SWIB19
Semantic Web in Libraries

# Smart Data for Digital Humanities

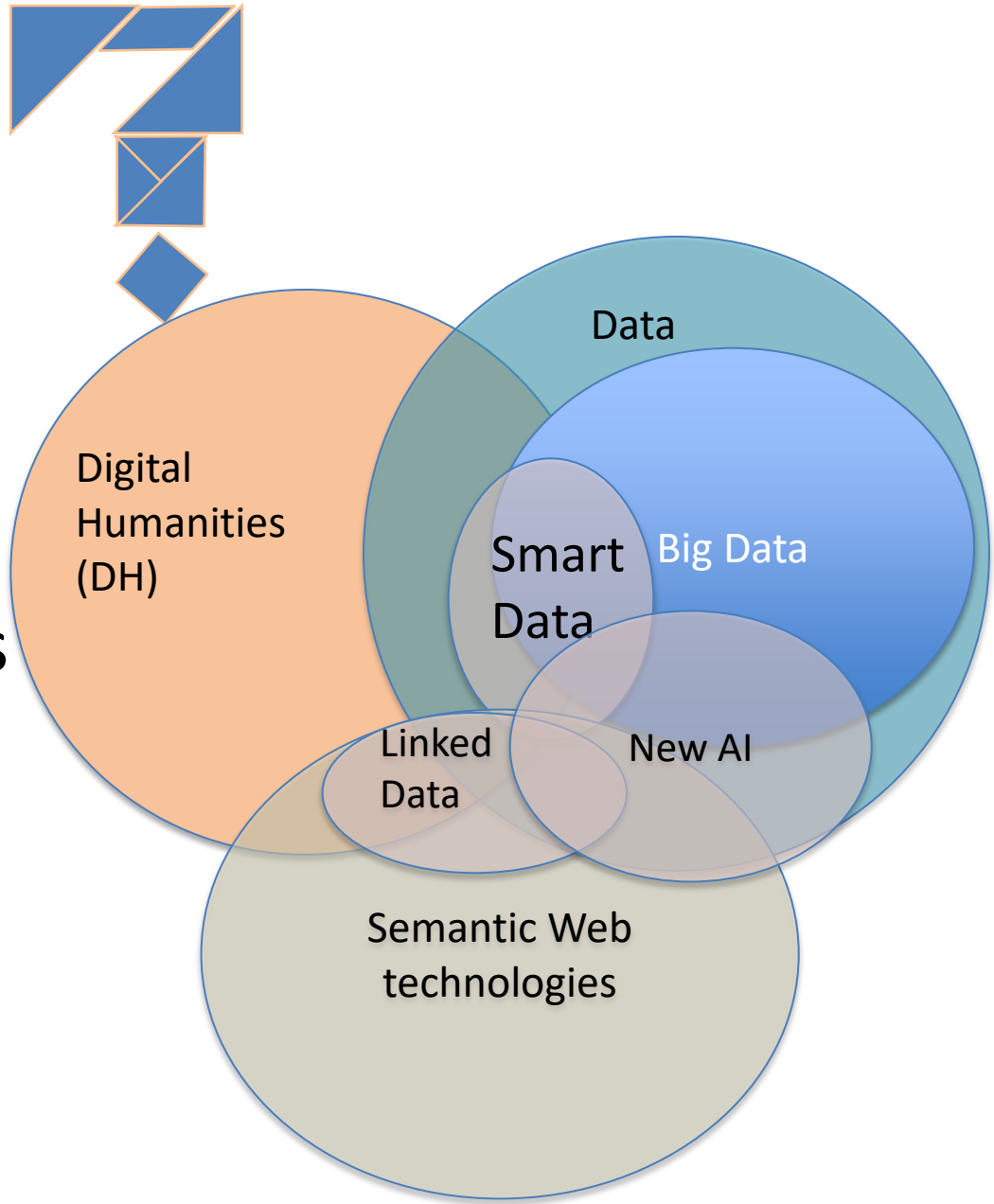

Marcia Lei Zeng

Kent State University

http://marciazeng.slis.kent.edu/

# *What do I mean…*

- Smart data
  - in the context of Big Data
- Digital Humanities
  - in the context of heritage institutions' data



# Outline

1. **Smart Data**
2. **Smart Data *in* DH**
3. **Smart Data *for* DH → in the context of LAMs' data**

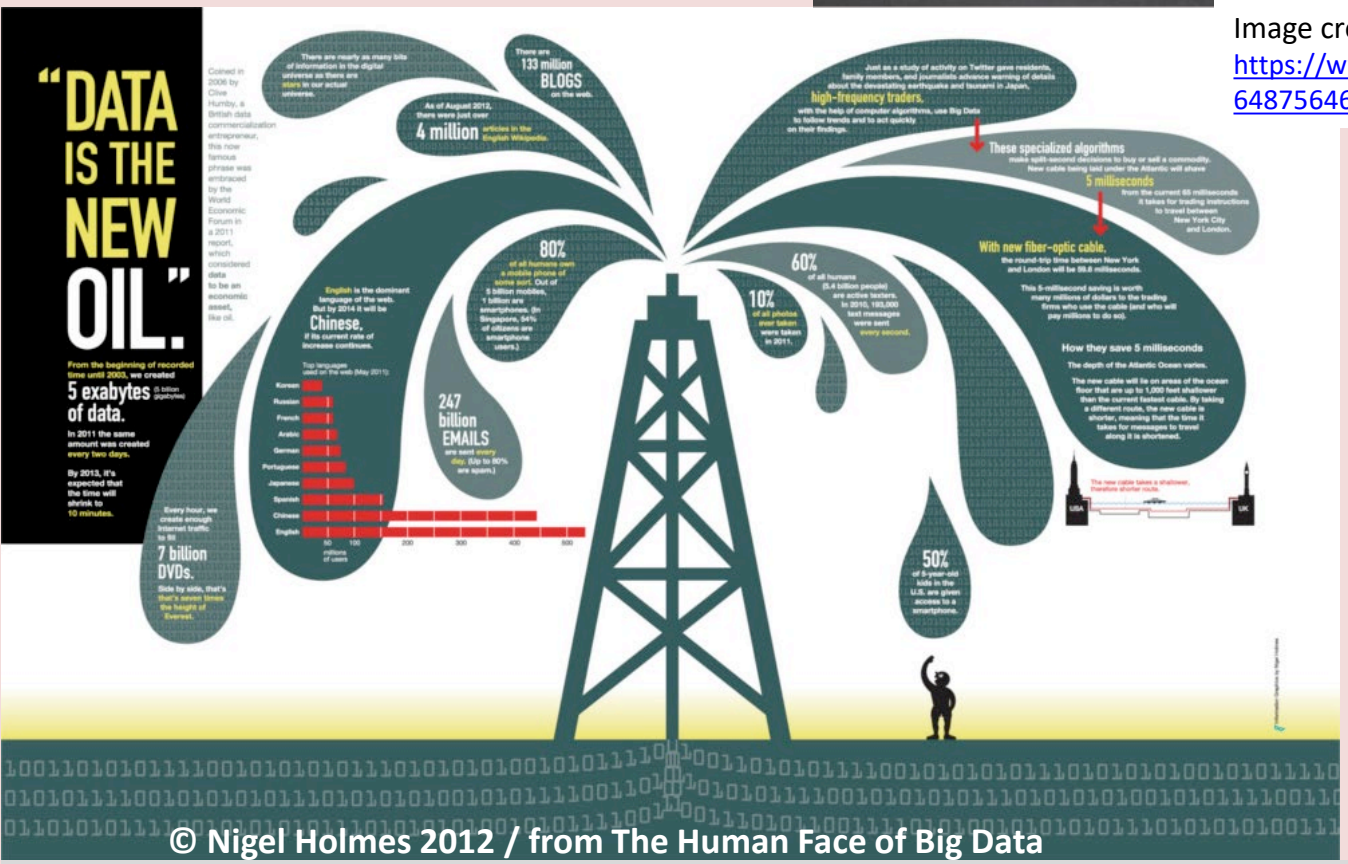
# 1. Smart Data

Why
What
How

WHY

# From Big Data to Smart Data


Image credit: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know




© Nigel Holmes 2012 / from The Human Face of Big Data

Image credit: https://i.pinimg.com/originals/68/4b/11/684b11b98a4bb4413b81d1feaed2e8e1.jpg

Only 3% were 'tagged' and 0.5% were analyzed in the *digital universe*. (Gantz & Reinsel, 2012, p. 3).

# Data is new oil

"However, in its raw form, data is just like crude oil; it needs to be refined and processed in order to generate real value. Data has to be cleaned, transformed, and analyzed to unlock its hidden potential."

TiECON East. 2014. *Data is new oil.*
http://www.tieconeast.org/2014/big-data-analytics

## Big Data is like Crude Oil { not new Oil }

Think of data as crude oil !

Big Data is like extracting the crude oil, transporting it in mega tankers, pumping it through pipelines and storing it in massive silos

But what about refining ?



Prithwis Mukerjee (2014). Introduction to Data Science
http://www.slideshare.net/prithwis/01-intro2-datascienceyantrajaalblog

WHAT

"Smart Data" means information that actually makes sense.

-- Lorentz, A. 2013. Big Data, Fast Data, Smart Data

Big Data is the problem and Smart Data is the solution.

-- Rice, E. 2014. Going from Big Data to Smart Data

The relationship between Big Data and Smart Data can be characterized as "what it is" and "what it is for" .

-- lafrate: *From Big data to Smart data.*

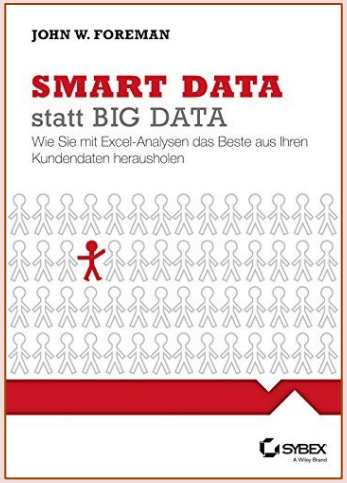
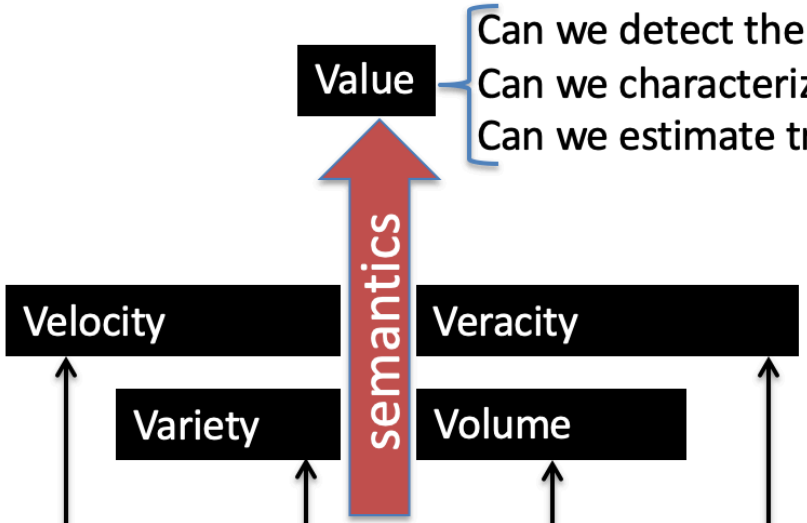



Image of Foreman, John W. and Jutta Schmidt. 2014. *Smart Data statt Big Data: Wie Sie mit Excel-Analysen das Beste aus Ihren Kundendaten herausholen* (German Edition)

KNO.E.SIS

# Big Data to Smart Data: Traffic Management example

Vehicular traffic data from San Francisco Bay Area aggregated from on-road sensors (numerical) and incident reports (textual)




Can we detect the onset of traffic congestion?
Can we characterize traffic congestion based on events?
Can we estimate traffic delays in a road network?

Representing prior knowledge of traffic lead to a focused exploration of this massive dataset

Every minute update of speed, volume, travel time, and occupancy resulting in 178 million link status observations, 738 active events, and 146 scheduled events with many unevenly sampled observations collected over 3 months.

http://511.org/

Slide from: Sheth, Amit. 2014. Transforming Big Data into Smart Data: Deriving Value via harnessing Volume, Variety and Velocity using semantics and Semantic Web. Keynote at 30th IEEE International Conference on Data Engineering (ICDE) 2014.

WHAT

# Smart Data in the context of Big Data

**Big Data**

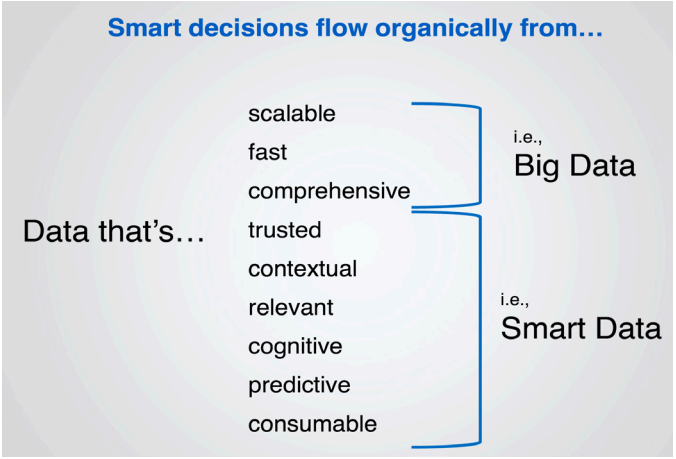
- **V**olume (data quantity)
- **V**elocity (data speed)
- **V**ariety (data types & nature)
- **V**ariability (data consistency)
- **V**eracity (data quality)

**V**alue

**Smart Data**

= The ability to achieve big insights from such data at any scale, great or small.

- trusted
- contextualized
- relevant
- cognitive
- predictive
- consumable




*Kobielus, James.* (2016)

*Source:* Zeng, M.L. 2017 DOI: 10.1515/jdis-2017-0001
*Compiled based on Kobielus, James.* (2016, June). *The Evolution of Big Data to Smart Data*. Keynote at Smart Data Online 2016.

WHAT

- Smart Data aims to filter out the noise and produce valuable data …
- Although unprecedentedly large amount of sensory data can be collected with the advancement of the Cyber-Physical-Social systems, the **key** is to explore **how** Big Data can become Smart Data and offer intelligence.

- SmartData 2019 : The 5th IEEE International Conference on Smart Data. CFP.


Image credit: https://cdait.gatech.edu/events/5th-ieee-international-conference-smart-data

HOW

# HOW to transform Big Data into Smart Data?


## Why Should You Attend?

The Smart Data Conference is designed to accommodate all levels of technical understanding. It will bring together emerging disciplines that are focused on more intelligent information gathering and analysis, including:

- Cognitive Computing
- Deep Learning
- Machine Learning
- Artificial Intelligence
- Predictive Analytics
- Graph Databases
- Machine Intelligence
- Voice Processing
- Semantic Technologies
- Autonomous Vehicles
- Big Data
- Data Science
- Internet of Things
- Text Analysis
- RDF
- Knowledge Graphs
- Contextual Computing
- Linked Data
- Deep Reasoning
- Ontologies
- JSON-LD
- Common Sense
- NLP
- Semantic Search

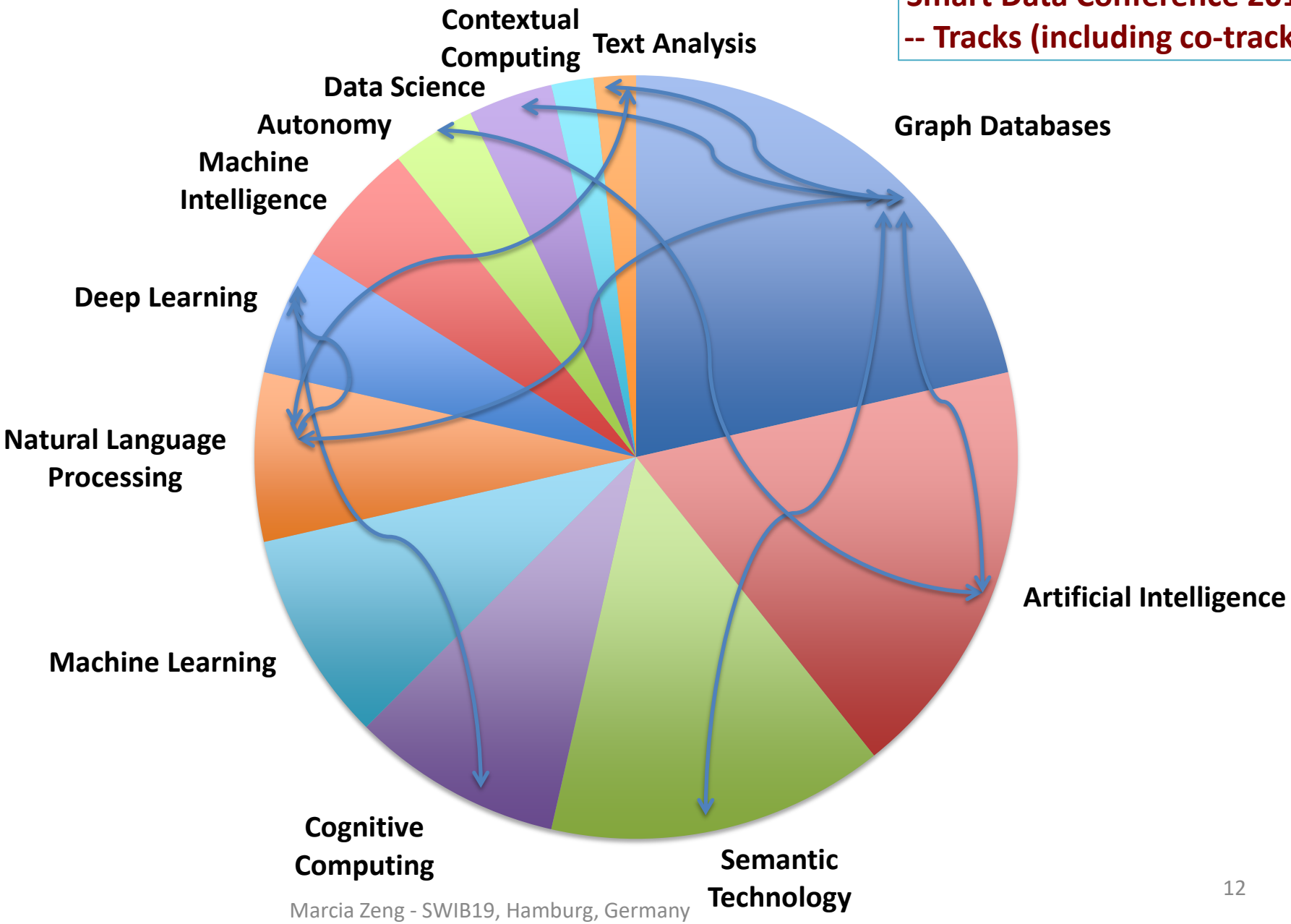


http://smartdata2017.dataversity.net/

HOW

# HOW to transform Big Data into Smart Data?

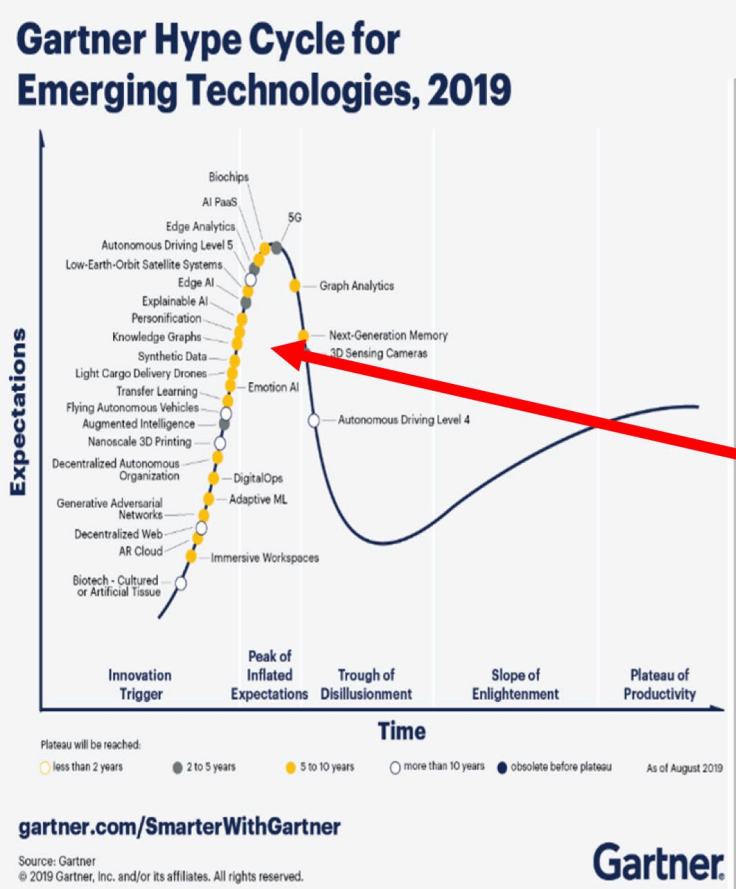
**Smart Data Conference 2017**
**-- Tracks (including co-tracks)**

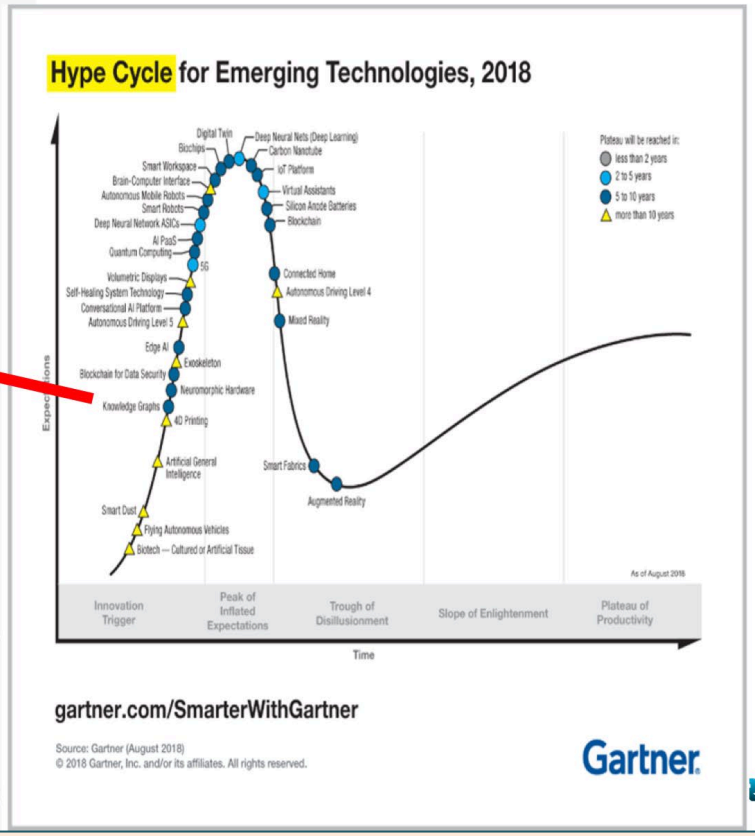


HOW

"The #knowledgegraph--**smart data** that can describe your business and its domains--is now eating software." - Morrison, Alan. Data Centric Design for SMART DATA & CONTENT Enthusiasts meetup, July 31, 2019, Chicago.

# Knowledge Graphs on the rise!

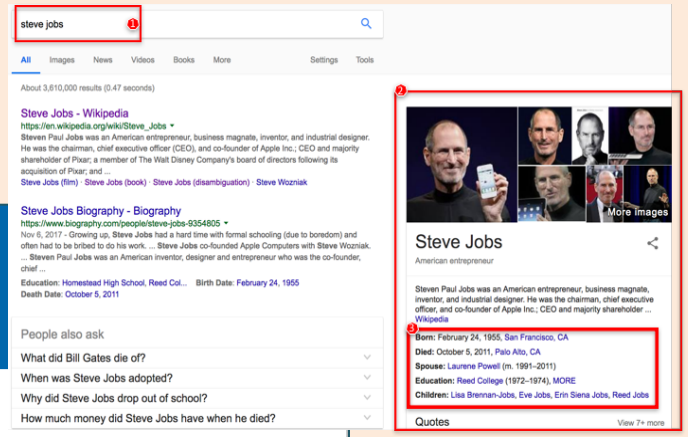



RANZ INC.

Aasman, Jans. Why Knowledge Graphs Now. Ontology Summit, Sept. 4. 2019. Slide 3.

HOW




# What is a Knowledge Graph?

- A system that tries to know and learn everything it can about an entity of interest to improve [internal processes, customer experience, health]
  - Yes, sometimes you can have more types of entities (product+customer, patient-provider)
- It always includes:
  - A (semantic) graph, ontologies, taxonomies
  - Identity management and a smart integration of silos of information
- It nearly always includes
  - Machine Learning, Natural Language Processing, Text classification
- And more and more we see
  - [taxonomy driven] Speech recognition

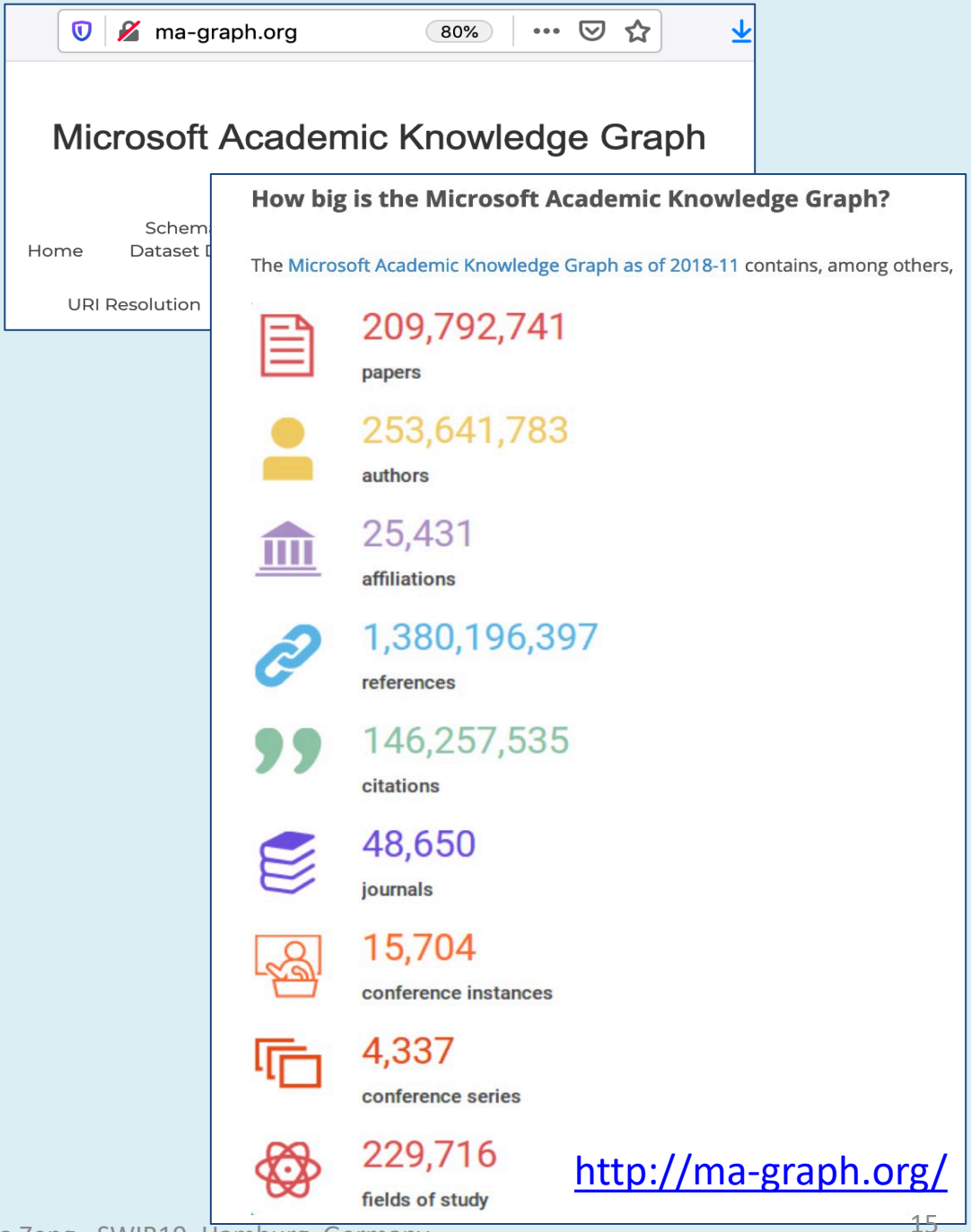
FRANZ INC.

Aasman, Jans . (2019). Why Knowledge Graphs Now. Ontology Summit, Sept. 4. 2019. Slide 36. http://ontologforum.org/index.php/ConferenceCall_2019_09_04

HOW

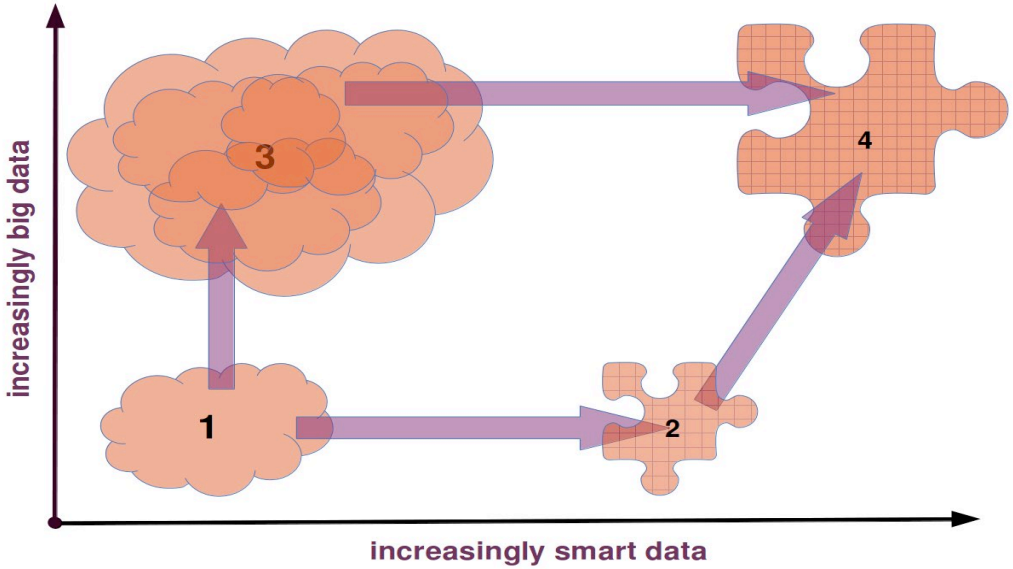
## The known Knowledge Graph builders (in USA)

- Apple
- Amazon.com
- Alphabet
- Microsoft Corp
- Facebook
- JPMorgan Chase
- Bank of America
- …



# What about data in the humanities?

Data has to be cleaned, transformed, and analyzed to unlock its hidden potential.

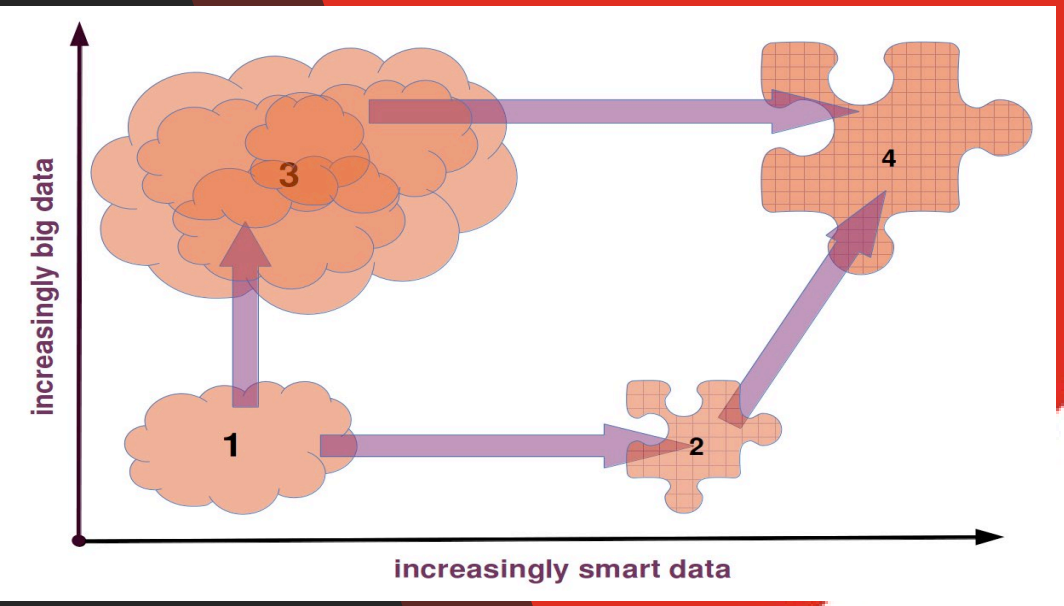



Schöch, Christof. 2013. **Big? Smart? Clean? Messy? Data in the humanities**. *Journal for Digital Humanities.* 2(3): 2-13.

Once tamed through organizing and integrating processes, large volumes of *unstructured, semi-structured,* and *structured data* are turned into "smart data" that reflect the research priorities of a particular discipline or field.

Smart data inquiries can then be used to provide comprehensive analyses and generate new products and services.

# 2. Smart Data in DH




- **WHO** in DH
- **WHICH** Data in DH
- **HOW** in DH

- **WHERE** is the distinctive mark?

# Digital Humanities

- The field is still expanding,
- The definitions are being debated, and
- The multifaceted landscape is yet to be fully understood.

- Most agree that initiatives and activities in digital humanities are at the **intersection** between the **humanities** and **digital information technology**.

- It is at this junction where digital technology will generate a paradigm shift in the near future, enabling scholars to identify major patterns in history, literature, and in the arts.


Image source: Katherine Hayles http://dtc-wsuv.org/wp/dtc375-scodi/katherine-hayles/

Ref:

- Svensson, P. 2010. The Landscape of Digital Humanities. *Digital Humanities Quarterly*. 4(1), 1938-4122
- Svensson, P. 2009. Humanities Computing as Digital Humanities. *Digital Humanities Quarterly.* 3(3), 1938-4122

# WHO in DH


## By Discipline

| | |
|---|---|
| Theory and Methodology of DH: | 97 |
| Arts and Cultural Studies: | 88 |
| Linguistics and Language Studies: | 75 |
| Literary and Philological Studies: | 75 |
| History: | 67 |
| Computer Science: | 62 |
| Human Language Technologies: | 43 |
| Library Science and Information Science: | 39 |
| Media and Communication Studies: | 33 |
| Social Sciences: | 24 |
| Archaeology: | 23 |
| Philosophy: | 13 |
| Other: | 13 |
| Ethnology: | 10 |
| Musicology: | 9 |
| Gender Studies: | 9 |
| Cognitive Science: | 6 |
| Legal Studies: | 6 |
| Theology and Religious Studies: | 5 |

https://registries.clarin-dariah.eu/courses/

*Image captured 2019-07-24.*

**DH refers to new modes of scholarship and institutional units for collaborative, transdisciplinary, and computationally engaged research, teaching, and publication.**


The Digging into Data Challenge (DiD) aims to address how "big data" changes the research landscape for the humanities and social sciences.

https://diggingintodata.org/

| Round / (year) | DiD Funders | DiD Funder countries | Winner # |
|---|---|---|---|
| **Round One /(2009)** | NEH, NSF, SSHRC, Jisc (4) | US, Canada, UK. | 8 |
| **Round Two/ (2011)** | NEH, NSF, SSHRC, Jisc, IMLS, AHRC, ESRC, NWO (8) | US, Canada, UK. Netherlands | 14 |
| **Round Three/ (2013)** | NEH, NSF, SSHRC, Jisc, IMLS, AHRC, ESRC, NWO CFI, NSERC (10) | US, Canada, UK. Netherlands | 14 |
| **Round Four /(2016)**<br><br>Renamed as the "T-AP Digging into Data Challenge" | NEH, NSF, SSHRC, Jisc, IMLS, AHRC, ESRC, NWO CFI, NSERC MINCyT, FAPESP, FRQ AKA, ANR, **DFG**, CONACYT, FCT.(18) | US, Canada, UK. Netherlands, Argentina, Brazil, Finland, France, Germany, Mexico, Portugal | 14 |
| **Total: 4 rounds** | **18 funders** from **11 countries** | | **50 winners** |

WHO in DH

# Domains/Areas of Interests || Resources || Approaches

Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016

DIGGING INTO DATA CHALLENGE

18 funders from 11 countries.

| Domains / Areas of Interests | Resources | Approaches |
|---|---|---|
| • activities in humanities & social science<br>• ancient language<br>• archaeology<br>• biodiversity<br>• child language development<br>• Colonisation of America<br>• comparative and epidemiological paradigm<br>• criminal intent<br>• debating<br>• early modern common placing<br>• economics<br>• English speech<br>• epidemiology<br>• film and media history<br>• financial system<br>• history<br>• human migration<br>• human rights violations<br>• information networks<br>• information patterns and behaviors<br>• journalism<br>• language evolution<br>• legal structures<br>• linguistics<br>• literary networks<br>• manuscripts provenance<br>• music<br>• musicology<br>• parliaments<br>• policy<br>• population<br>• railroad<br>• social sceince<br>• sociological theory<br>• standards of living<br>• storytelling traditions and story repertoires<br>• trading and financial markets<br>• vocabularies | • audio (music) recordings<br>• cuneiform tablets (Mesopotamia )<br>• folklore collections<br>• GDP per capita<br>• geographical data<br>• GitHub<br>• journals<br>• Knowledge Graphs<br>• Knowledge Organization Systems<br>• letters<br>• linguistics databases<br>• manuscripts<br>• manuscripts (pre-modern European)<br>• maps<br>• medical images<br>• medieval charters<br>• multilingual classic text<br>• music info<br>• news about terrorism<br>• newspapers<br>• open access publications<br>• papyrus documents<br>• passages<br>• poetry<br>• population databases<br>• proceedings<br>• quotations<br>• records in indigenous style<br>• records in Spanish<br>• signs<br>• social media<br>• speech datasets<br>• speech recordings<br>• speeches<br>• spoken language collections<br>• tweets, political<br>• video data<br>• writing pieces | • annotation<br>• comparative analysis<br>• computational analysis<br>• computing<br>• corpus building<br>• cross datasets analysis<br>• cross-datasets searching<br>• cross-linguistic annotation<br>• data management<br>• data mining<br>• image processing<br>• indexing<br>• linking<br>• machine coding<br>• machine learning<br>• machine translation<br>• metadata aggregation<br>• metadata analysis<br>• metadata auto-generation<br>• metadata extraction<br>• natural language processing (NLP)<br>• protocols development<br>• spatial-temporal correlation<br>• speech mining<br>• text analysis<br>• visualization |

Source: compiled based on the short descriptions available at https://dev.diggingintodata.org/awards

**Domains/Areas of Interests** || Resources || Approaches
Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016

- activities in humanities & social science
- ancient language
- archaeology
- biodiversity
- child language development
- Colonisation of America
- comparative and epidemiological paradigm
- criminal intent
- debating
- early modern common placing
- economics
- English speech
- epidemiology
- film and media history
- financial system
- history
- human migration
- human rights violations
- information networks
- information patterns and behaviors
- Journalism
- language evolution
- legal structures
- linguistics
- literary networks
- manuscripts provenance
- music
- musicology
- parliaments
- policy
- population
- railroad
- social sceince
- sociological theory
- standards of living
- storytelling traditions and story repertoires
- trading and financial markets
- vocabularies

Source: compiled based on the short descriptions available at https://dev.diggingintodata.org/awards

WHICH data

# "DATA" ≠ "DIGITAL DATA"

Data = "reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing"

for example,

- a sequence of bits,
- a table of numbers,
- the characters on a page,
- the recording of sounds made by a person speaking,
- a moon rock specimen,
- … …

Information = "Any type of knowledge that can be exchanged. In an exchange, it is represented by data."

-- *The Reference Model for an Open Archival Information System (OAIS)*
Consultative Committee for Space Data Systems, 2012, p 1-10 and p 1-12

Data are representations of observations, objects, or other entities used as evidence of phenomena for the purpose of research or scholarship.

- Borgman, 2015, *Big data, little data, no data: Scholarship in the networked world*, p 28.

WHICH data

# – Example from Digital Universe

"Digital Universe" -- a measure of all the digital data created, replicated, and consumed in a single year

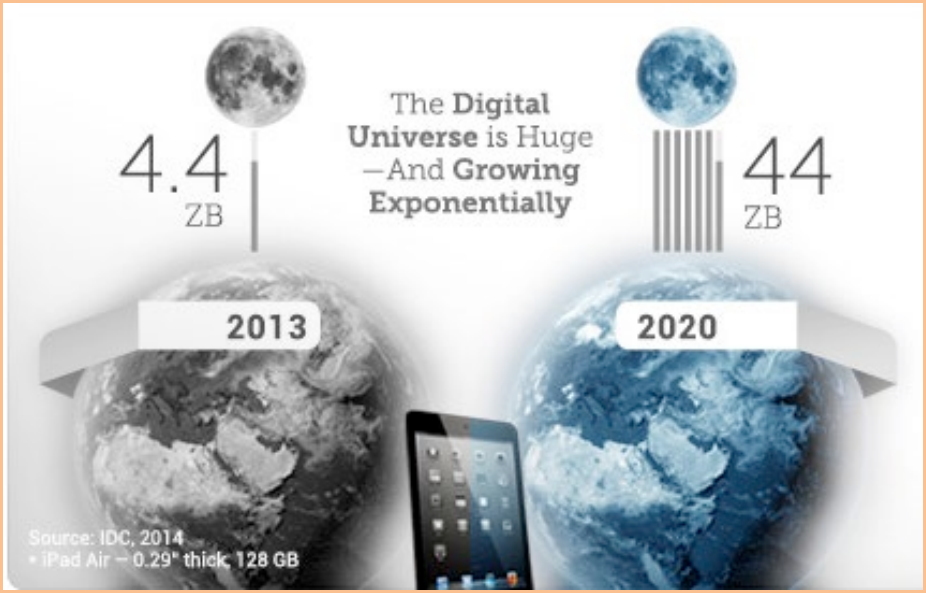



Image credit: http://emergingtechblog.emc.com/wp-content/uploads/2014/04/Digital-Universe1.png

## Data used by the Digital Universe is made up of:

- images and videos on mobile phones uploaded to YouTube,
- digital movies populating the pixels of our high-definition TVs,
- banking data swiped in an ATM,
- security footage at airports and major events such as the Olympic Games,
- subatomic collisions recorded by the Large Hadron Collider at CERN,
- transponders recording highway tolls,
- voice calls zipping through digital phone lines, and
- texting as a widespread means of communications

-- Gantz & Reinsel, 2012.
"The digital universe in 2020", p1.

# WHICH data

Data used by the digital universe is made up of:

- images and videos on mobile phones uploaded to YouTube,
- digital movies populating the pixels of our high-definition TVs,
- banking data swiped in an ATM,
- security footage at airports and major events such as the Olympic Games,
- subatomic collisions recorded by the Large Hadron Collider at CERN,
- transponders recording highway tolls,
- voice calls zipping through digital phone lines, and
- texting as a widespread means of communications

A primary challenge in applying the Smart Data approach to DH:

- ➢ the availability of data resources
  - for those in need of historical data
    - that one <u>could not be</u> obtained through web crawling or scraping.

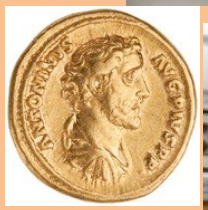
Such a “digital universe”
<u>may not be</u> the <u>major</u> or <u>only source</u>
for humanities researchers.

# WHICH data in DH

Through LAMs: ***unstructured data*** found in documents and other information-bearing objects

- are available in the largest quantity,
- have the most diversity in type, nature, and quality, and
- are the most challenging to process.

The nature of such data is quite different from, e.g., the data used by the "digital universe."


*Revisit*

## Domains/Areas of Interests || Resources || Approaches

Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016

18 funders from 11 countries.

| Domains / Areas of Interests | Resources | Approaches |
|---|---|---|
| • activities in humanities & social science<br>• ancient language<br>• archaeology<br>• biodiversity<br>• child language development<br>• Colonisation of America<br>• comparative and epidemiological paradigm<br>• criminal intent<br>• debating<br>• early modern common placing<br>• economics<br>• English speech<br>• epidemiology<br>• film and media history<br>• financial system<br>• history<br>• human migration<br>• human rights violations<br>• information networks<br>• information patterns and behaviors<br>• journalism<br>• language evolution<br>• legal structures<br>• linguistics<br>• literary networks<br>• manuscripts provenance<br>• music<br>• musicology<br>• parliaments<br>• policy<br>• population<br>• railroad<br>• social sceince<br>• sociological theory<br>• standards of living<br>• storytelling traditions and story repertoires<br>• trading and financial markets<br>• vocabularies | • audio (music) recordings<br>• cuneiform tablets (Mesopotamia )<br>• folklore collections<br>• GDP per capita<br>• geographical data<br>• GitHub<br>• journals<br>• Knowledge Graphs<br>• Knowledge Organization Systems<br>• letters<br>• linguistics databases<br>• manuscripts<br>• manuscripts (pre-modern European)<br>• maps<br>• medical images<br>• medieval charters<br>• multilingual classic text<br>• music info<br>• news about terrorism<br>• newspapers<br>• open access publications<br>• papyrus documents<br>• passages<br>• poetry<br>• population databases<br>• proceedings<br>• quotations<br>• records in indigenous style<br>• records in Spanish<br>• signs<br>• social media<br>• speech datasets<br>• speech recordings<br>• speeches<br>• spoken language collections<br>• tweets, political<br>• video data<br>• writing pieces | • annotation<br>• comparative analysis<br>• computational analysis<br>• computing<br>• corpus building<br>• cross datasets analysis<br>• cross-datasets searching<br>• cross-linguistic annotation<br>• data management<br>• data mining<br>• image processing<br>• indexing<br>• linking<br>• machine coding<br>• machine learning<br>• machine translation<br>• metadata aggregation<br>• metadata analysis<br>• metadata auto-generation<br>• metadata extraction<br>• natural language processing (NLP)<br>• protocols development<br>• spatial-temporal correlation<br>• speech mining<br>• text analysis<br>• visualization |

Source: compiled based on the short descriptions available at https://dev.diggingintodata.org/awards

# WHICH data in DH

Domains/Areas of Interests || **Resources** || Approaches Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016

Source: compiled based on the short descriptions available at https://dev.diggingintodata.org/awards

## Resources

- audio (music) recordings
- cuneiform tablets (Mesopotamia )
- folklore collections
- GDP per capita
- geographical data
- GitHub
- journals
- Knowledge Graphs
- Knowledge Organization Systems (KOS)
- letters
- linguistics databases
- manuscripts
- manuscripts (pre-modern European)
- maps
- medical images
- medieval charters
- multilingual classic text
- music info
- news about terrorism
- newspapers
- open access publications
- papyrus documents
- passages
- poetry
- population databases
- proceedings
- quotations
- records in indigenous style
- records in Spanish
- signs
- social media
- speech datasets
- speech recordings
- speeches
- spoken language collections
- tweets, political
- video data
- vocabularies
- writing pieces

**HOW** in DH

Domains/Areas of Interests || Resources || **Approaches** Expressed in the Project Descriptions of *Digging into Data Challenge* Round 1-4, 2009-2016

| Domains / Areas of Interests | Resources | Approaches |
|---|---|---|
| • activities in humanities & social science<br>• ancient language<br>• archaeology<br>• biodiversity<br>• child language development<br>• Colonisation of America<br>• comparative and epidemiological paradigm<br>• criminal intent<br>• debating<br>• early modern common placing<br>• economics<br>• English speech<br>• epidemiology<br>• film and media history<br>• financial system<br>• history<br>• human migration<br>• human rights violations<br>• information networks<br>• information patterns and behaviors<br>• jonrnalism<br>• language evolution<br>• legal structures<br>• linguistics<br>• literary networks<br>• manuscripts provenance<br>• music<br>• musicology<br>• parliaments<br>• policy<br>• population<br>• railroad<br>• social sceince<br>• sociological theory<br>• standards of living<br>• storytelling traditions and story repertoires<br>• trading and financial markets<br>• vocabularies | • audio (music) recordings<br>• cuneiform tablets (Mesopotamia )<br>• folklore collections<br>• GDP per capita<br>• geographical data<br>• Github<br>• journals<br>• Knowledge Graphs<br>• Knowledge Organization Systems<br>• letters<br>• linguistics databases<br>• manuscripts<br>• manuscripts (pre-modern European)<br>• maps<br>• medical images<br>• medieval charters<br>• multilingual classic text<br>• music info<br>• news about terrorism<br>• newspapers<br>• open access publications<br>• papyrus documents<br>• passages<br>• poetry<br>• population databases<br>• proceedings<br>• quotations<br>• records in indigenous style<br>• records in Spanish<br>• signs<br>• social media<br>• speech datasets<br>• speech recordings<br>• speeches<br>• spoken language collections<br>• tweets, political<br>• video data<br>• writing pieces | • annotation<br>• comparative analysis<br>• computational analysis<br>• computing<br>• corpus building<br>• cross datasets analysis<br>• cross-datasets searching<br>• cross-linguistic annotation<br>• data management<br>• data mining<br>• image processing<br>• indexing<br>• linking<br>• machine coding<br>• machine learning<br>• machine translation<br>• metadata aggregation<br>• metadata analysis<br>• metadata auto-generation<br>• metadata extraction<br>• natural language processing (NLP)<br>• protocols development<br>• special-temporal correlation<br>• speech mining<br>• text analysis<br>• visualization |

Source: compiled based on the short descriptions available at https://dev.diggingintodata.org/awards

## Approaches


- annotation
- comparative analysis
- computational analysis
- computing
- corpus building
- cross-datasets analysis
- cross-datasets searching
- cross-linguistic annotation
- data management
- data mining
- image processing
- indexing
- linking
- machine coding
- machine learning
- machine translation
- metadata aggregation
- metadata analysis
- metadata auto-generation
- metadata extraction
- natural language processing (NLP)
- protocols development
- spatial-temporal correlation
- speech mining
- text analysis
- visualization

# "COMPLEXITIES"

## - *DIGITAL HUMANITIES 2019* theme

- "This theme has a multifaceted connection with *Digital Humanities* scholarship."
  - The Alliance of Digital Humanities Organizations (ADHO) CFP
- DH2019, July 2019, Utrecht University, Netherlands
  - 1000+ delegates from 50+ countries;
  - participants are from all kinds of disciplines and professions.

Let's look at three cases, from this year, and from different kinds of reports: article, movie, and museum exhibition.

# Cases

- A recent article: "Ancient Rome: A genetic crossroads of Europe and the Mediterranean."
  Antonio, Margaret L., et al. ***Science*** 366.6466 (2019): 708-714.
  https://science.sciencemag.org/content/366/6466/708
- A recent documentary movie: **The Violence Paradox**, NOVA. Nov.20, 2019 https://www.pbs.org/wgbh/nova/video/the-violence-paradox/
- A recent exhibition **"Leonardo da Vinci: 500 Years of Genius"**
  at Denver Museum of Nature & Science, March-Aug., 2019.
  https://grandeexhibitions.com/the-secrets-of-mona-lisa/#explore
  & NOVA Documentary: "Decoding Da Vince"
  https://www.pbs.org/wgbh/nova/video/decoding-da-vinci/

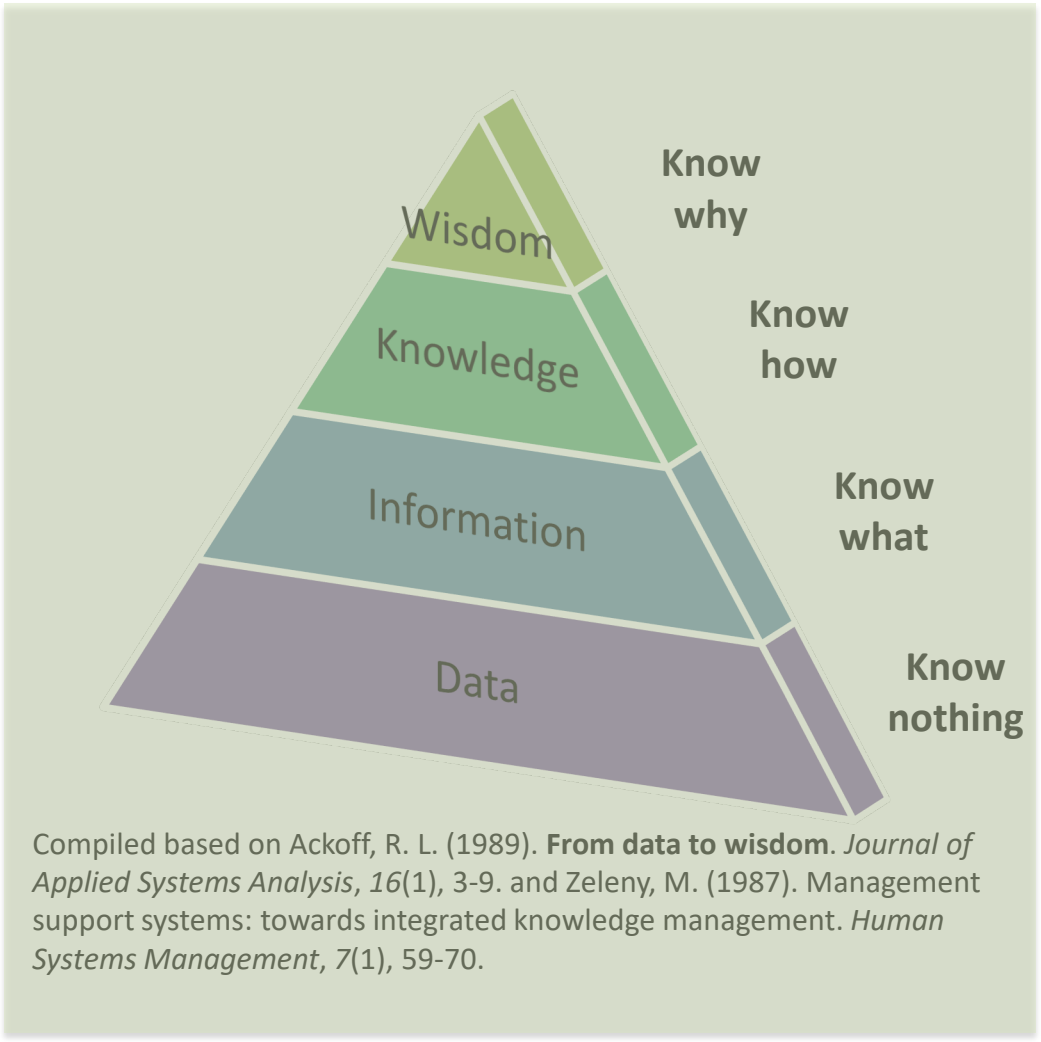
# Smart Data in DH

## WHERE is the distinctive mark?

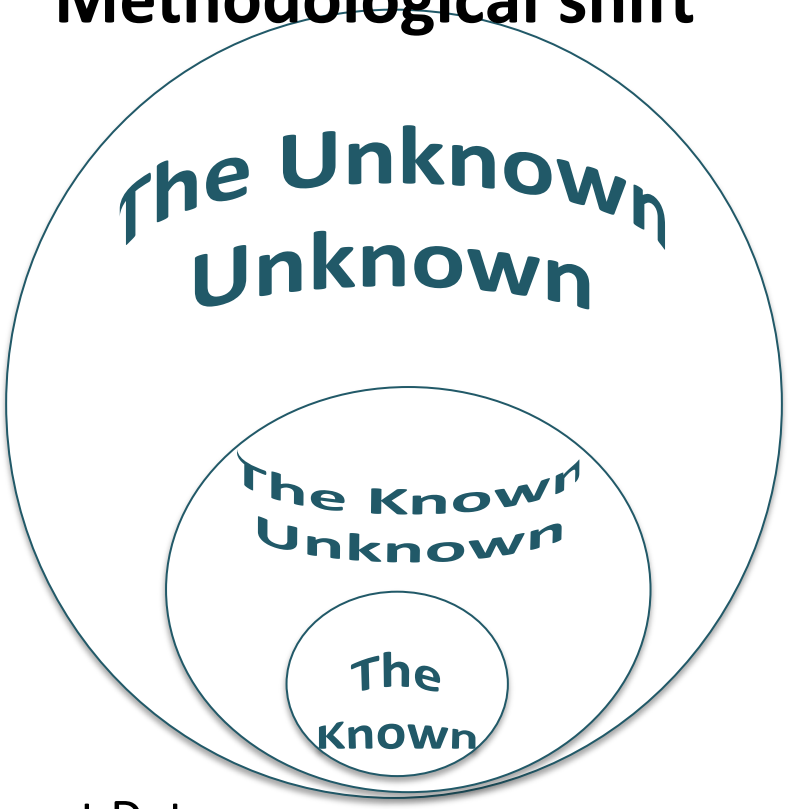
Technological shift?

Methodological shift?

# Methodological shift




Compiled based on Ackoff, R. L. (1989). **From data to wisdom**. *Journal of Applied Systems Analysis*, *16*(1), 3-9. and Zeleny, M. (1987). Management support systems: towards integrated knowledge management. *Human Systems Management*, *7*(1), 59-70.

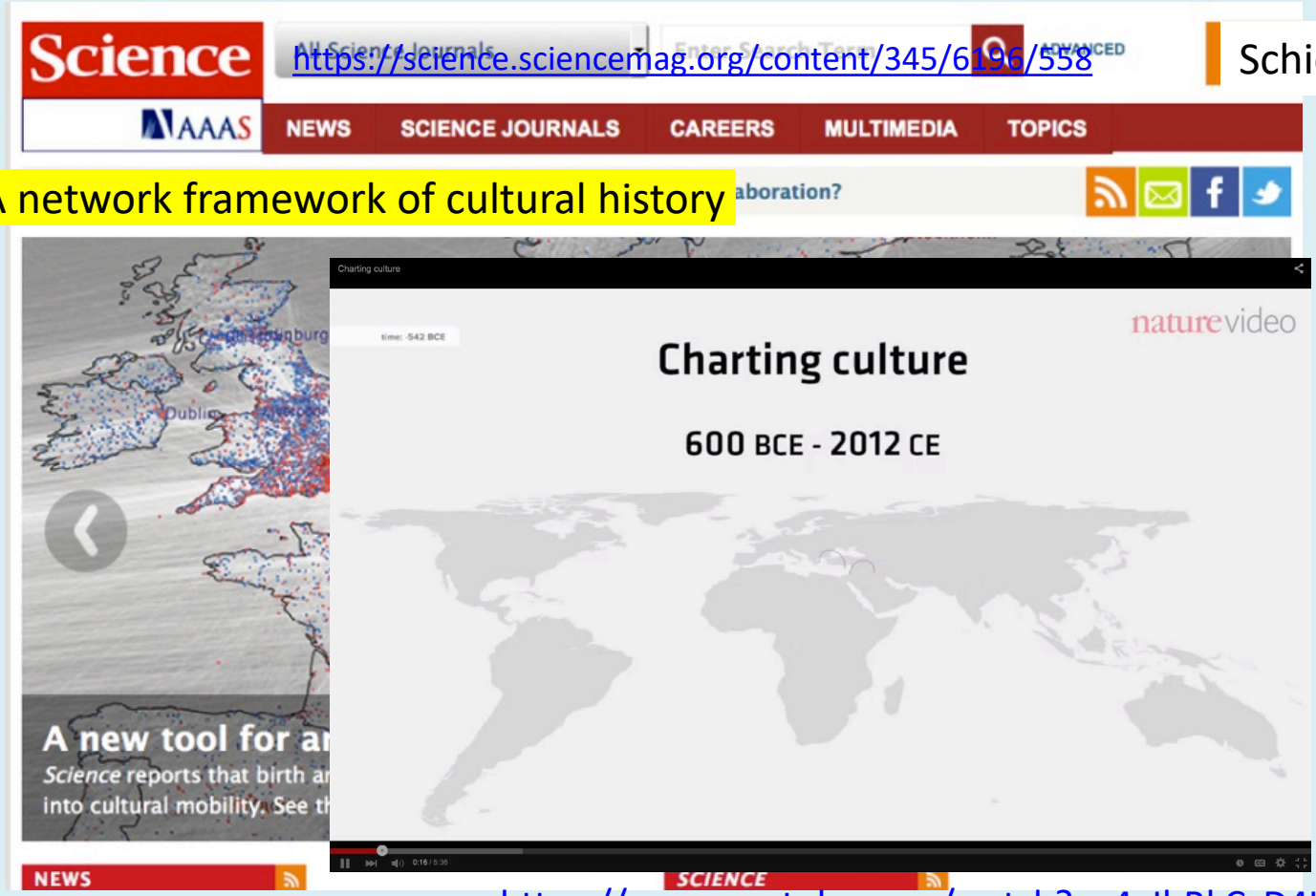



Smart Data

- based on Big Data's methodology
- assumes the ability to reveal the *unknown-unknowns*
- instead of taking the approach that one knows to do something in order to prove or disapprove the *known-unknowns*.

**Let's see another example and think:**

- Why does *Science* magazine publish something on humanities?
- Why is the video "Charting culture" provided by *Nature*?
- What kind of data did the project use?




https://science.sciencemag.org/content/345/6196/558

Schich *et al.*, 2014a, 2014b

Birth and death data of notable individuals reveal interactions between culturally relevant locations over two millennia.

*see also* video

https://www.youtube.com/watch?v=4gIhRkCcD4U

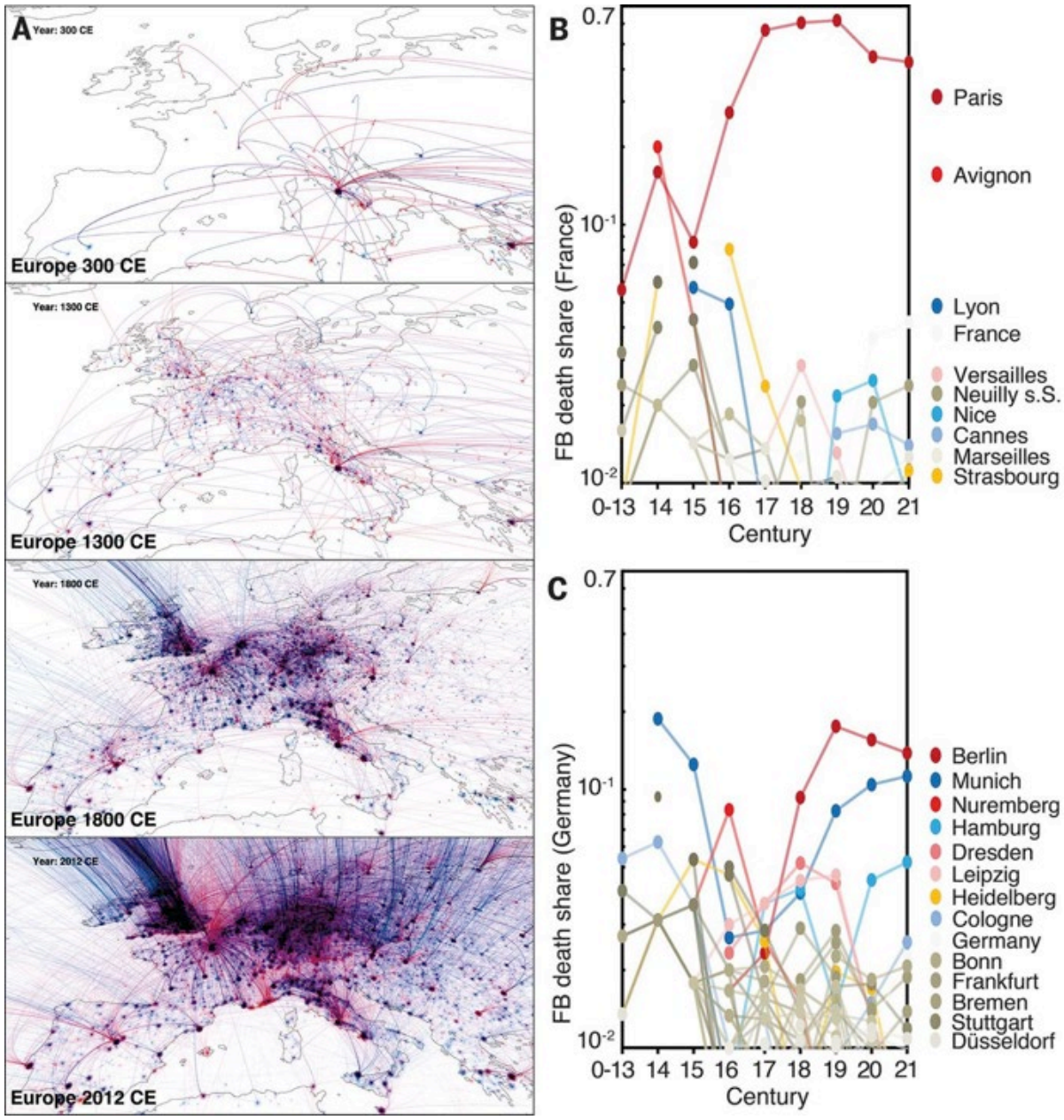


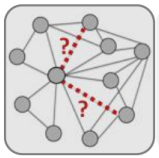

**Fig. 3**
**The visualization of birth-death network dynamics offers a meta-narrative of cultural history.**

Schich *et al.*, 2014a

http://science.sciencemag.org/content/345/6196/558.full

# Discovery is about the non-obvious

Discovery creates new knowledge (hitherto unknown / unrecorded in domain docs, data sources). New knowledge is surprising and anomalous. It could be formally abstracted in several forms:



**New link between entities**: A new side effect of a drug, a new potential emerging company as an acquisition target or sales lead, a non-obvious but relevant case / law applicable in a particular trial / proceeding, a person of interest in a terrorist attack. (Link Prediction, Relationship Discovery, Relationship Ranking)

**A potential new important entity in the domain**: e.g. a new material for display technologies, a new investor for a particular investment area etc. (Entity Discovery, Entity Recommendation, Entity Ranking)

**Changing significance of an existing entity:** it's changing relationships/attributes/metrics. e.g. increasing stake by an investor in an organization, increasing interaction between a person of interest and some criminal in an intelligence gathering scenario, decreasing number complaints for a particular product or service in a retail / consumer scenario. (Trend analysis, Distribution Analysis, Anomaly Detection)

Search and Exploration look for knowledge already available in the knowledge sources available to the system. They are necessary for Discovery but not sufficient

Confirmation of facts known to an SME of a domain is not discovery (listing all possible side effects of a drug which may have been mined from structured or unstructured data)

Taylor, Jamie. (Google - Knowledge Graph Schema Team.) 2018. Knowledge Graph. Panel "Enterprise-scale knowledge graphs". International Semantic Web Conference (ISWC) 2018.

Revisit:

# Methodological shift

Smart Data

- is based on Big Data's methodology that
- assumes the ability to reveal the *unknown-unknowns*
- instead of taking the approach that one knows to do something in order to prove or disapprove the *known-unknowns*.


Turning Big Data into Smart Data

(Cont.) 3. Smart Data for DH
→in the context
of
LAMs' data

**HOW to support DH**

1) providing LAM data
2) serving humanities research

Two major types of challenges

# Which process is the most challenging and difficult?

**Deciding the scope of research data.**

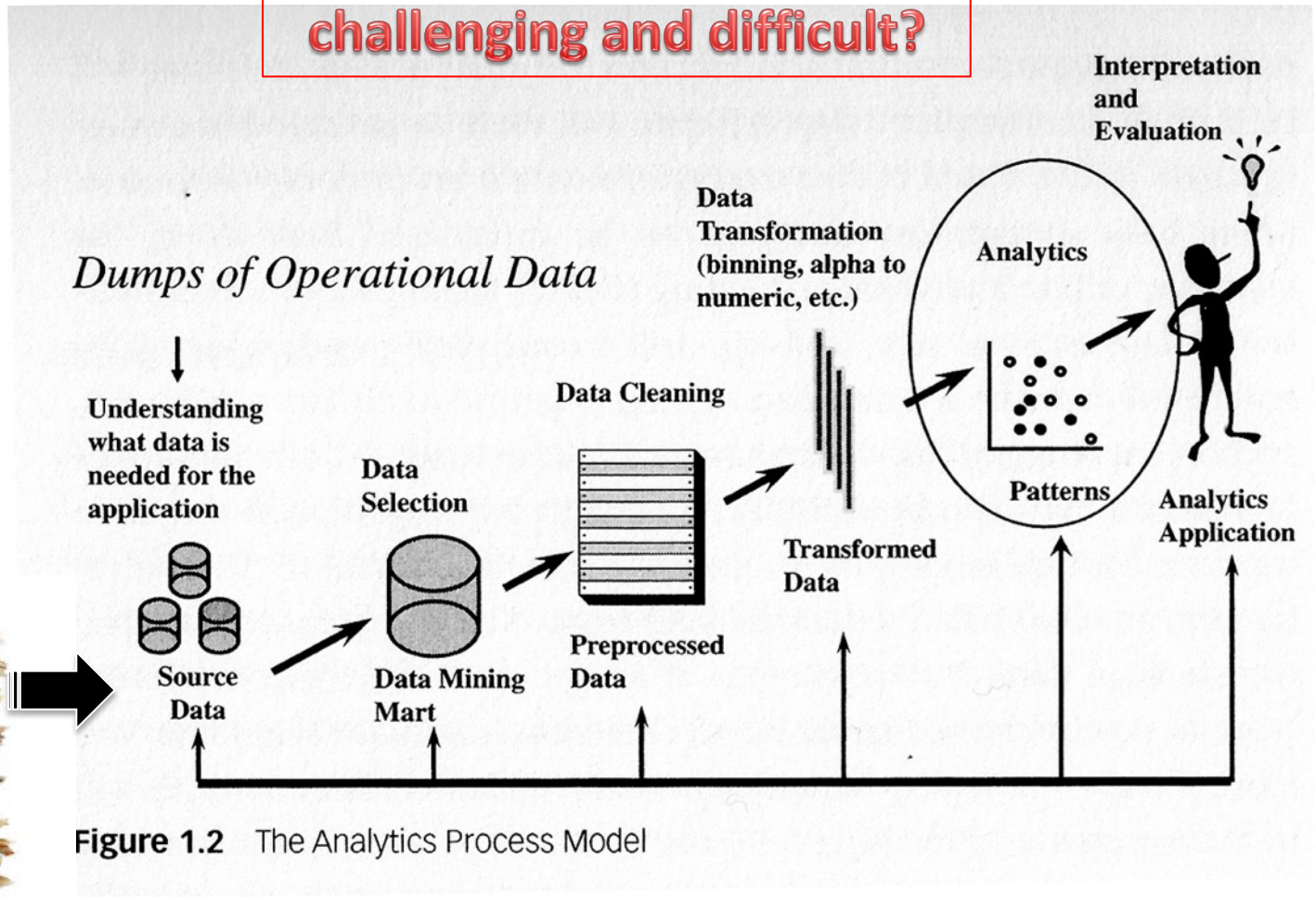
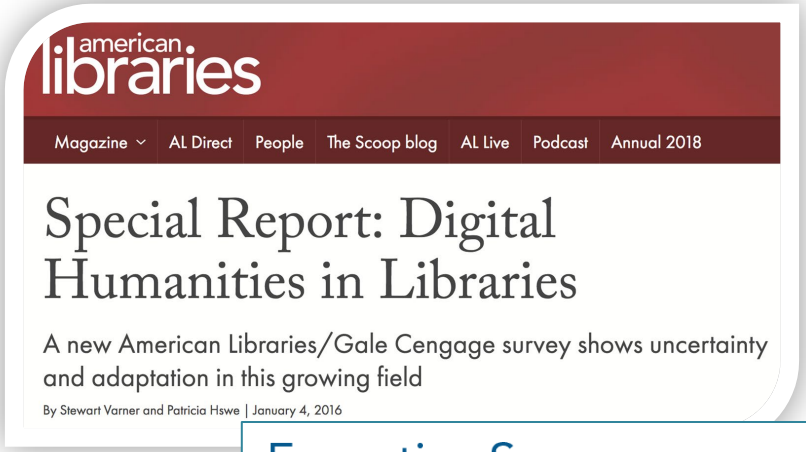



**Figure 1.2** The Analytics Process Model

Figure image credit: Bart Baesens, 2014. Analytics in a Big Data World: The Essential Guide to Data Science and its Applications, Figure 1.2.



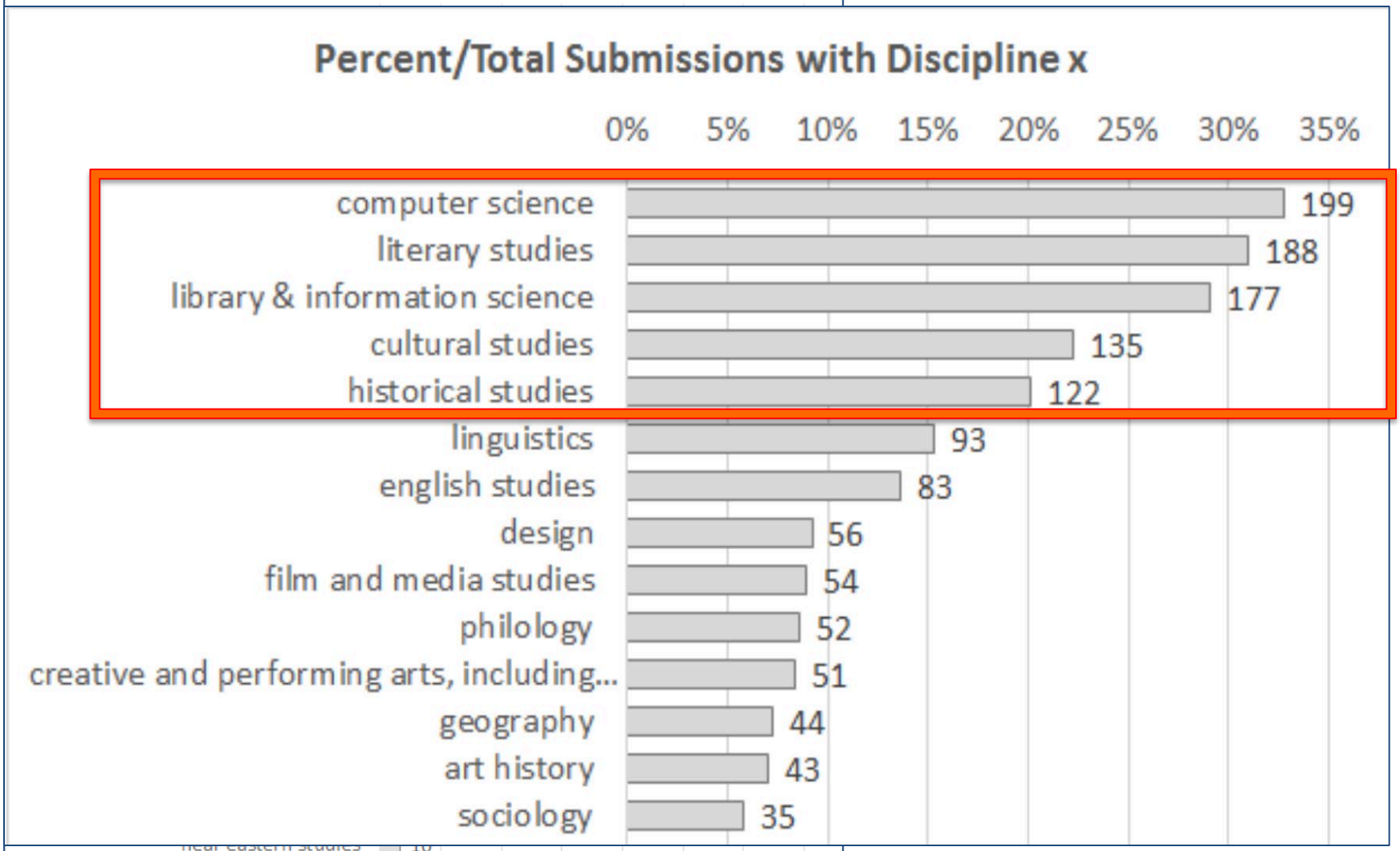

# Digital Humanities

Faculty Survey Results – December 2015

## Executive Summary

- Almost all faculty think that support in digital humanities elevates the importance of academic libraries.
  - Six out of ten faculty feel that a digital humanities center belongs in the library digital collections center.
  - In addition to having an institutional repository, over half of faculty also saw the role of the library to include advocating for coordinated digital support, packaging existing services together, or creating avenues for scholarly use of metadata.
  - Further, in their own work, over half also saw the role of the library to include general support, being a liaison for existing library services, providing training on tools, or helping find available sources.
  - The main reasons faculty felt that it elevates the importance included seeing the library as more of a digital center, renewing the library as a central place for research, and demonstrating the value of the skills of the librarian.

- Varner, S. & Hswe, P. 2016. "Special report: Digital humanities in libraries." *American Libraries* (2016-01-04).

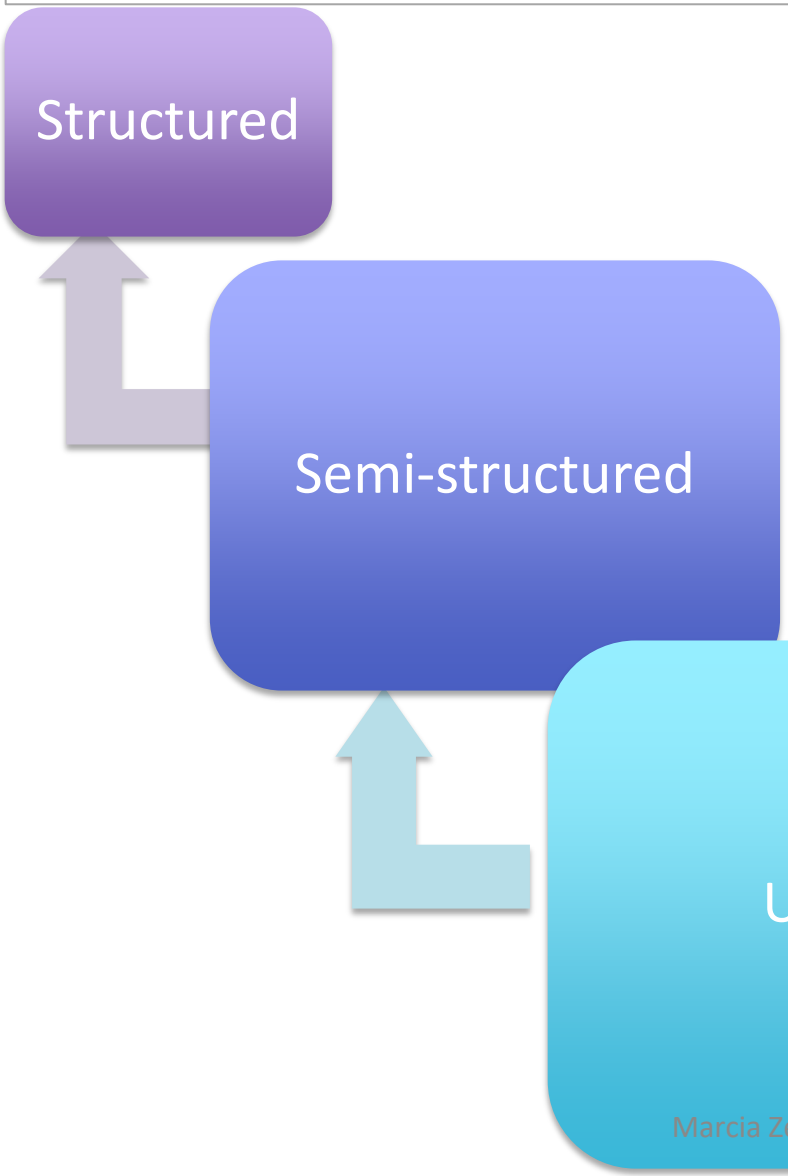



translation studies 9
theology 9
law 8
italian studies 8
disability studies 6

*Source:* Weingart, S. (2017). Submissions to DH2017 (pt.1) [Web log post]. http://scottbot.net/submissions-to-dh2017-pt-1/

# Data resources usually served through LAMs

## LAM Data Examples



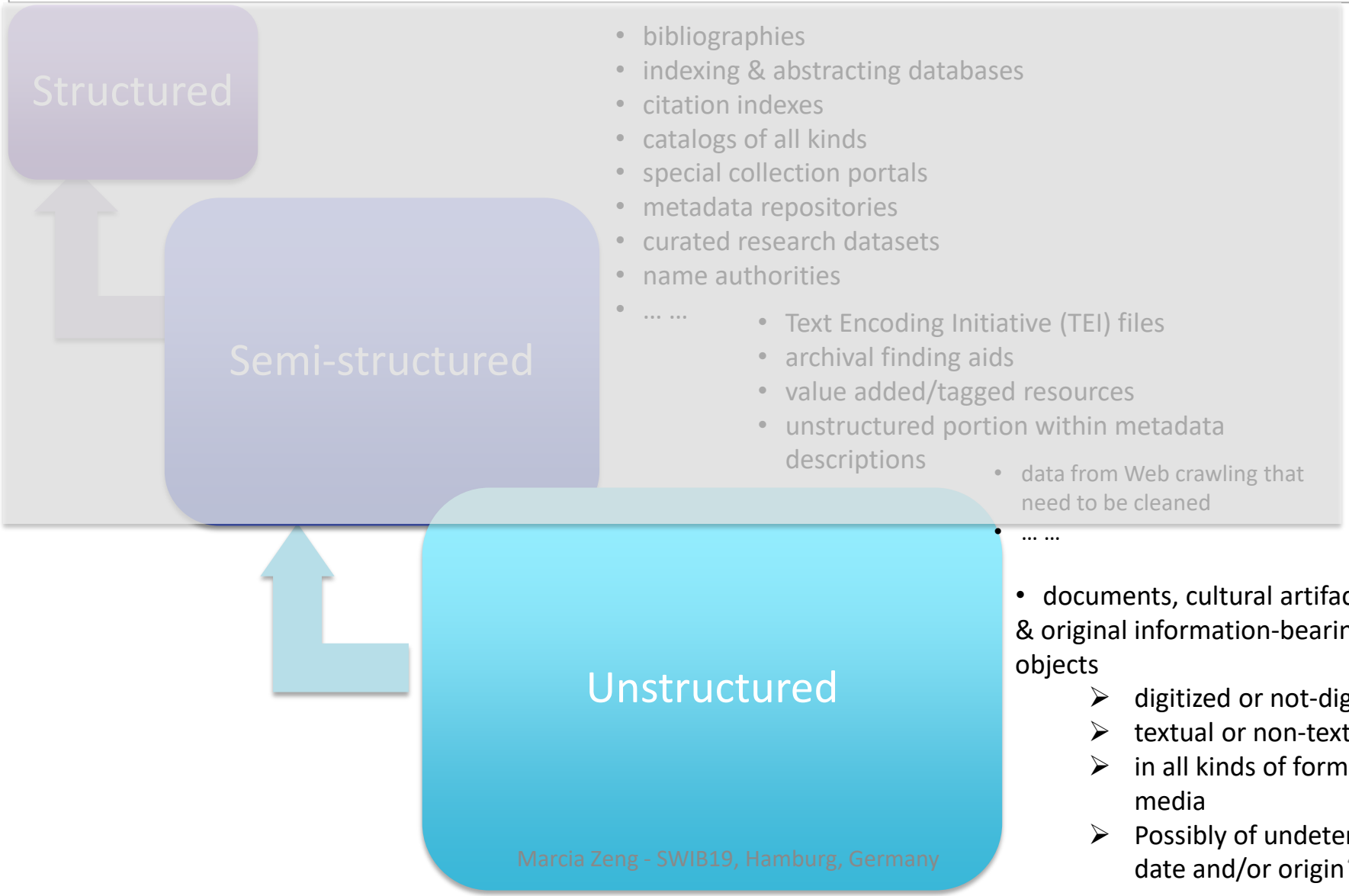

- bibliographies
- indexing & abstracting databases
- citation indexes
- catalogs of all kinds
- special collection portals
- metadata repositories
- curated research datasets
- name authorities
- … …

- Text Encoding Initiative (TEI) files
- archival finding aids
- value added/tagged resources
- unstructured portion within metadata descriptions
- data from Web crawling that need to be cleaned
- … …

- documents, cultural artifacts, & original information-bearing objects
  - ➢ digitized or not-digitized
  - ➢ textual or non-textual
  - ➢ in all kinds of formats and media
  - ➢ Possibly of undetermined date and/or origin

# Data resources usually served through LAMs

## LAM Data Examples




- bibliographies
- indexing & abstracting databases
- citation indexes
- catalogs of all kinds
- special collection portals
- metadata repositories
- curated research datasets
- name authorities
- … …

- Text Encoding Initiative (TEI) files
- archival finding aids
- value added/tagged resources
- unstructured portion within metadata descriptions

- data from Web crawling that need to be cleaned
- … …

- documents, cultural artifacts, & original information-bearing objects
  - digitized or not-digitized
  - textual or non-textual
  - in all kinds of formats and media
  - Possibly of undetermined date and/or origin

# Stage 1. Digitization & Documentation

Image credit: Wikipedia

E.g., 2002 -- 2011, US IMLS Grants to States program

- supplied $980 million
- to support increased access to digital information,
- including $67 million toward the digitization of local history and special collections.

**Smart Data + DH** approach emphasizes:

- Transforming <u>unstructured data</u>
  to → <u>structured data</u> & <u>semi-structured data</u>

# Three Perspectives
## -- from the **creation of the structured data**

1 Production

- **Documentation / Metadata**
  - ***Descriptive*** *metadata*
    - IS-ness
    - Categories/collection/division
  - ***Administrative*** *metadata*
    - 'Bio' of the object
      - Possession
      - Communication history
      - Preservation/updating
    - Assessment
    - Legal documentation
  - ***Structural/technical*** *metadata*
    - Size
    - Components, part-whole
    - Order
    - File formats

Content 2

**Metadata**
- ***Descriptive*** *metadata*
  - OF-ness
  - ABOUT-ness
- **Relationships**

**Index**

**Markup** (embedded)

**Knowledge graph**

**Ontology (=Knowledge base)**

Onsite | Online (Web + Mobile)

3 **Audiences' receiving interests**

- ***Use*** **Metadata**

*cited, researched, tagged, searched, shared, followed, time spent, liked …*

# Stage 2. Datafication

**Structured Data Trending:**

- Machine ~~readable~~ understandable data
- Machine ~~readable~~ actionable data
- Accurate (no error) data in the processes of interlinking, citing, transferring, rights-permission, use and reuse
- One → to → Many uses and high efficiency processing data

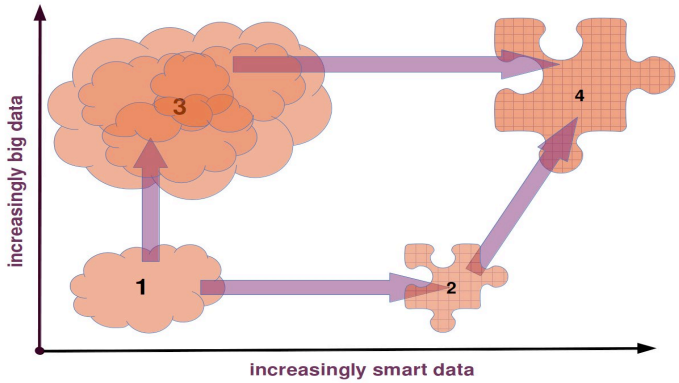



Image: Christof Schöch, 2013

Refer to specific cases at: Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. *El profesional de la información*, 28(1), e280103. https://doi.org/10.3145/epi.2019.ene.03

# Stage 3. Contextualization

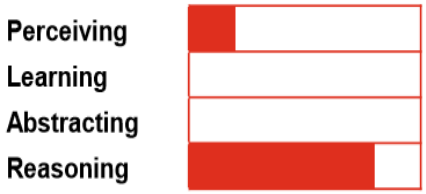
## The key opportunity – Large-scale integration and model-driven intelligence in a de-siloed and de-duplicated way

**Previously dominant**

**Rule-based systems (includes KR)**

Handcrafted knowledge" is the term DARPA uses; rule-based programming + procedure replication in process automation, + some knowledge representation (KR)

- Strong on logical reasoning in specific concrete contexts
  - Procedural + declarative programming + set theory, etc.
  - Deterministic
- Can't learn or abstract
- Still exceptionally common and useful




**Example:** Consumer tax software

+

**On the rise and rapidly improving**

**Statistical machine learning**

- Probabilistic
- From Bayesian algorithms to neural nets (yes, deep learning also)
- Strong on perceiving and learning (classifying, predicting)
- Weak on abstracting and reasoning
- Quite powerful in the aggregate but individually (instance by instance) unreliable
- Can require lots of data


**Example:** Facial recognition using deep learning/neural nets

=

**Nascent, just beginning**

**Contextualized, model-driven approach**

- Contextualized modeling approach-allows efficiency, precision and certainty
- Combines power of deterministic, probabilistic and description logic
- Allows explanations to be added to decisions
- Accelerates the training process with the help of specific, contextual human input
- Takes less data


Example: Explains first how handwritten letters are formed so machines can decide-less data needed, more transparency.

John Launchbury of DARPA (https://www.youtube.com/watch?v=N2L8AqkEDLs), Estes Park Group and PwC research, 2017



Morrison, Alan. (2019). Data-centric design and the knowledge graph. Slide 15.

# Smart Data in the context of Big Data

**Big Data**

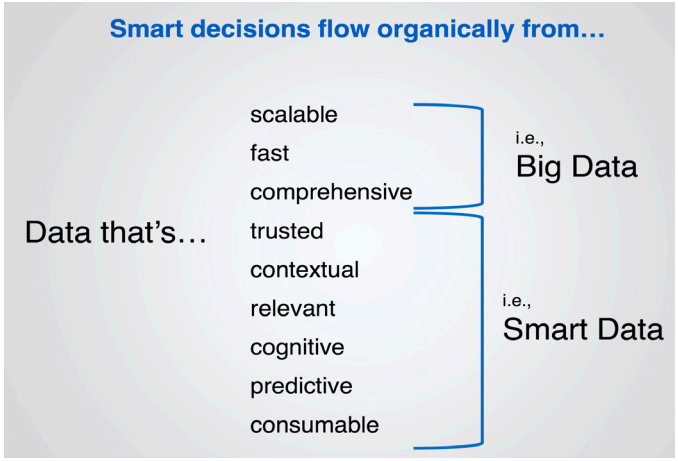
**V**olume (data quantity)
**V**elocity (data speed)
**V**ariety (data types & nature)
**V**ariability (data consistency)
**V**eracity (data quality)

**V**alue

**Smart Data**

= The ability to achieve big insights from such data at any scale, great or small.

- trusted
- contextualized
- relevant
- cognitive
- predictive
- consumable




*Kobielus, James.* (2016)

*Source:* Zeng, M.L. 2017 DOI: 10.1515/jdis-2017-0001
*Compiled based on Kobielus, James.* (2016, June). *The Evolution of Big Data to Smart Data*. Keynote at Smart Data Online 2016.

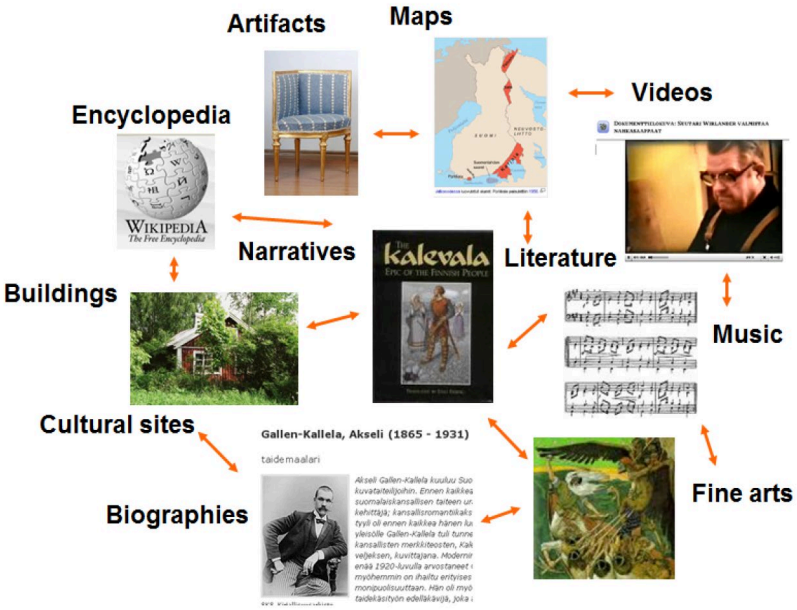
(Cont.) 3. Smart Data for DH
→in the context
of
LAMs' data

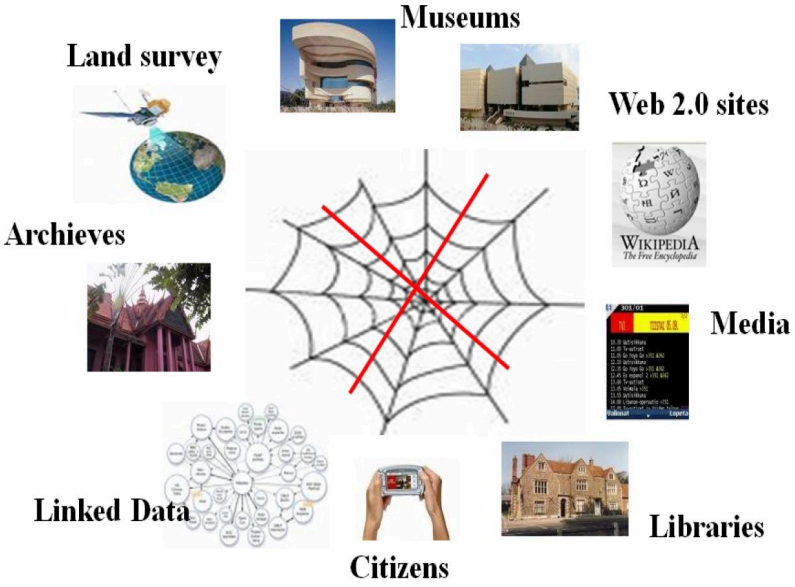
**HOW to support DH**

1) providing LAM data
2) serving humanities research

Two major types of challenges

# a) Challenges for publishing Cultural Heritage (CH) collections on the Web

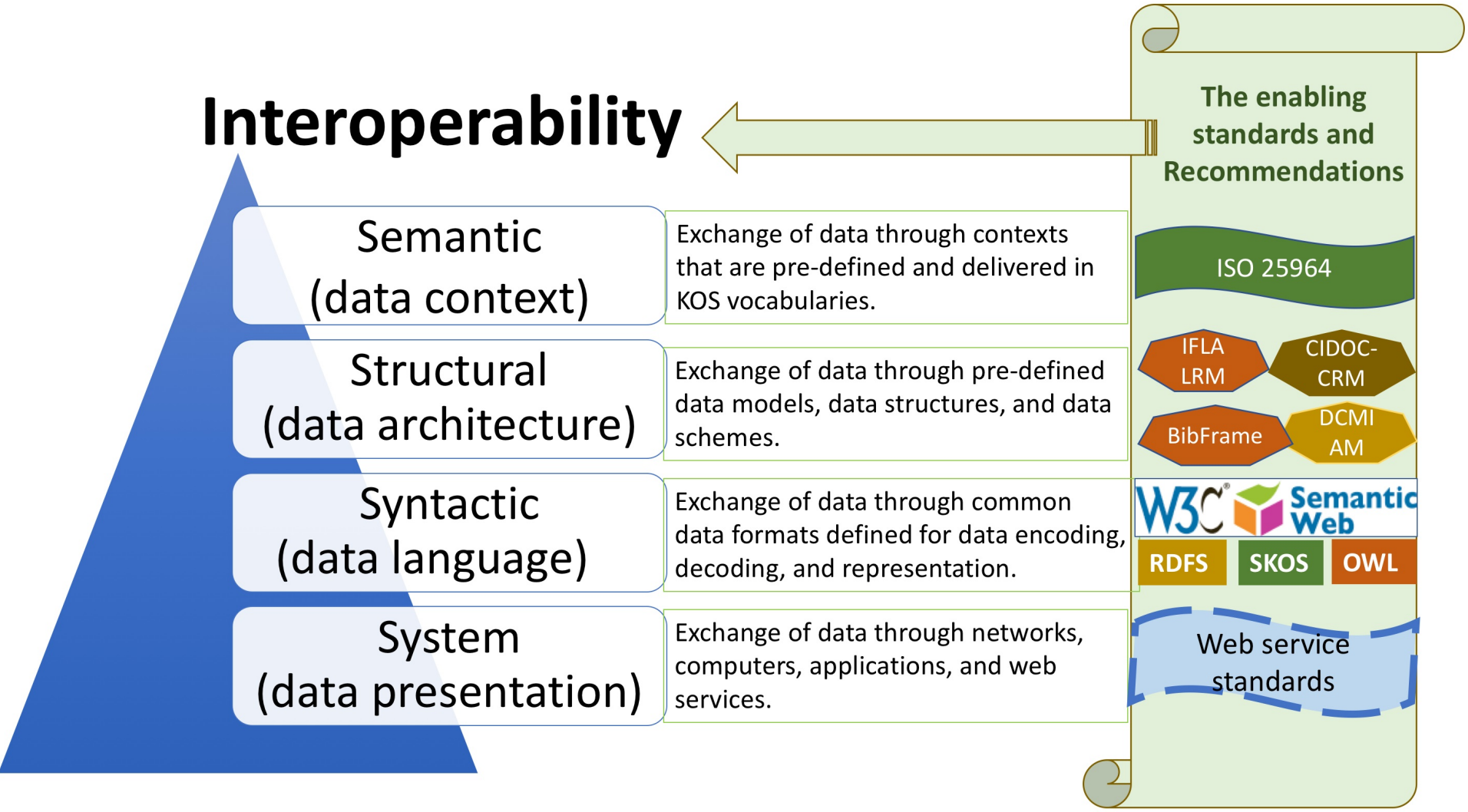



HOW TO make the heterogeneous CH contents semantically interoperable … across the data silos → WHERE the standards and best practices are not compatible with each other

- Hyvönen, Eero. 2016. “Cultural Heritage Linked Data on the SemanticWeb: Three Case Studies Using the Sampo Model” VIII Encounter of Documentation Centres of Contemporary Art. Artium, Vitoria-Gasteiz, Spain, October 19-20, 2016.

# Interoperability





Zeng, Marcia L. 2018. Interoperability. *ISKO Encyclopedia of Knowledge Organization (IEKO).* http://www.isko.org/cyclo/interoperability

Zeng, Marcia L. *SEMANTIC INTEROPERABILITY - Notes and bookmarks Bookmarks*. http://metadataetc.org/LOD/interoperability.html

# Approach: Sampo Model

a semantic model based on a shared ontology infrastructure in the middle


"Sampo" is a mythical machine or thing in Finnish folklore that gives you everything, but nobody really knows what it is.

. It includes mutually aligned metadata and shared domain ontologies, modeled using SW standards.


- Hyvönen, Eero. 2016. "Cultural Heritage Linked Data on the SemanticWeb: Three Case Studies Using the Sampo Model" VIII Encounter of Documentation Centres of Contemporary Art. Artium, Vitoria-Gasteiz, Spain, October 19-20, 2016.
- Hyvönen, Eero. [2019]. Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery. *Semantic Web*. [in press]

# Example:
## The Sampo model and the portals

## WarSampo

https://www.sotasampo.fi/en/

### Data

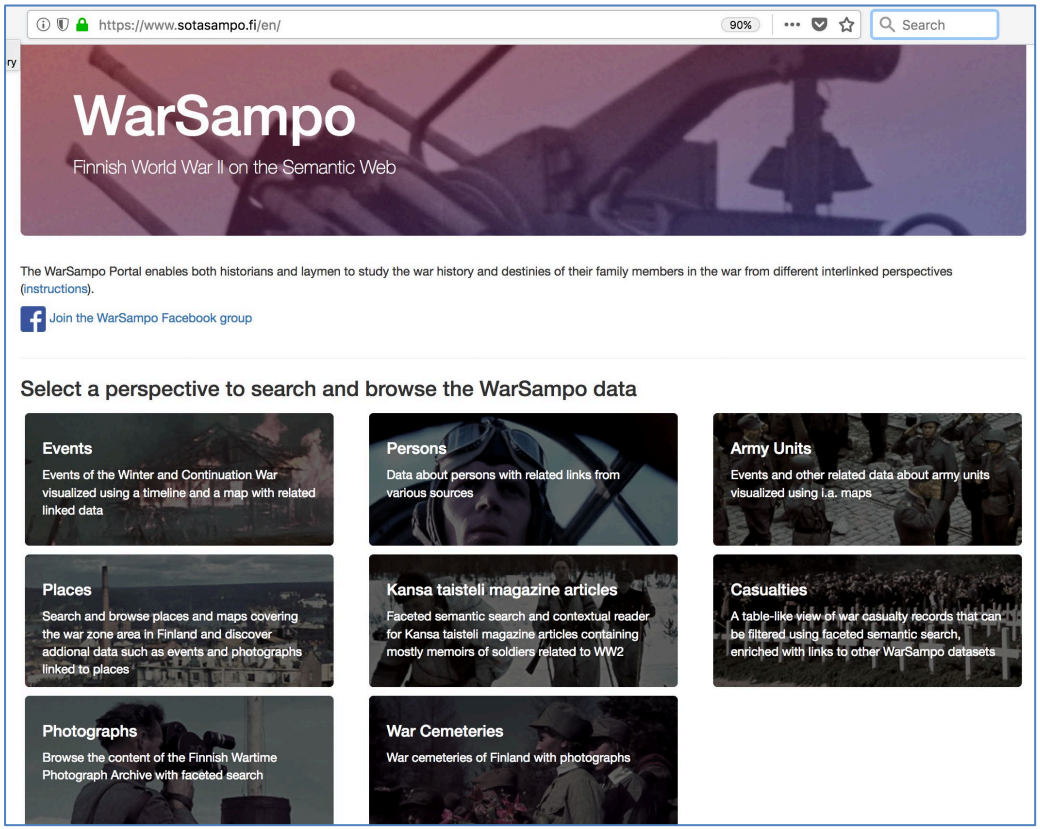
- 8 datasets from different organizations
- 7.6 million triples in a SPARQL endpoint

**Ontologies and knowledge bases**

- the **troops** and their hierarchies,
- **persons** with their ranks and roles,
- **place** -- historical places,
- **event** -- battles, politics, and other war time incidents,
- **time** periods,
- **weapons**,
- **vessels**, and
- a **subject matter** ontology.

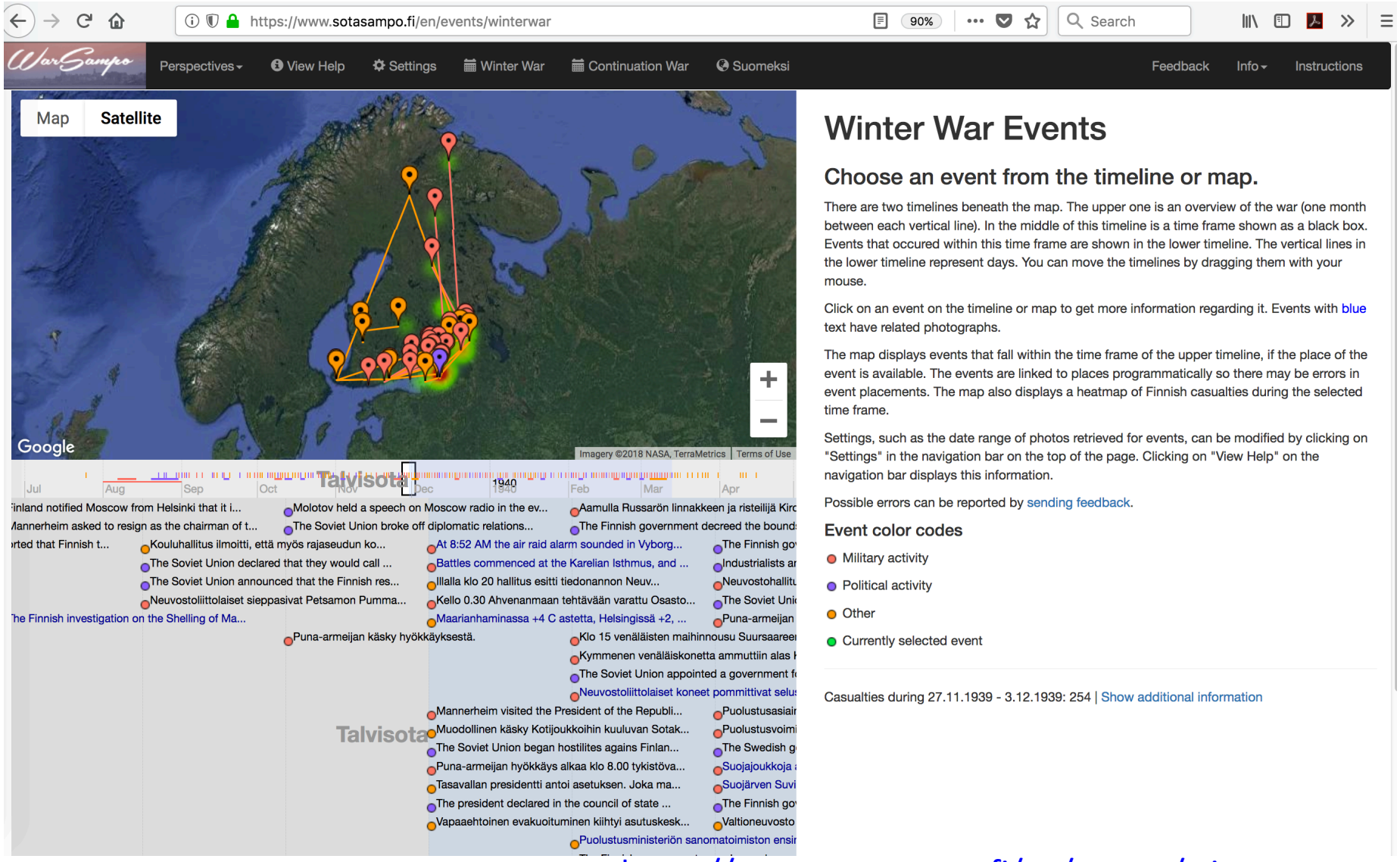
https://seco.cs.aalto.fi/publications/2016/hyvonen-et-al-warsa-dh2016.pdf

**Portal**




- Published in November 2015
  - over 10,000 distinct visitors during the first 3 days
- WarSampo is one of the two winners at the 2017 LODLAM Challenge

# WarSampo




https://www.sotasampo.fi/en/events/winterwar

# Example:

## The Sampo model and the portals

### BiographySampo

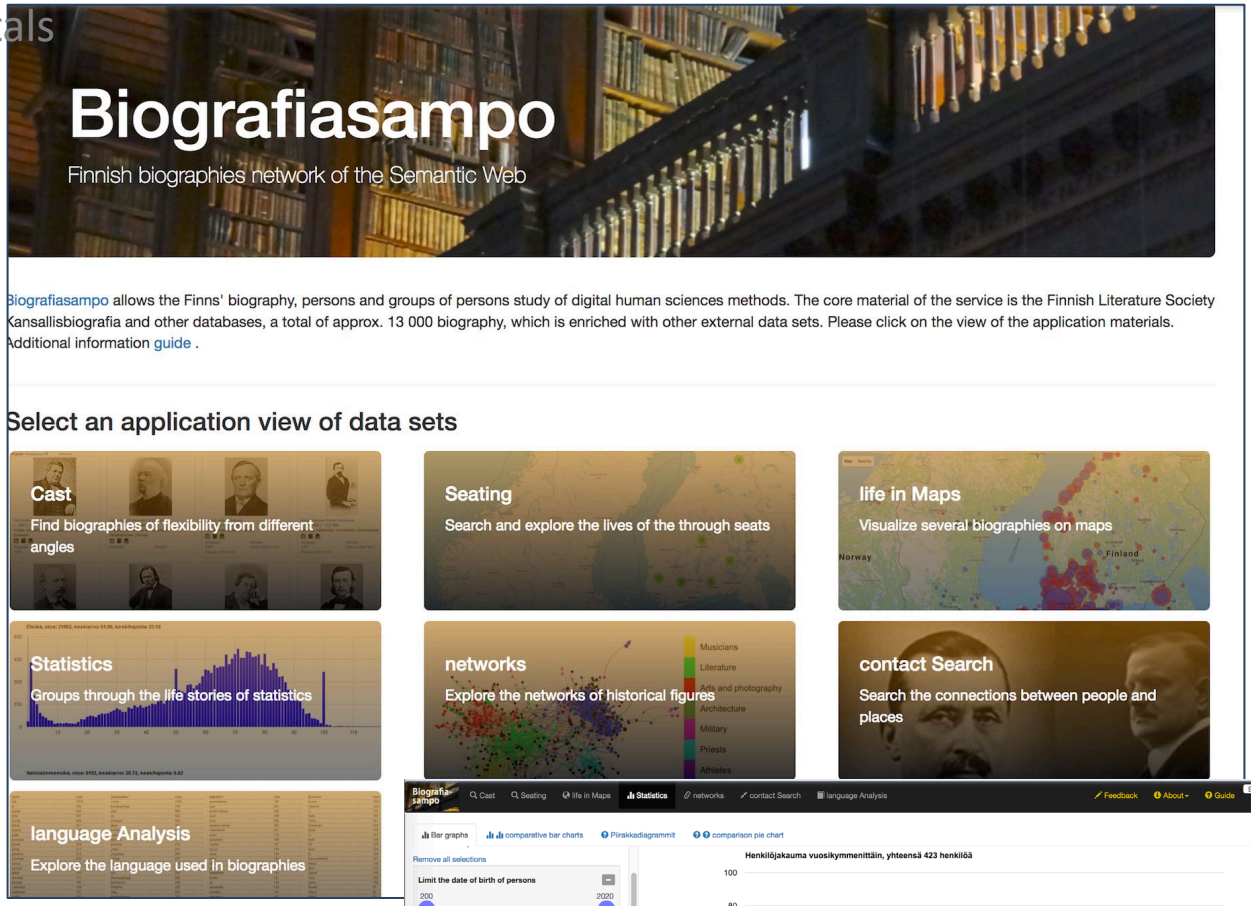
http://biografiasampo.fi/

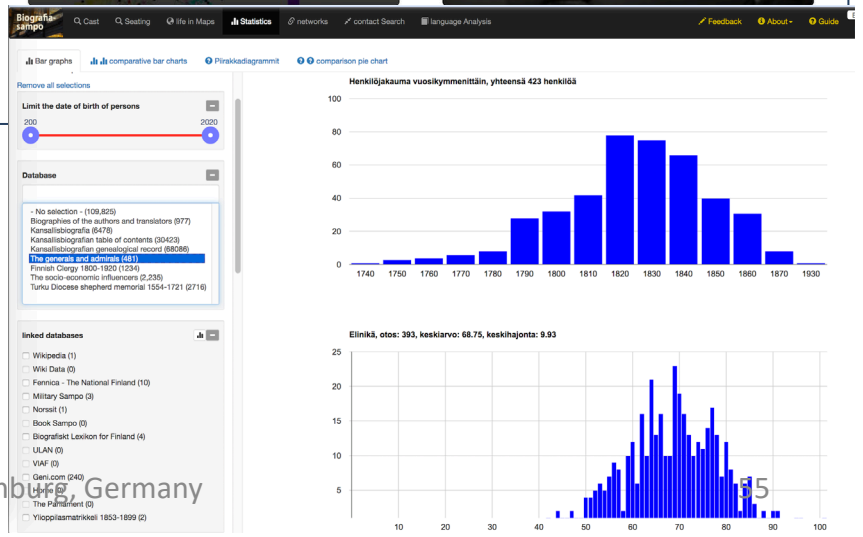
Finnish National Biographies on the Semantic Web

Based on extracting knowledge graphs from 13 000 biographies (multiple language choices.)

**View by:**

- Case
- Seating
- Life in maps
- Statistics
- Networks
- Contact search
- Language analysis




Video: “ BiographySampo - Artificial Intelligence Reading Biographies for the Semantic Web”
https://vimeo.com/328419960

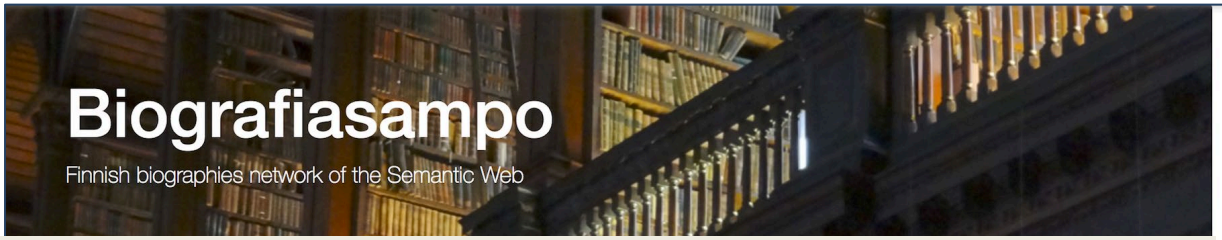



Fig. Comparing the life charts of two target groups, admirals and generals (left) and clergy (right) of the historical Grand Duchy of Finland (1809–1917).

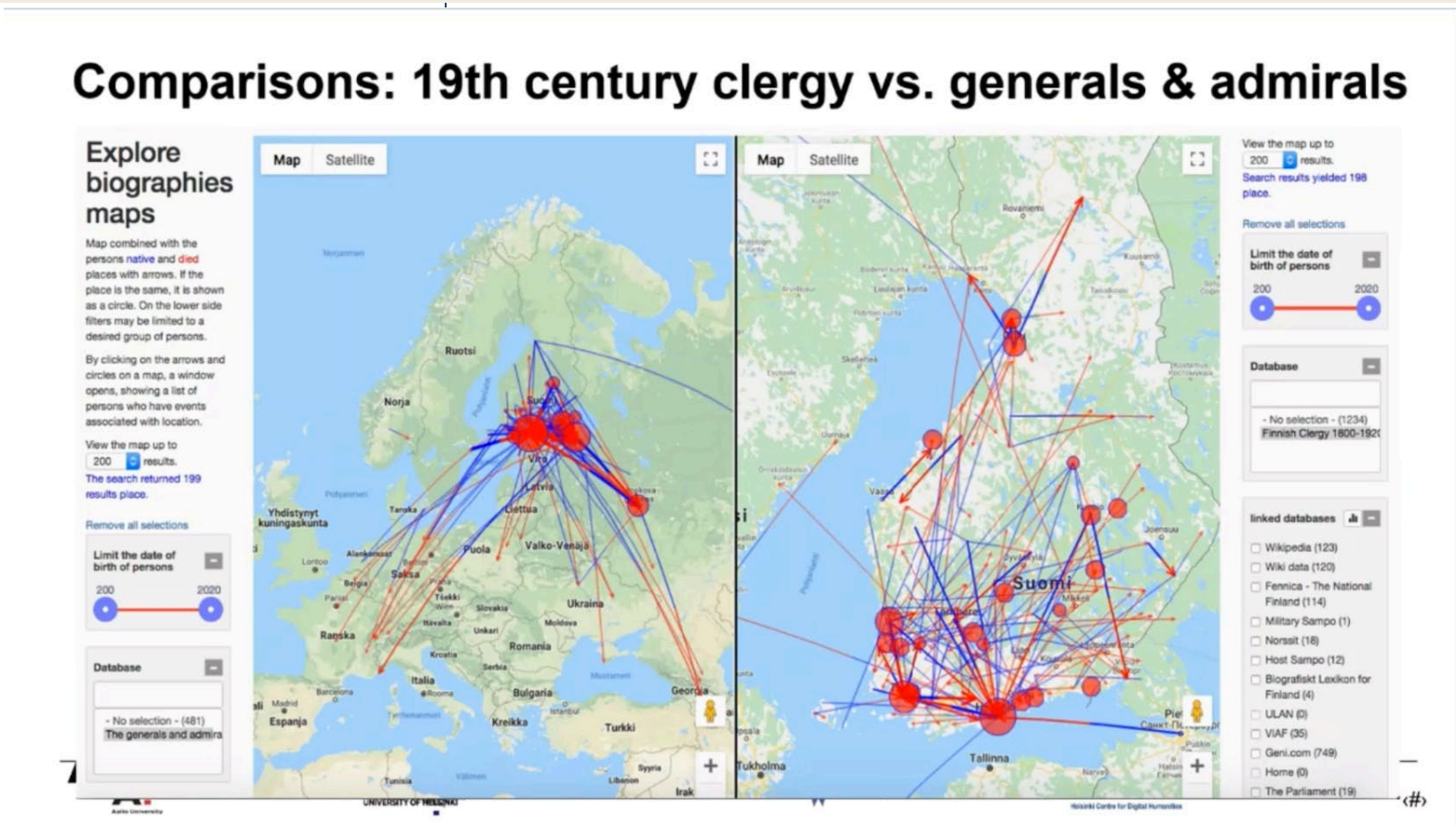



Image source: BiographySampo - Artificial Intelligence Reading Biographies for the Semantic Web. https://vimeo.com/328419960

# Example:

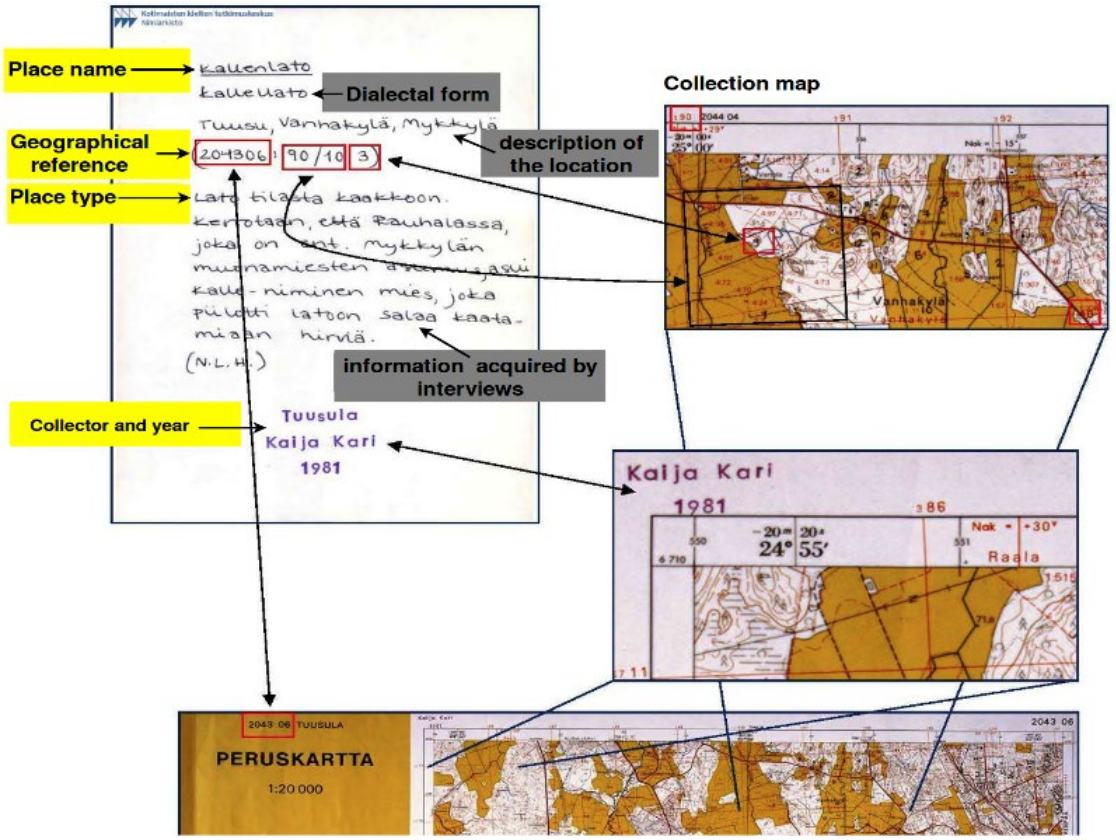
## The Sampo model and the portals

**NameSampo**
http://nimisampo.fi

- A Linked Open Data Infrastructure and Workbench for Toponomastic Research

• Based on over 2 million places names collected in Finland and beyond.

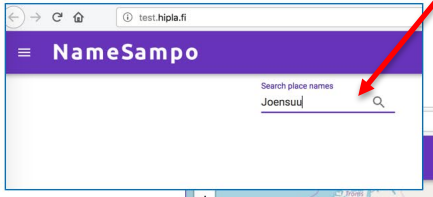
•The objective of the project was to convert the place name entry slips, collection maps, and the attribute and metadata related to them into digital format.



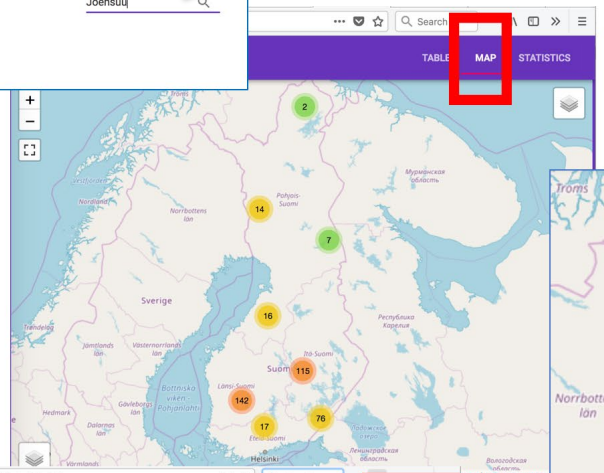

https://seco.cs.aalto.fi/projects/nimisampo/

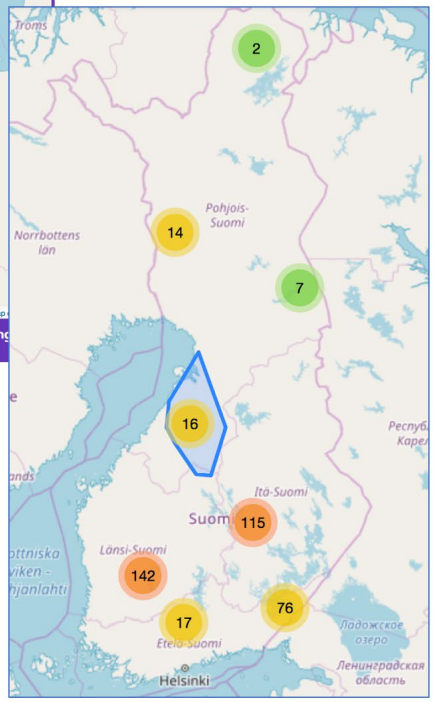
The service at http://nimisampo.fi got 9500 users during the first weekend.

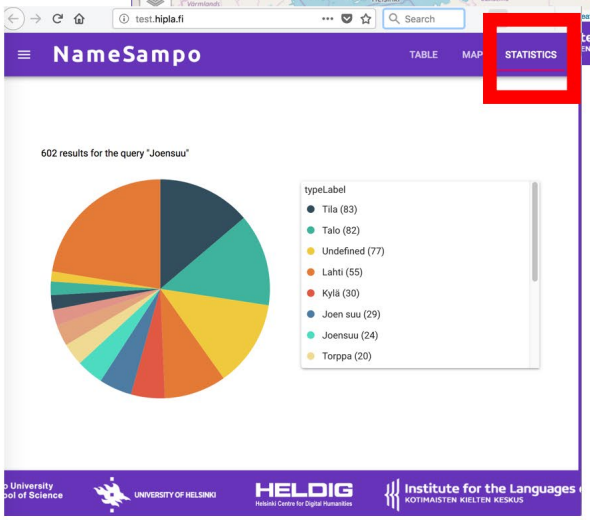
# Checking the place: Joensuu



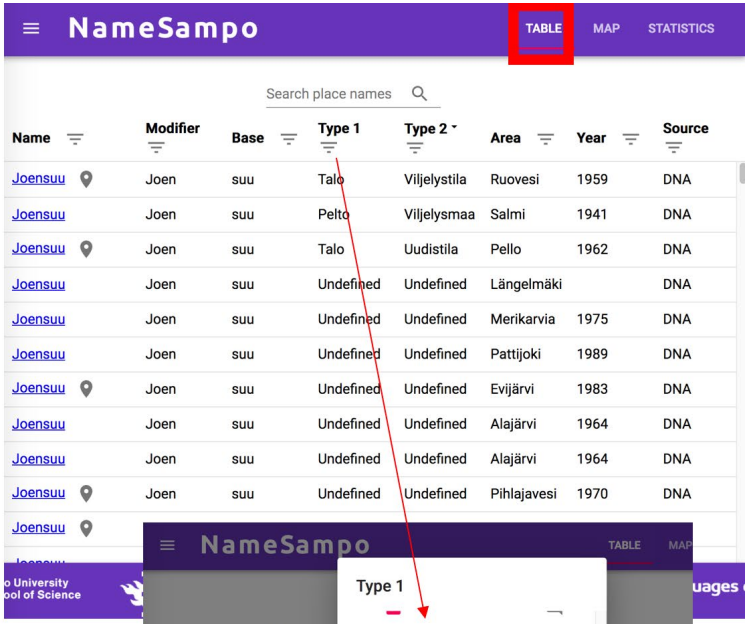

http://test.hipla.fi/

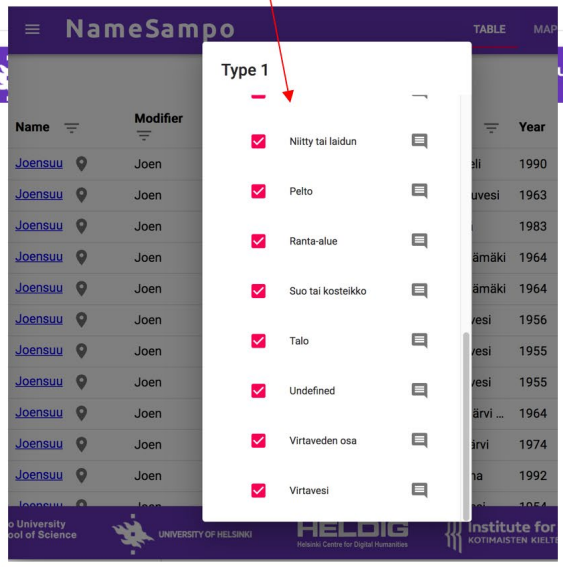



http://test.hipla.fi/



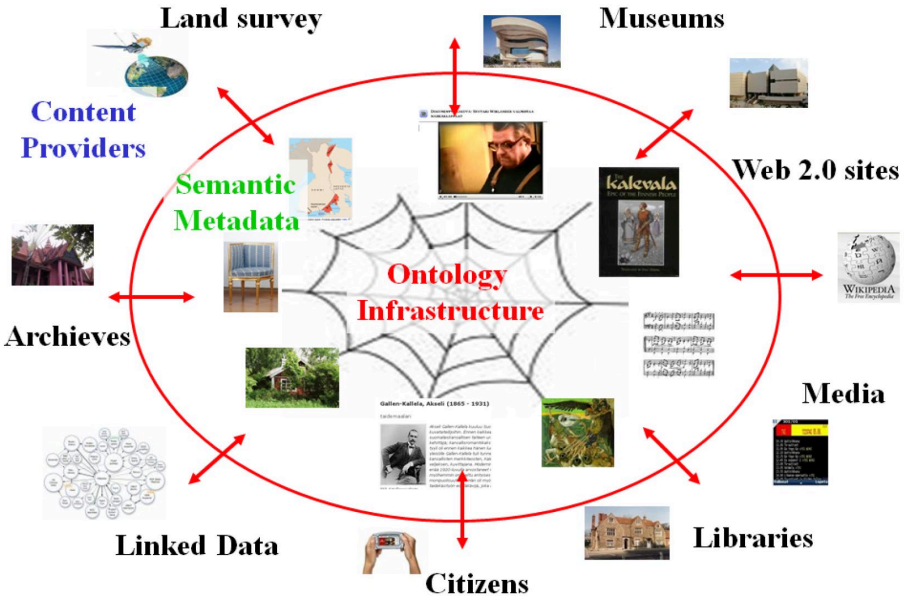
# Sampos development (2008-2020)

- **CultureSampo** – Finnish Culture on the Semantic Web (2008)
- **TravelSampo** - Mobile Contextualized Services of Cultural Tourism (2011)
- **BookSampo** – Finnish Fiction Literature on the Semantic Web (2011) [2 million users in 2018]
- **WarSampo** – Finnish WW2 on the Semantic Web (2015) [230 000 users in 2018]
  - » + War Cemeteries application [2017]
  - » + War Prisoners application [to be published 29.11.2019]
- **BiographySampo** - Finnish Biographies on the Semantic Web (2018) [15 000 users]
- **NameSampo** – Linked Data Workbench for Toponomastic Research (2019) [30 000 users]
- **WarVictimSampo 1914-1922** – National War History, with National Archives of Finland, to be published 20.11.2019
- **MMM – Manuscript Migration Mapper**, medieval manuscripts with Oxford, Pennsylvania, IRHT (Paris), to be published in 23.11.2019
- **FindSampo** – Archaeology & Citizen Science of metal detectorists, with Finnish Heritage Agency (forthcoming 2020)
- **LawSampo** – Finnish Legislation and Case Law, with Ministry of Justice of Finland (forthcoming 2020)
- **AcademySampo** – Historical Finnish Academic People in 1640–1899 (forthcoming 2020)

Source: Hyvönen, Eero, 2019-11-18 email conversation. Paper to be published in 2020.

**List of 20+ ontologies**: https://seco.cs.aalto.fi/ontologies/

# The Take-aways from the Sampo model




**LAMs Data's direction →**

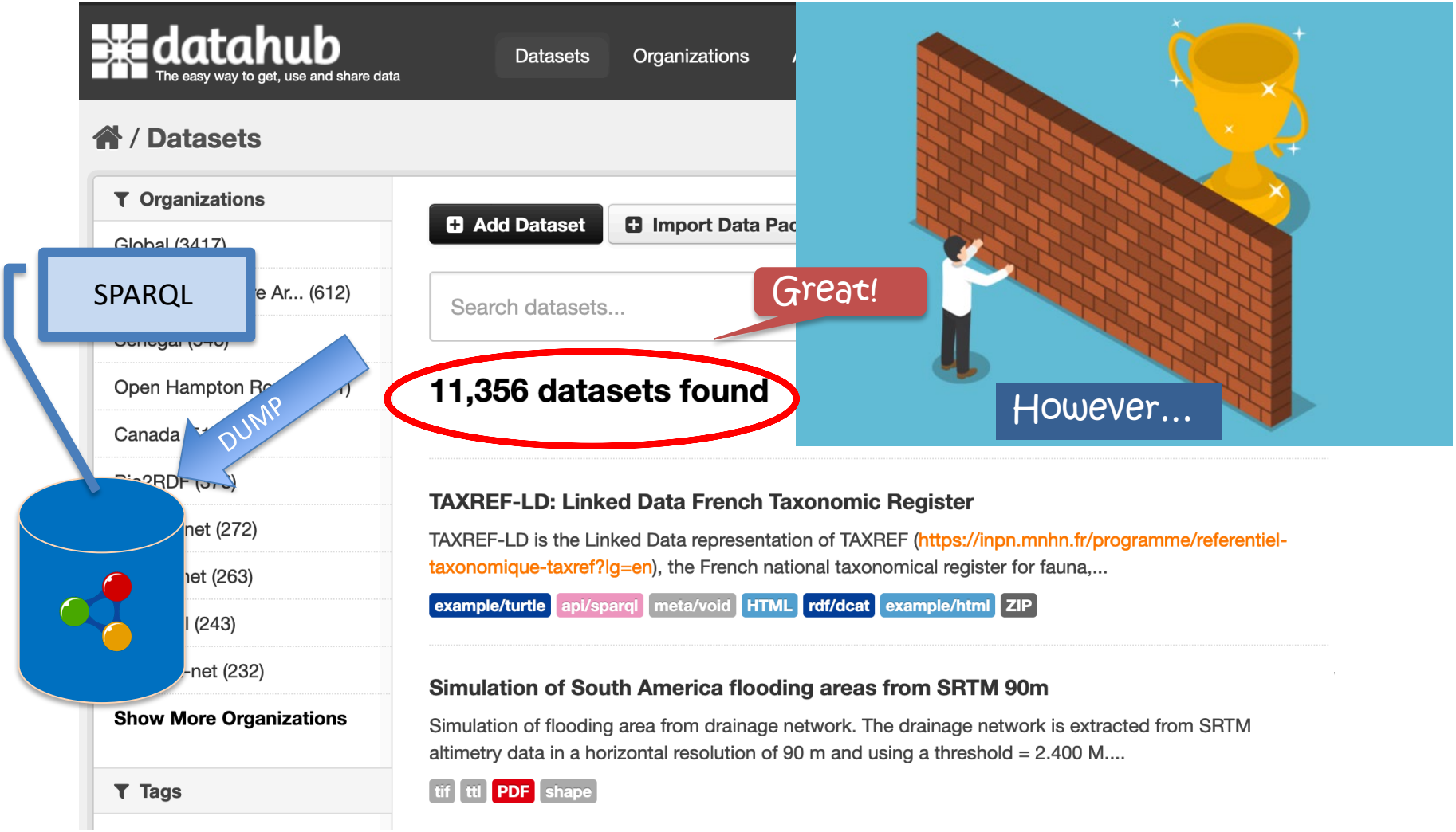
Data Silos → Data Lakes → **Data Planets** → **Data Universe**

Serving humanities research

Two major types of challenges

- a) Challenges for publishing cultural heritage collections on the Web
- b) Challenges for easier access & effective usage of LOD datasets by humanities researchers

# b) Challenges for easier access & effective usage of LOD datasets by humanities researchers



**Data Access**

Constructing a user-friendly environment for SPARQL query

# *Quo vadis,* LOD? Problem of data reuse after data conversion

| Questions | Possible Solutions |
| --- | --- |
| Not knowing the data content and what could be queried | Indication of SPARQL examples of each possible querying/ research needs |
| Not skilled with SPARQL language | Giving instruction of step-by-step SPARQL query reasoning |
| Not having sense with queried data results; Not knowing benefits gaining from the queried results | Implementation of tools for data visualization |

Sophy S. Chen, Lu-yen Lu. Linked Open Data and Possible Approaches for Constructing a free Platform for Application: Cases from ASCDC's LOD Datasets, Dublin Core 2019 Conference, 26. September, 2019; Seoul, KOREA

**What is needed for action:**

- The  LOD Datasets need appropriate practices for:
  - Effective usages by users (human and machine)
- The barrier lies in the **communication about** the datasets rather than in their own structures, formats, or contents.
- Need to take the full advantages of LOD datasets as knowledge bases (i.e., not just the published RDF triple stores).
- Follow best practices and share products for LOD delivery
  - W3C *Data On the Web Best Practices* recommendation.2017. https://www.w3.org/TR/dwbp/
  - FAIR principles https://www.go-fair.org/fair-principles/

Image credit: https://www.liquidanalytics.com/blog/2018/7/13/overcome-your-ai-barriers

# Cases

- Effort: Liberate users from unfamiliar query languages & provide effective analytics outcomes
  - Case 1: Online Coins of the Roman Empire (OCRE), http://numismatics.org/ocre/
- Effort: Provide query examples to users, reveal the rich contents of the datasets
  - Case 2: Wikidata Query Service
  - EXAMPLE. Query : *Map of economists in PM20* by place of birth* Query available at: https://www.wikidata.org/wiki/Wikidata:WikiProject_20th_Century_Press_Archives/Use_cases
- Effort: Enable users to master queries through template, Turn LOD KOS into easy-to-use knowledge bases
  - Case 3. Gerry Vocabularies
  - http://vocab.getty.edu/queries

# The Take-aways

Efforts

Towards Easier Access & Effective Usage of LOD Datasets

See also: three cases.

1. Aggregated RDF triples are presented with user-understandable interface (e.g., DBpedia).

2. Provide query examples to users
Reveal the rich contents of the datasets (e.g., WIKIDATA )

3. Enable users to master queries through template, turning LOD KOS into easy-to-use knowledge bases (e.g., Getty Vocabularies LOD).

4. Liberate users from unfamiliar query languages & provide effective analytics outcomes (e.g., OCRE)

   Code-free quantitative analysis tools, comprehensive comparison and exploration options, user-friendly webpages

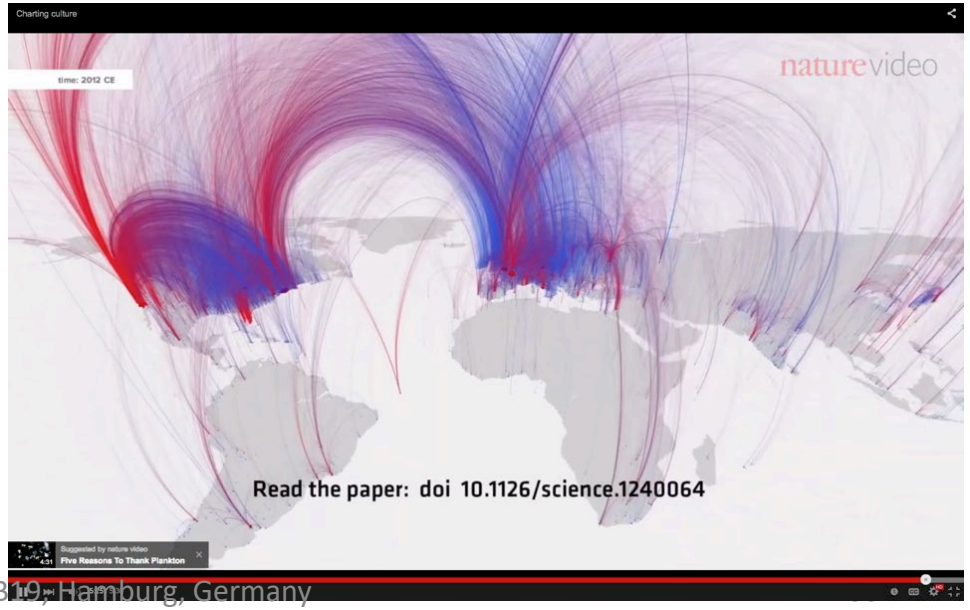
5. Expose datasets to machines, transfer structured data into knowledge graphs

# Conclusion (1/2)

**Advanced technologies now allow researchers :**

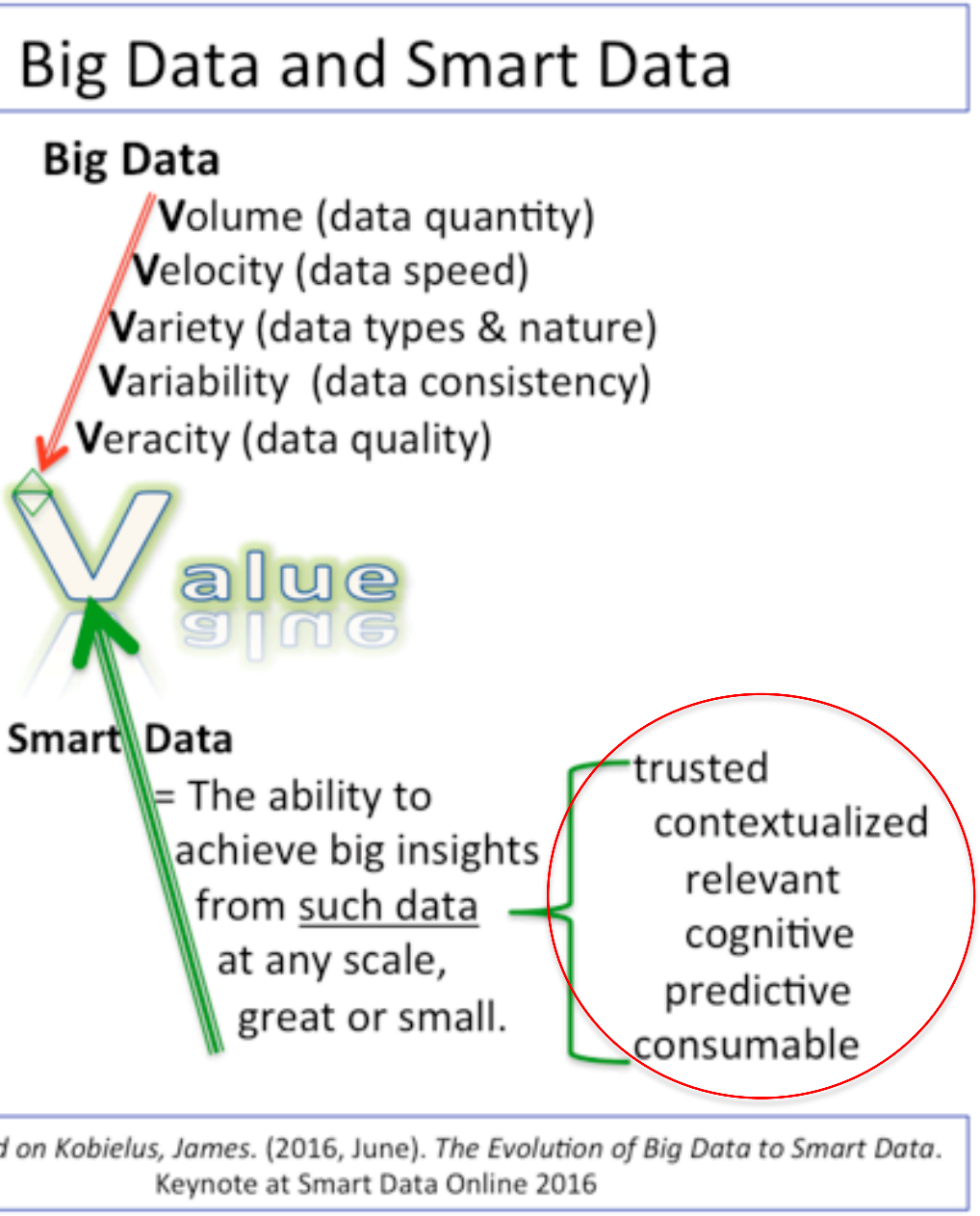
**(under the umbrella of Big Data and the Semantic Web)**

- to access and reuse large volumes of diverse data,
- to discover patterns and connections formerly hidden from view,
- to reconstruct the past,
- to discover impacts in real and virtual environments, and
- to bring the complex intricacies of innovations to light,
  all as never before.

# Conclusion (2/2)

- Challenges and opportunities co-exist.
- It is certain that Smart Data will continue to have extraordinary value in digital humanities.
- Data provided by LAMs and cultural heritage institutions are treasures for all humanities researchers.






Zeng 2017 DOI: 10.1515/jdis-2017-0001

SWIB19
Semantic Web in Libraries

# Smart Data for Digital Humanities

Thank you!


Marcia Lei Zeng

Kent State University

http://marciazeng.slis.kent.edu/

# References (1)


- Ackoff, R. L. (1989). From data to wisdom. Journal of Applied Systems Analysis, 16(1), 3-9.
- Anderson, C. (2008). The end of theory: the data deluge makes the scientific method obsolete. Wired 16(7). Retrieved from: https://www.wired.com/2008/06/pb-theory/
- Antonio, M. L., et. al. (2019). Ancient Rome: A genetic crossroads of Europe and the Mediterranean. Science, 366(6466), 708-714.
- Aasman, Jans . (2019). Why Knowledge Graphs Now. Ontology Summit, Sept. 4. Retrieved from: http://ontologforum.org/index.php/ConferenceCall_2019_09_04
- Borgman, C. (2015). Big data, little data, no data: Scholarship in the networked world. Cambridge, MA: MIT Press.
- Borne, K. (2013, April 4). Big data, small world: Kirk Borne at TEDxGeorgeMasonU [Video file]. Retrieved from https://www.youtube.com/watch?v=Zr02fMBfuRA
- Burdick, A., Drucker, J., Lunenfeld, P., Presner, T., & Schnapp, J. (2012). Digital_Humanities. Cambridge, MA: MIT Press.
- Chen, S. & Lu, L. Linked Open Data and Possible Approaches for Constructing a free Platform for Application: Cases from ASCDC's LOD Datasets, Dublin Core 2019 Conference, 2019 Setp, Seoul, KOREA
- Consultative Committee for Space Data Systems. (2012). Reference model for an open archival information system (OAIS): Recommended practice (CCSDS 650.0-M-2: Magenta Book). Washington, DC. CCSDS. Retrieved from http://public.ccsds.org/publications/archive/650x0m2.pdf
- DATAVERSITY Education, LLC. (2017). Smart Data Conference (website). Retrieved from http://smartdata2017.dataversity.net
- Digging into Data Challenge. (n.d.). Retrieved from https://diggingintodata.org/
- Gardner, D. (2012). An ocean of data [Introduction]. In R. Smolan & J. Erwitt (Eds.), The human face of big data (14-17). Sausalito, CA: Against All Odds Productions.
- Gantz, J., & Reinsel, D. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the Far East. IDC iView, December 2012, 1-16. Retrieved from http://www.dedupecentral.co.uk/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

# References (2)


- Humby, C. (2006). Data is the new oil. Talk given at the Association of National Advertisers (ANA) Senior Marketer's Summit, Kellogg School. [Source: M. Palmer. (2006 Nov. 3). Data is the New Oil (Web log post). Retrieved from http://ana.blogs.com/maestros/2006/11/data_is_the_new.html]
- Hyvönen, E. (2016). Cultural Heritage Linked Data on the SemanticWeb: Three Case Studies Using the Sampo Model. VIII Encounter of Documentation Centres of Contemporary Art. Vitoria-Gasteiz, Spain, Oct. 19-20.
- Hyvönen, E. [2019]. Using the Semantic Web in Digital Humanities: Shift from Data Publishing to Data-analysis and Serendipitous Knowledge Discovery. *Semantic Web*. [in press]
- Iafrate, F. (2015). From Big Data to Smart Data. London: ISTE Ltd., and Hoboken, NJ: John Wiley & Sons, Inc.
- Kaplan, F. (2015). A map for big data research in digital humanities. Frontiers in Digital Humanities, 2, p. 1. DOI 10.3389/fdigh.2015.00001
- Kobielus, J. (2016, June). The evolution of big data to smart data [PowerPoint slides]. Keynote at Smart Data Online 2016.
- Mayer-Schönberger, V. & Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. New York, NY: Eamon Dolan/Houghton Mifflin Harcourt.
- Morrison, Alan. (2019). Data-centric design and the knowledge graph. Data Centric Design for SMART DATA & CONTENT Enthusiasts meetup, July 31, 2019, Chicago. Retrieved from https://www.slideshare.net/AlanMorrison/datacentric-design-and-the-knowledge-graph
- Mukerjee, P. (2014). Introduction to data science [PowerPoint slides]. Retrieved from http://www.slideshare.net/prithwis/01-intro2-datascienceyantrajaalblog
- National Endowment for the Humanities (NEH). [2016]. Grants. Retrieved from https://www.neh.gov/grants
- Schich, M., Song, C., Ahn, Y.Y., Mirsky, A., Martino, M., Barabási, A.L., & Helbing, D. (2014a). A network framework of cultural history. Science, 345(6196), 558-562.
- Schich, M., Song, C., Ahn, Y. Y., Mirsky, A., Martino, M., Barabási, A. L., & Helbing, D. (2014b, July 31). Charting culture. Nature Video [Video file]. Retrieved from https://www.youtube.com/watch?v=4gIhRkCcD4U
- Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the humanities. Journal of Digital Humanities, 2(3), 13.

# References (3)


- Sheth, A. (2014, March). Transforming big data into smart data: Deriving value via harnessing volume, variety and velocity using semantics and Semantic Web, [PowerPoint Slides]. Keynote at 30th IEEE International Conference on Data Engineering (ICDE) 2014. Retrieved from http://ieeexplore.ieee.org/document/6816634/ DOI: 10.1109/ICDE.2014.6816634
- Svensson, P. (2009). Humanities computing as digital humanities. Digital Humanities Quarterly, 3(3). Retrieved from http://digitalhumanities.org/dhq/vol/3/3/000065/000065.html
- Svensson, P. (2010). The landscape of digital humanities. Digital Humanities Quarterly, 4(1). Retrieved from http://digitalhumanities.org/dhq/vol/4/1/000080/000080.html
- TiECON East. (2014). Data is the new oil. Retrieved from http://www.tieconeast.org/2014/big-data-analytics
- Taylor, Jamie. (2018). Knowledge Graph. Panel “Enterprise-scale knowledge graphs”. International Semantic Web Conference 2018. Retrieved from: iswc2018.semanticweb.org/wp-content/uploads/2018/10/Panel-all.pdf
- Van Ruyskensvelde, S. (2014). Towards a history of e-ducation? Exploring the possibilities of digital humanities for the history of education. Paedagogica Historica, 50(6), 861-870.
- Wallis, R. (2016, July). Contextual computing with knowledge graphs and the web of entities. Presentation at Smart Data Online 2016.
- Weingart, S. (2016). Submissions to DH2016 (pt.1) [Web log post]. Retrieved from http://www.scottbot.net/HIAL/index.html@tag=dhconf.html
- Weingart, S. (2017). Submissions to DH2017 (pt.1) [Web log post]. Retrieved from http://scottbot.net/submissions-to-dh2017-pt-1/
- Zeleny, M. (1987). Management support systems: Towards integrated knowledge management. Human Systems Management, 7(1), 59-70.
- Zeng, M. L. (2017). Smart data for digital humanities. Journal of Data and Information Science. 2(1), 1-12. DOI: 10.1515/jdis-2017-0001
- Zeng, M. L. (2018). Interoperability. ISKO Encyclopedia of Knowledge Organization (IEKO). http://www.isko.org/cyclo/interoperability
- Zeng, M. L. (2019). Semantic enrichment for enhancing LAM data and supporting digital humanities. El profesional de la información, 28(1), e280103. https://doi.org/10.3145/epi.2019.ene.03