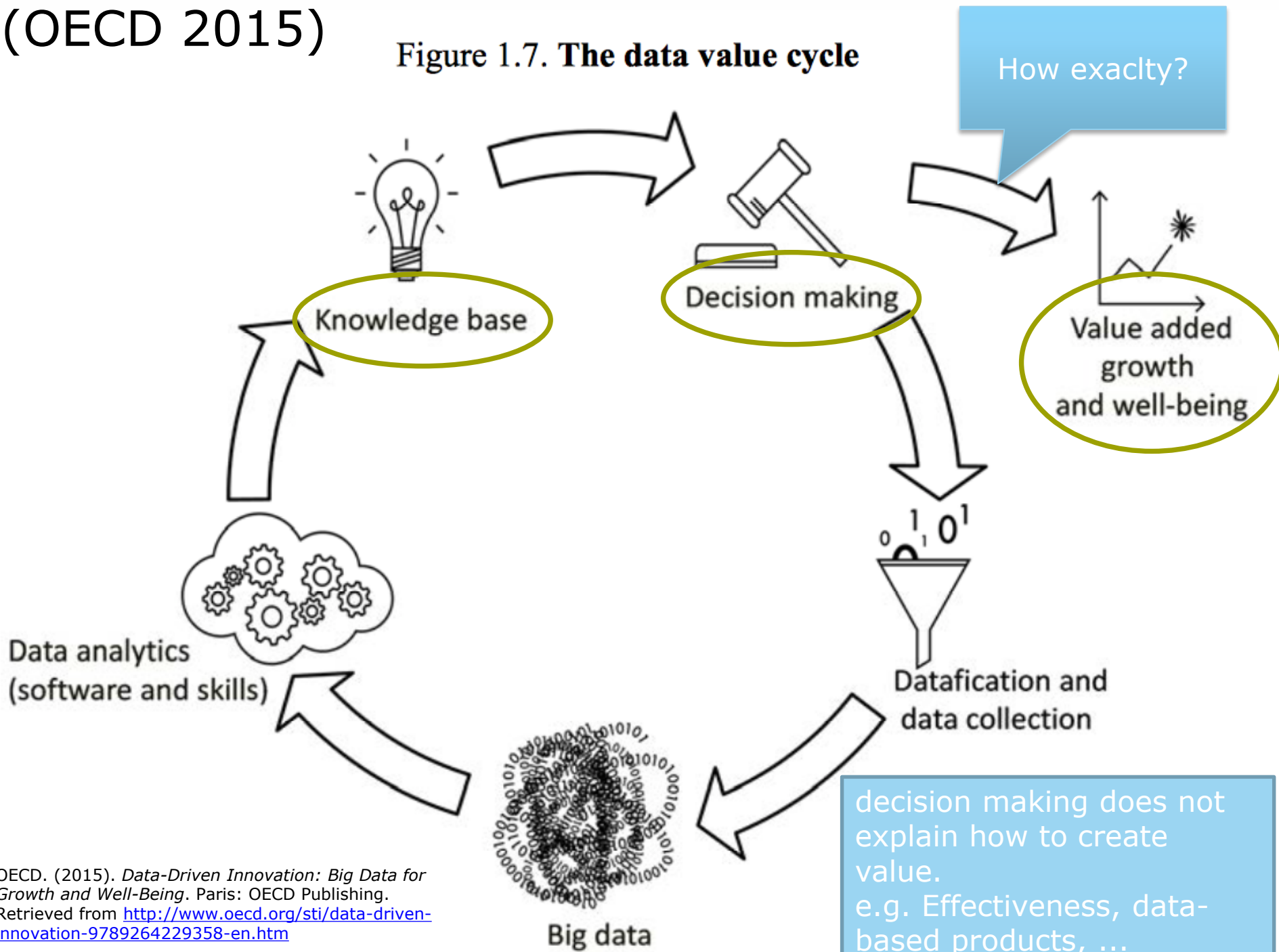


**Figure 1: NIST Big Data Reference Architecture**

NIST. (2015). *NIST Big Data Interoperability Framework: Volume 2, Taxonomies* (NIST Special Publication No. 1500-1). NIST Big Data Public Working Group. Retrieved from <http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-2.pdf>

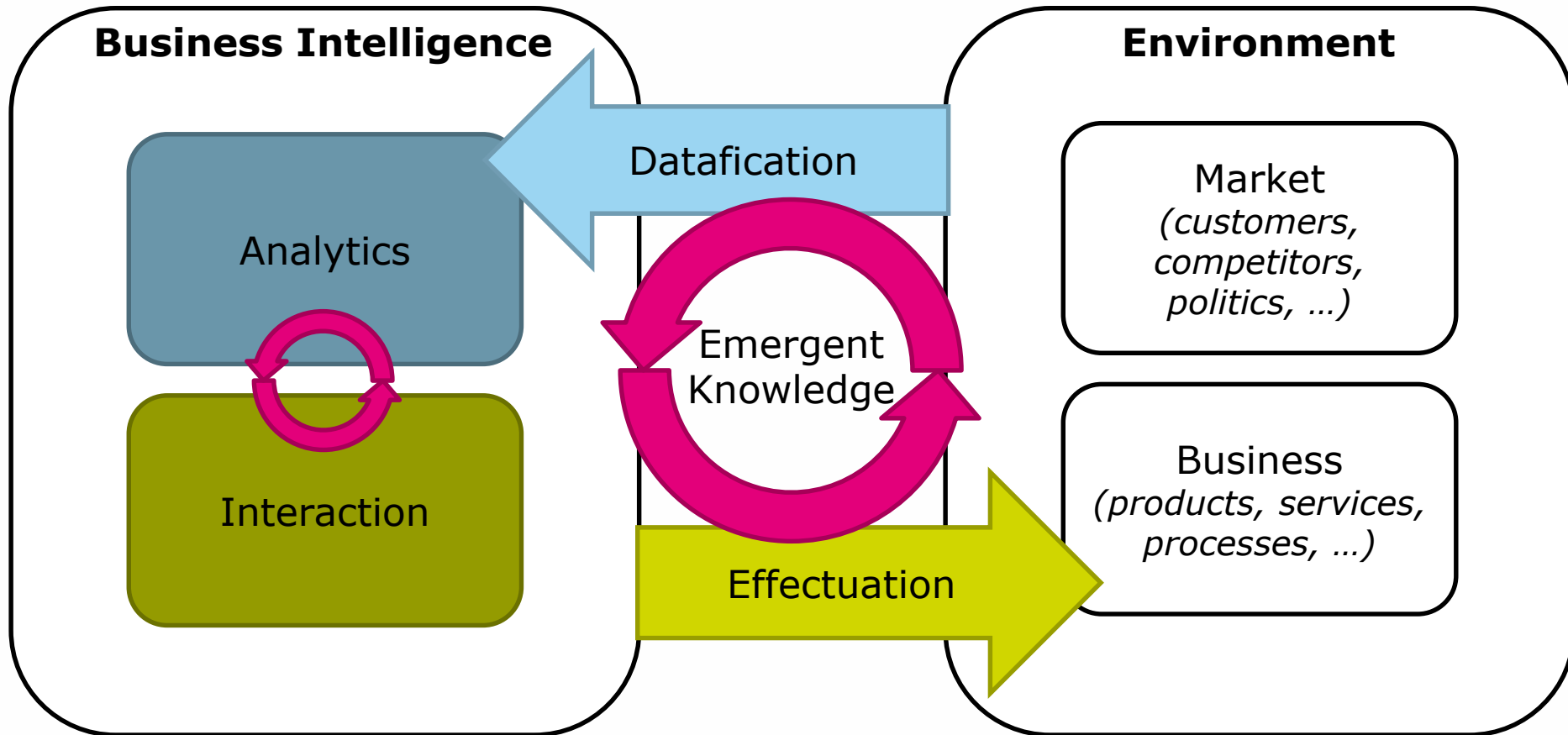
(OECD 2015)

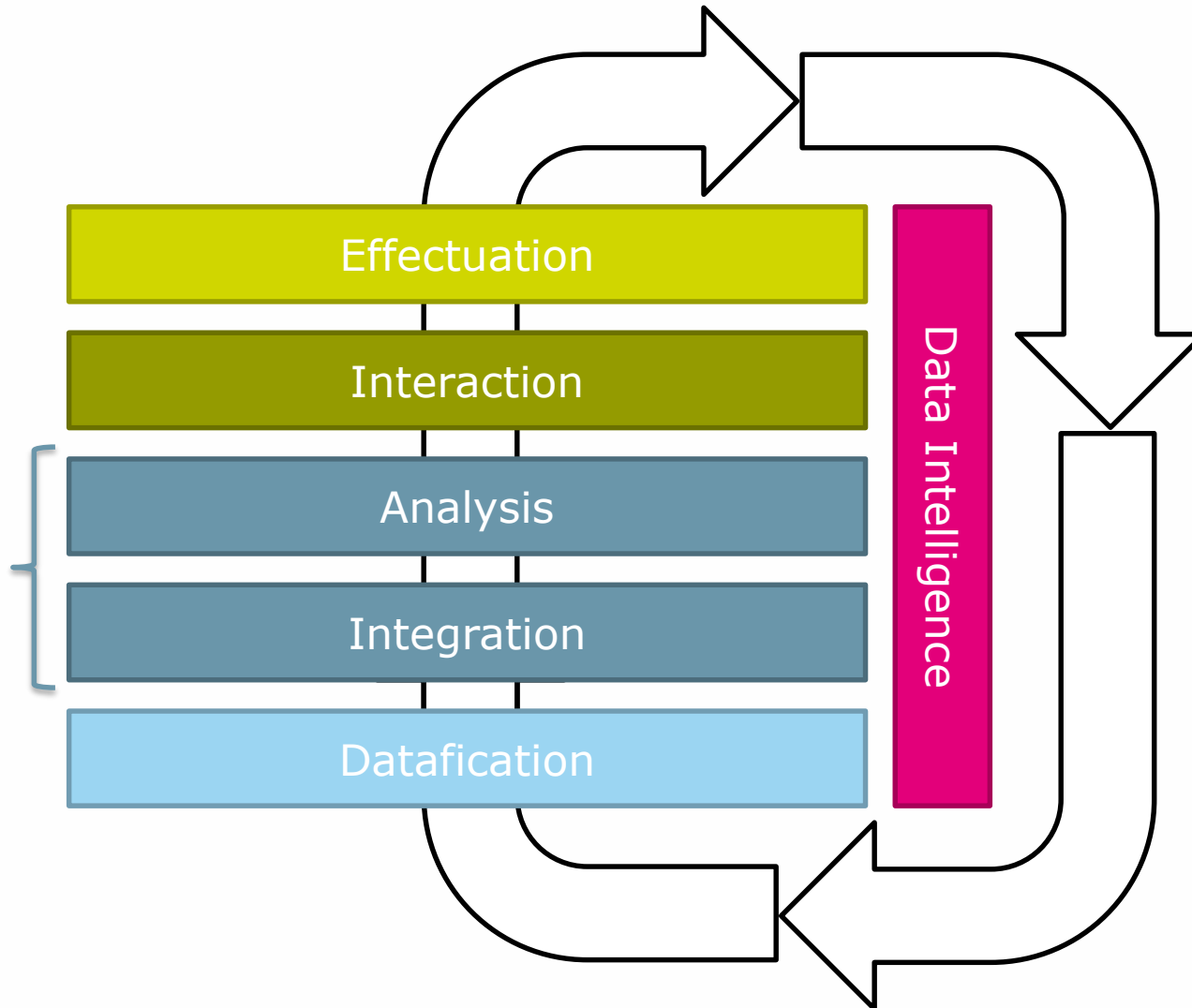
Figure 1.7. The data value cycle



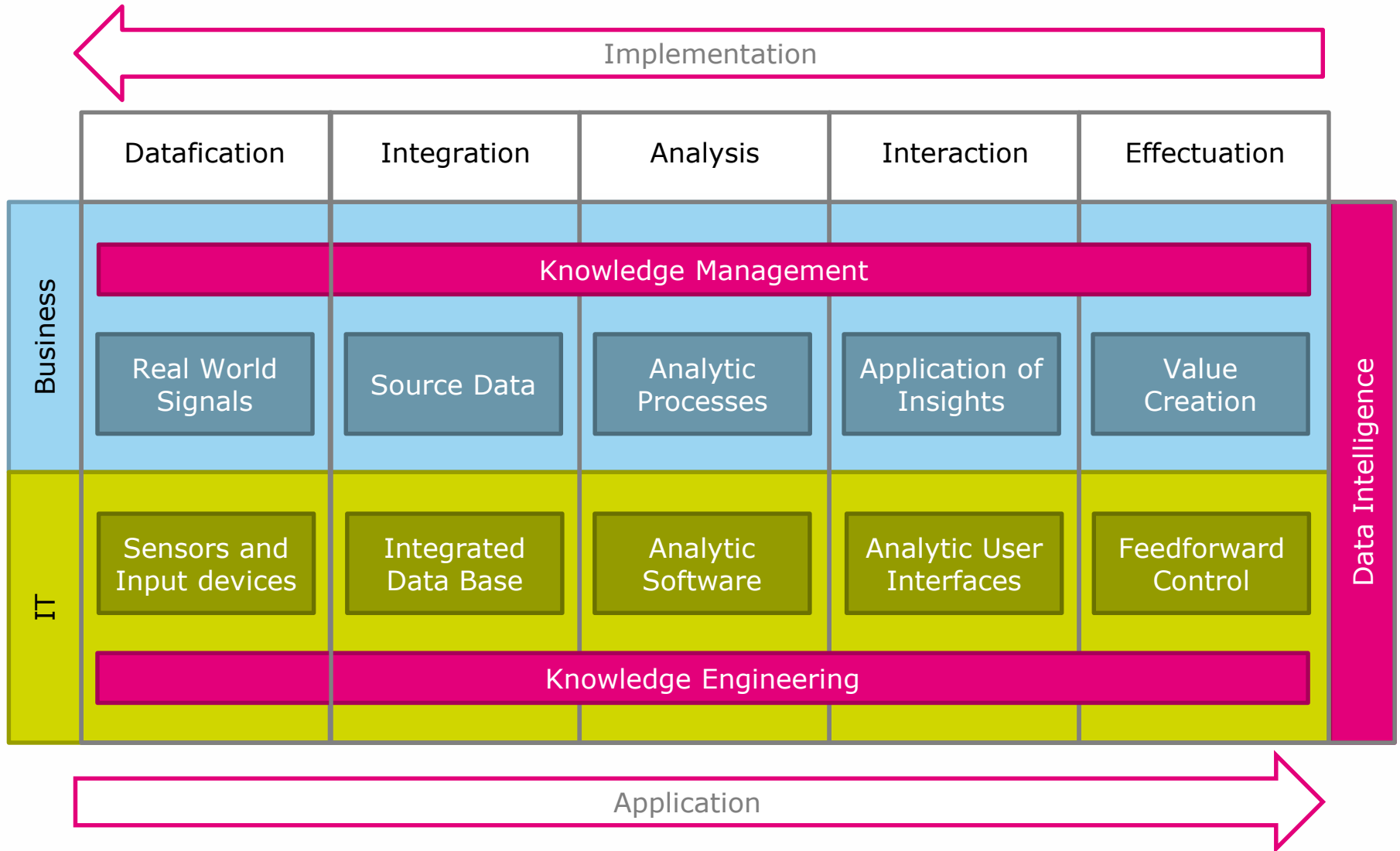


## Business Intelligence (BI) is a *cognitive socio-technical system*

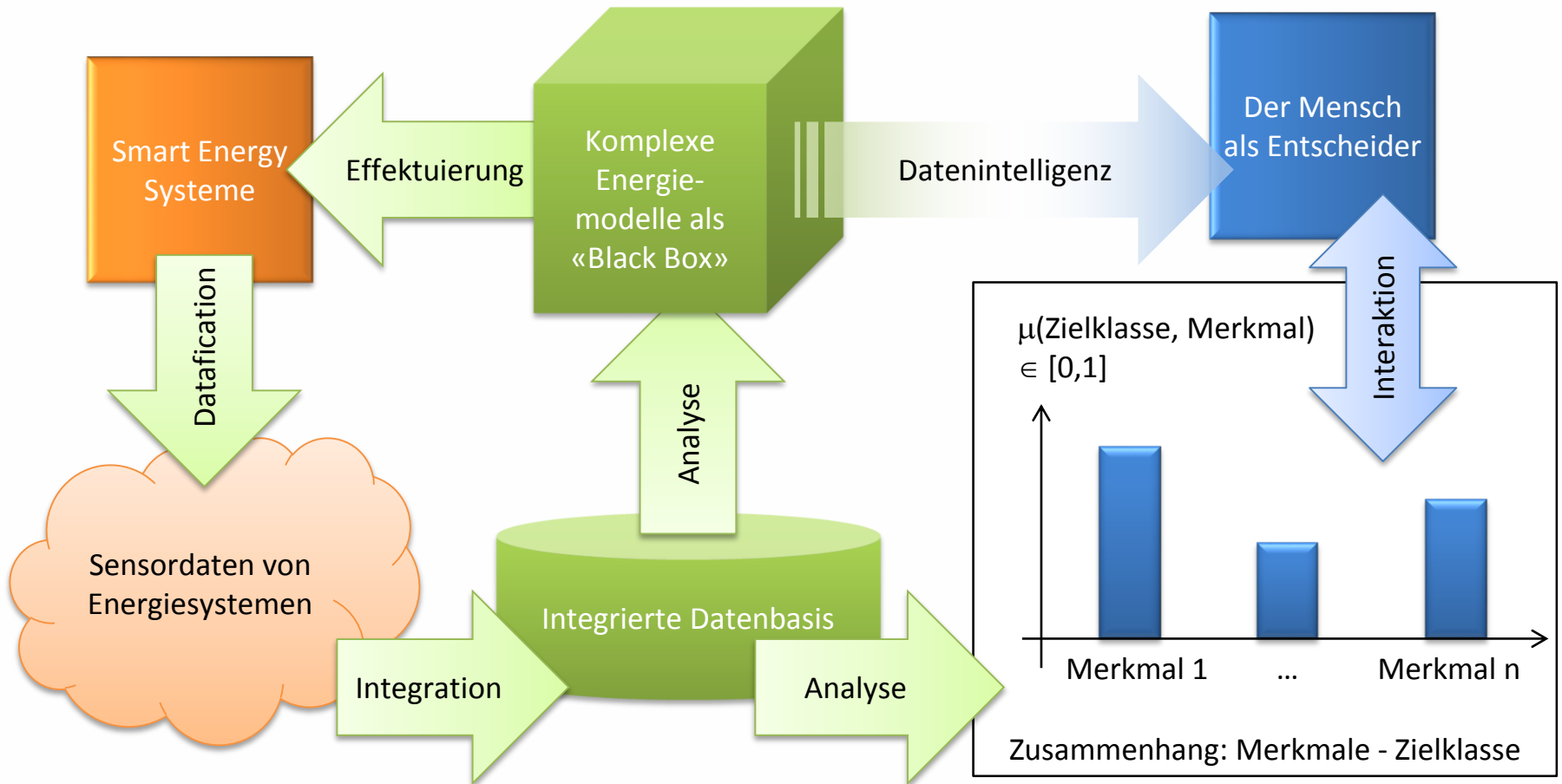




# BDM<sup>cube</sup> for business IT alignment in big data management



## Case Study: enersis suisse AG



[2] Kaufmann, M., Koller, T., Kurochina, D., Stoffel, K., Hemmje, M. (2017). Induktive Datenvisualisierung für Smart Energy mit dem IFC-Filter: Fallstudie enersis suisse AG. To appear in: Meier, A., Portmann, E. (Ed.) Smart City - Strategie, Governance & Projekte. (to appear)

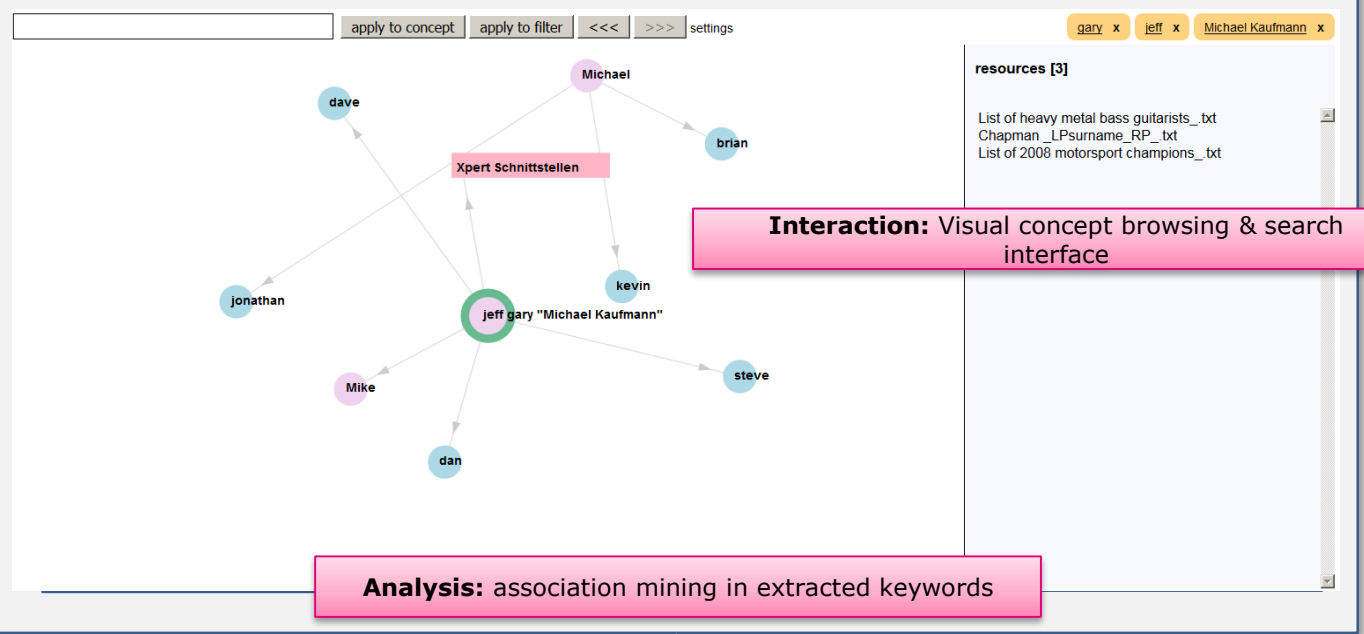
[3] Kaufmann, M., Meier, A., & Stoffel, K. (2015). IFC-Filter: Membership function generation for inductive fuzzy classification. Expert Systems with Applications, 42(21), 8369–8379. <http://doi.org/10.1016/j.eswa.2015.06.034>

# Case Study FIVE Informatik AG: Integration of Predefined Ontologies and Emergent Knowledge Graphs in a Concept Brower

**Effectuation:** Evaluated as not appropriate for FIVE. Emergent associations often useless or distracting. FIVE missed direct editing, known structure, facets, workflow

**Intelligence:** emergent & encoded knowledge

## C. Visualization as knowledge map in the search engine



**Interaction:** Visual concept browsing & search interface

**Analysis:** association mining in extracted keywords

**Integration:** Search engine index



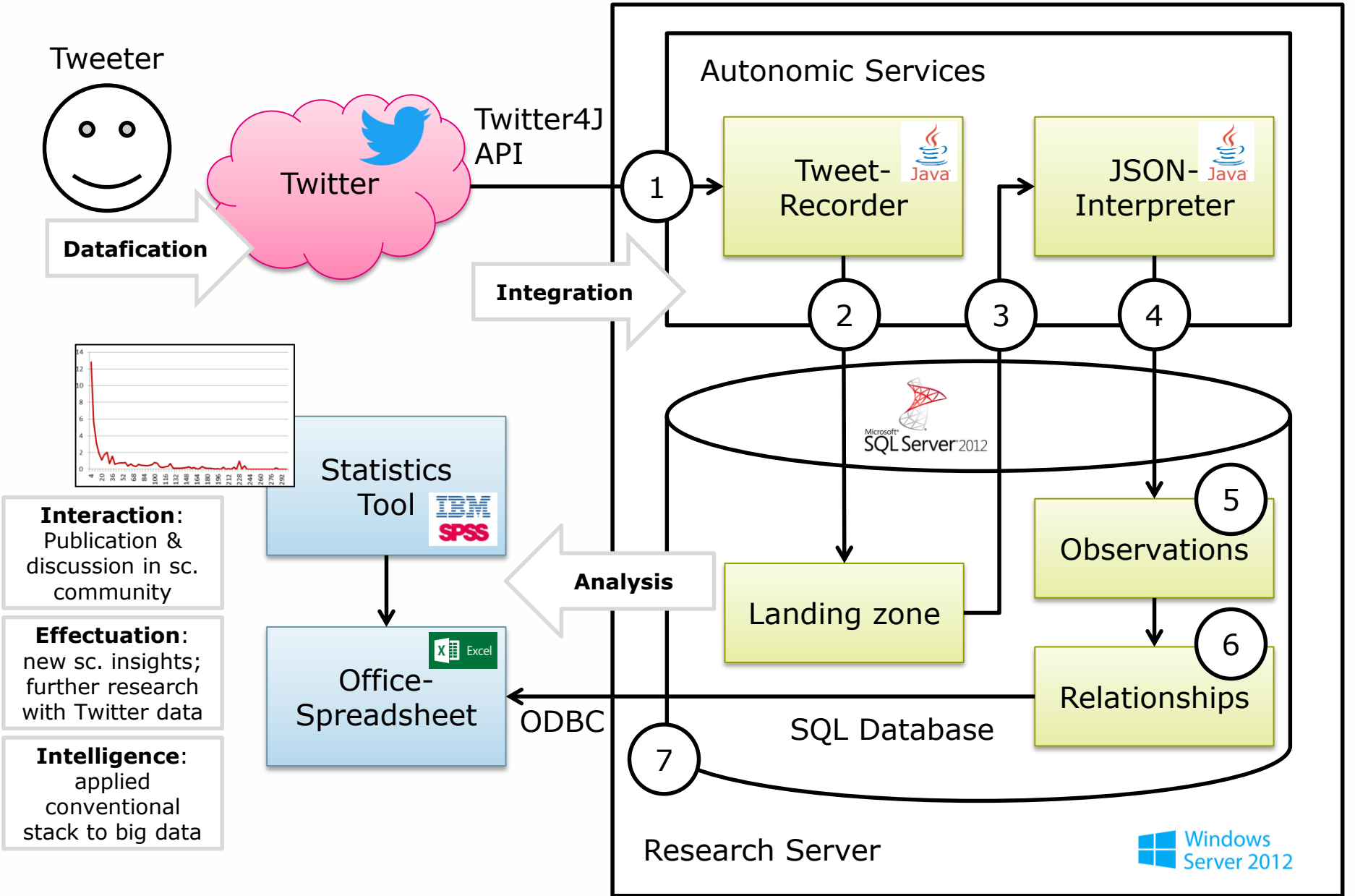
**Datafication:** business processes



B. Enterprise ontology: top-down structuring

Wilke, G., Emmenegger, S., Lutz, J., & Kaufmann, M. (2016). Merging Bottom-Up and Top-Down Knowledge Graphs for Intuitive Knowledge Browsing. In T. Andreasen et al. (Eds.), *Flexible Query Answering Systems 2015* (pp. 445–459). Springer  
[http://link.springer.com/chapter/10.1007/978-3-319-26154-6\\_34](http://link.springer.com/chapter/10.1007/978-3-319-26154-6_34)

# Case study: Twitter Data Research @ HSLU

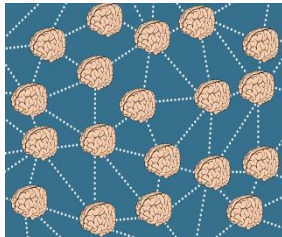


Delbiaggio, K., Hauser, C., Kaufmann, M. (2016). "The proximity Bias of Communication Recorded on Twitter in Switzerland", in I. Bernhard (Ed.) Uddevalla Symposium 2016, Geography, Open Innovation, Diversity and Entrepreneurship. Revised papers presented at the 19th Uddevalla Symposium, June 30th to July 2nd at Birkbeck University of London, United Kingdom.

## IDS Datenwelten Hochschule Luzern

- Der interdisziplinäre Schwerpunkt (IDS) Datenwelten ist der Big Data Think Tank der Hochschule Luzern.

Experten  
Netzwerk



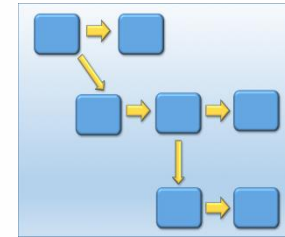
X-Lab



Events,  
Hackathons



Interdisziplinäre  
Projekte



- Der IDS Datenwelten möchte anhand konkreter Beispiele aufzeigen, **wie aus Daten Werte geschaffen werden können.**
- Die Besonderheit des IDS liegt in seiner **interdisziplinären Sicht** auf die nachhaltige Wertschöpfung aus Daten.

## Forschungsgruppe Data Intelligence

- Datenintelligenz ist die Kompetenz einer Person oder einer Organisation, Wissen und Fähigkeiten aufgrund von Daten zu erlangen und anzuwenden.
- Die Forschungsgruppe Data Intelligence entwickelt, erforscht und evaluiert Methoden, mit denen sich Sensordaten erfassen, Wissen aus Daten gewinnen und die Interaktion mit Daten vereinfachen lässt.



Peter Sollberger



Michael Kaufmann



Mariia Astrakhantceva



Dozent/in Data Intelligence





**Partnerschaft mit dem Big Data Seminar  
und der Forschungsgruppe Data Intelligence  
der Hochschule Luzern – Informatik**

Lucerne University of  
Applied Sciences and Arts

**HOCHSCHULE  
LUZERN**

Informatik

[m.kaufmann@hslu.ch](mailto:m.kaufmann@hslu.ch)  
[hslu.ch/bigdata](http://hslu.ch/bigdata)

- ☑ Sponsern Sie das Big Data Seminar
- ☑ Werden Sie Gastgeber beim nächsten Event
- ☑ Schaffen Sie Innovation in F&E Projekten mit uns