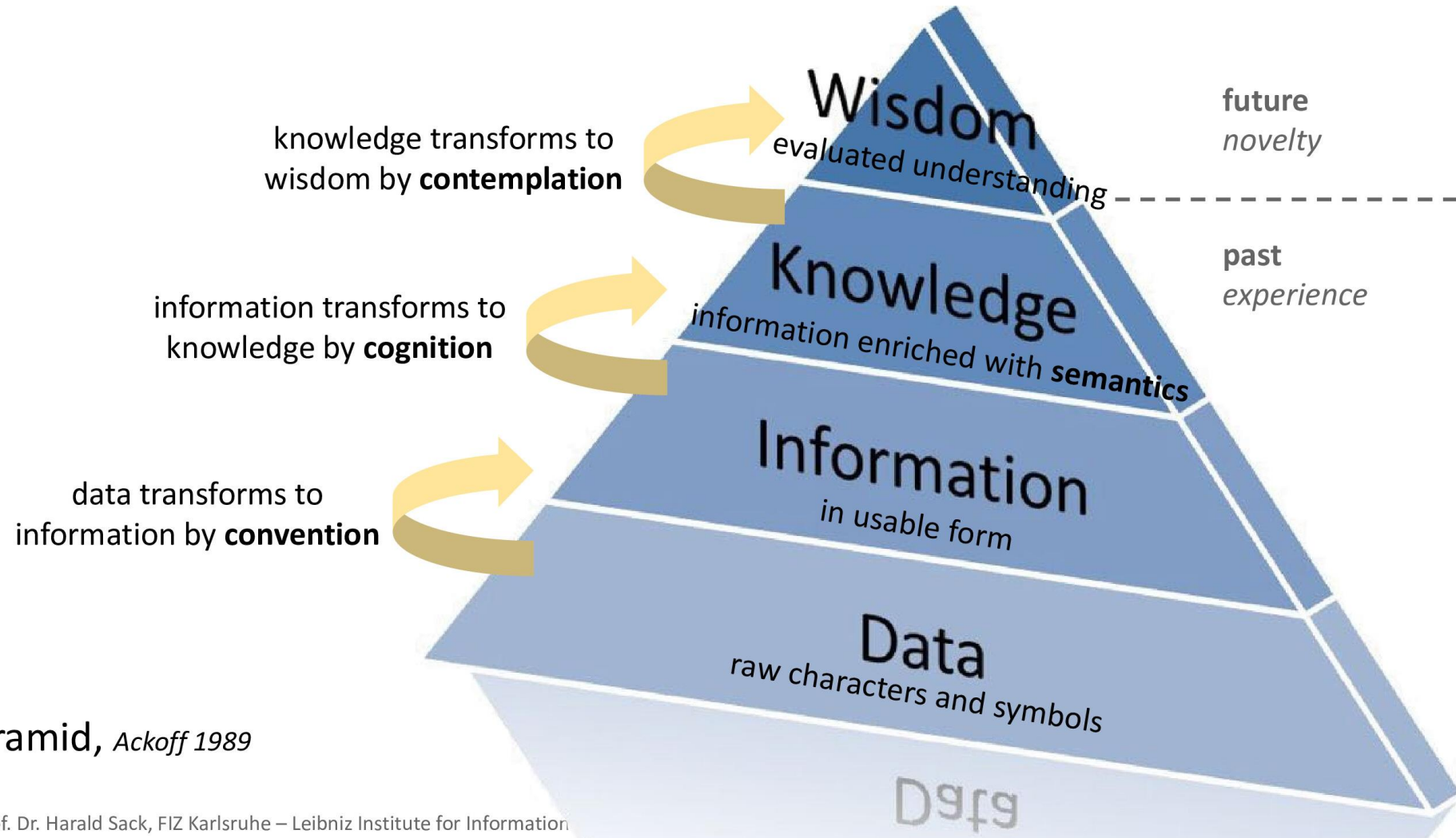


# WHERE IS THE KNOWLEDGE?

Krzysztof Kutt, PhD  
Knowledge in AI Systems  
WFAIS UJ

# Data, Information, Knowledge



DIKW Pyramid, Ackoff 1989

# SOURCES OF KNOWLEDGE

1. Experts
2. Knowledge bases
3. Data mining & machine learning
4. Data enrichment

# EXPERTS

The primary source  
of knowledge

# DOMAIN EXPERTS

- People want to collaborate
- There is a need for methods and tools that work
- It's not only about KRR methods

	Past	Present
Design	<b>KE expert</b> with domain expert access	KE expert paired with <b>domain expert(s) and community</b>
Population	<b>KE expert</b> learns domain	KE and <b>domain experts</b> determine the vocabulary
Evolution	<b>KE expert</b> heavily involved	KE expert involved in tools customization for <b>domain experts</b>
Tool users	Trained in <b>Computer Science</b>	Trained in <b>Domain Sciences</b>
Application users	Well understood group	Diverse and evolving group
Reuse	Well thought out	Expect the unexpected

[My ICD](#)[ICD Content](#)[Category Notes and Discussions](#)[Reviews](#)[Change History](#)[Manage Hierarchy](#)[Export and Import](#)

**ICD Categories**

Create Watch Branch ▾ Search: '29E' 'Roseola infantum'

- 07 VII Diseases of the eye and adnexa 9 2773
- 08 VIII Diseases of the ear and mastoid process 7
- 09 IX Diseases of the circulatory system 4 2285
- 10 X Diseases of the respiratory system 3 1385
- 11 XI Diseases of the digestive system 4 4381
- 12 XII Diseases of the skin 10 6775
  - LA Infections and infestations affecting the skin
    - LA0 Viral infections affecting the skin 2 2
    - LA00 Pox virus infections of the skin 2
    - LA01 Herpes virus infection of skin and mucous membranes 2
    - 29A0 Herpes simplex infection of skin and mucous membranes 2
    - LA011 Varicella zoster infection of skin 2
    - LA018 Other human herpes virus infections 2
    - 29E Roseola infantum 4
    - LA02 Human papilloma virus infection of skin 2
    - LA05 Skin disorders related to HIV and other immunodeficiencies 2
    - LA07 Viral exanthems 1 17
    - LA08 Miscellaneous skin disorders resulting from viral infections 2
    - LA09 Miscellaneous dermatoses with suspected viral etiology 2

**Details for 29E Roseola infantum**

**Title & Definition** Classification Properties Terms Clinical Description

Manifestation Properties Causal Properties Temporal Properties Severity Properties

Functioning Properties Specific Condition Properties Treatment Diagnostic Criteria

ICD-10 Notes and Hints ICD-10 Linearizations **Editorial Information**

**ICD-10 Code** ? B08.2 ✕

**Sorting label** 29E ✕

**ICD Title** ? Roseola infantum ✕

**Short Definition** ?

Text			
An acute, short-lived, viral disease of infants and young children characterized by a high fever at onset that drops to normal after 3-4 days and the concomitant appearance of a macular or maculopapular rash that appears first on the trunk	✕	2	

## ICD-11 BY WHO

Joint efforts by medics in WebProtege



# Professor catalog of the University of Leipzig | catalogus professorum Lipsiensis

- Home
- Epochs
- Faculties
- Professors of the day
- Rectors and deans
- entire directory

Epochs > [1919-1932](#) and [1933-1945](#)  
 Faculties > [Faculty of Arts I, University of Leipzig - Faculty of Philological-Historical Division \(1920-1951\)](#)  
 Data: [Resource](#) | [RDF](#) | [PDF](#) | [Printer-friendly](#)

## Prof. Dr. phil. Schücking *Levin* Ludwig

### Life

b. 5/29/1878 in Burgsteinfurt  
 d. 12/10/1964 in Farchant  
 PND: [117124931](#)

### Curriculum vitae



Source: Private collection

### Study

- 1897-1901 Study: modern languages and art history in Freiburg, Göttingen, Berlin and Munich
- 1904 Habilitation for English Language and Literature at the University of Göttingen  
Title of work: broad set of shortcut in Beowulf.
- 1901 Promotion to Dr. phil. in English Philology at the University of Göttingen  
Title of work: English Material Relations of the Italian comedy *Lilly*

Professors at the University of Leipzig – their life and work  
[\(link\)](#)

# CATALOGUS PROFESSORUM LIPSIENSIIUM

Full text search

- Background
- Information in English
- Abbreviations
- Literature





WIKIPEDIA  
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

- Interaction
- Help
- About Wikipedia
- Community portal

- Recent changes
- Contact page
- Jobs
- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Wikidata item

Main Page [Talk](#)

Read [View source](#) [View history](#)

## Welcome to [Wikipedia](#),

the [free encyclopedia](#) that [anyone can edit](#).  
5,132,278 articles in [English](#)

- [Arts](#)
- [Biography](#)
- [Geography](#)
- [History](#)
- [Mathematics](#)
- [Science](#)
- [Society](#)
- [Technology](#)
- [All portals](#)

### From today's featured article



Replicas of the "Twin Sisters", cannons used in the Battle of San Jacinto

The **Runaway Scrape** was the 1836 escape of Texas residents from the encroaching Mexican Army of Operations under the command of [Antonio López de Santa Anna](#) during the [Texas Revolution](#). Civilian evacuations

began on the Gulf Coast in January after the vanguard of the Mexican army crossed the Rio Grande to quell the insurrection of American colonists and Tejanos (Mexicans born in Texas). Weeks later, news of the Battle of the Alamo and the Goliad massacre created a state of panic. Santa Anna ordered the Texas commander-in-chief of raw recruits who had little or no combat experience. Fleeing civilians moved in tandem with Houston's troops for protection, as he sought a safe training camp for his soldiers. The pursuing Mexican army had orders to execute all rebel combatants, and it cut a swath of destruction in its search for them. After a mere three weeks training near

### In the news

- American singer and songwriter **Prince** is found dead at the age of 57.
- At least 64 people are killed and more than 340 others injured in **an attack** in [Kabul](#), Afghanistan.
- Ethiopians [Lemi Berhanu Hayle](#) and [Atsede Baysa](#) (pictured) win the men's and women's [Boston Marathon](#), respectively.
- A **magnitude 7.8 earthquake** hits Ecuador, killing at least 480 people and injuring more than 4,000 others.
- **Two earthquakes** in Japan injure more than 3,000 people across Kyushu, Japan.



Atsede Baysa

# WIKIPEDIA

You've probably heard of it before...

### Ongoing events

**Recent deaths:** [Chyna](#) · [Victoria Wood](#) · [Estelle Balet](#) · [Bill Gray](#)



# CROWDSOURCING

Human-labeled data, e.g.:

1. Google reCAPTCHA
2. Amazon Mechanical Turk

User feedback and reviews

# KNOWLEDGE BASES

They are already there

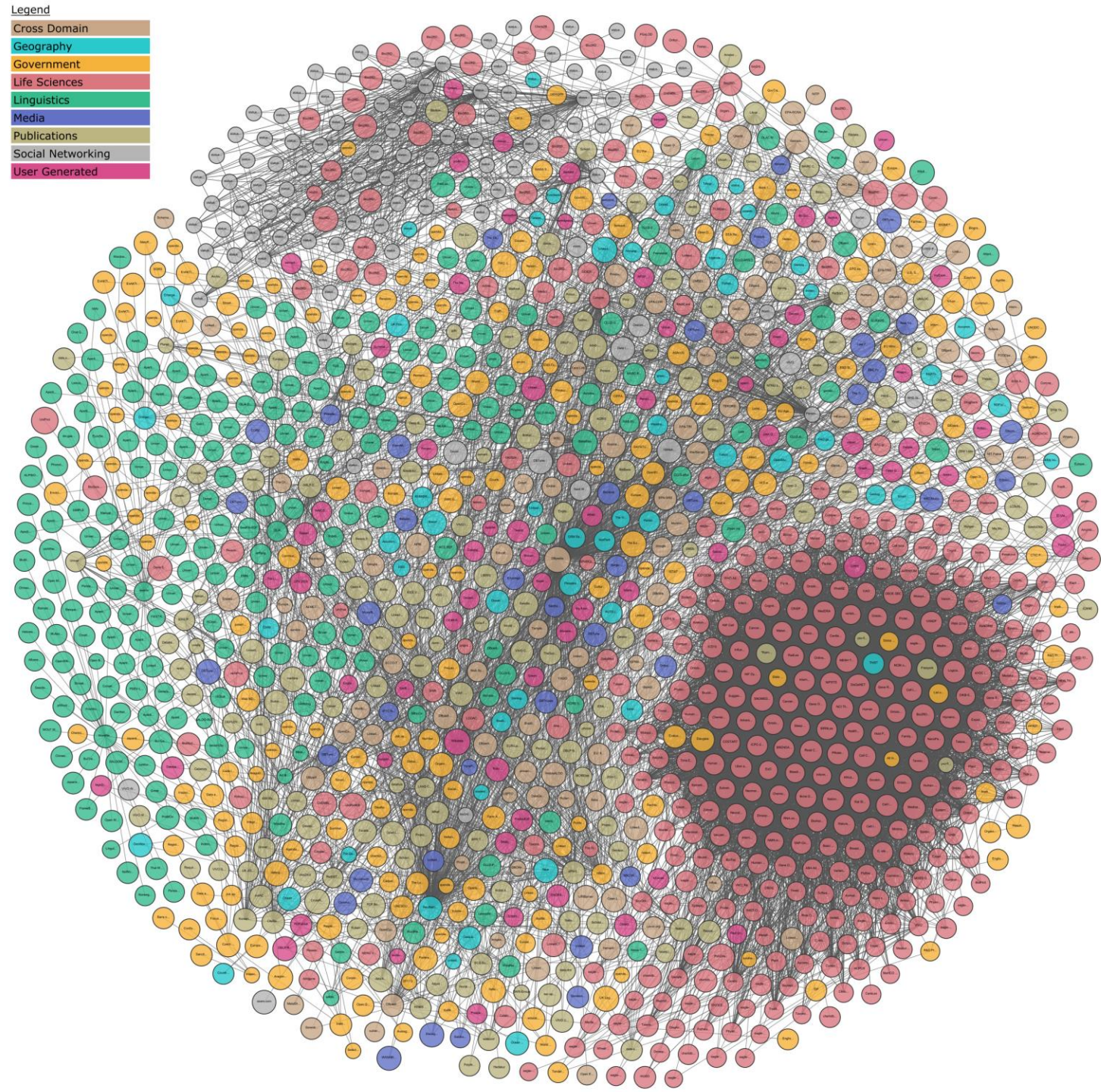
# KNOWLEDGE BASES

1. Knowledge graphs, e.g., Google Knowledge Graph, Wikidata, DBpedia
2. Ontologies, e.g., WordNet, Cyc
3. Scientific databases (e.g., chemistry, genomics, bio-sciences, agro-sciences)
4. APIs / Open Data, e.g., OpenWeatherMap, Google Maps, Hugging Face

# LINKED OPEN DATA

Provides **wider context**

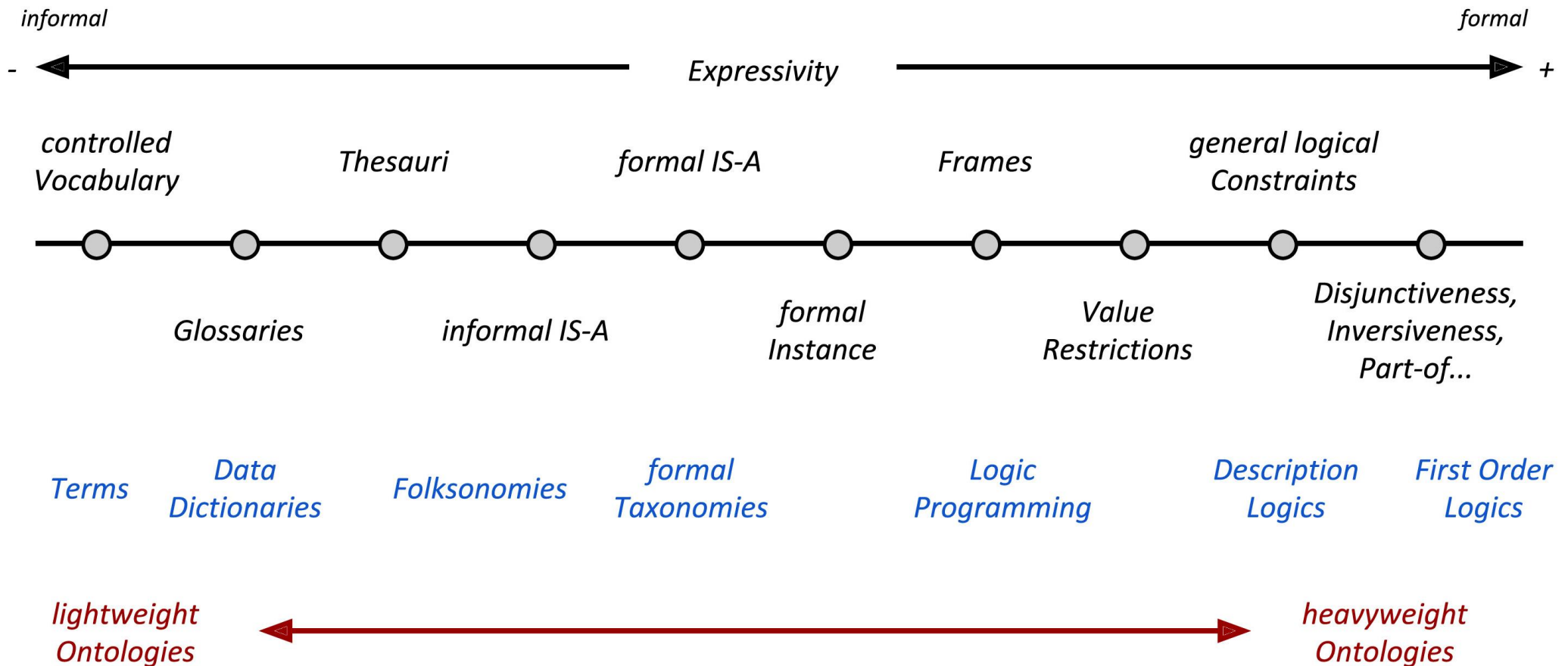
- Wikidata, DBpedia
- Interconnections between bases = **better knowledge!**





# Ontology Types and Categories

according to their level of Generality

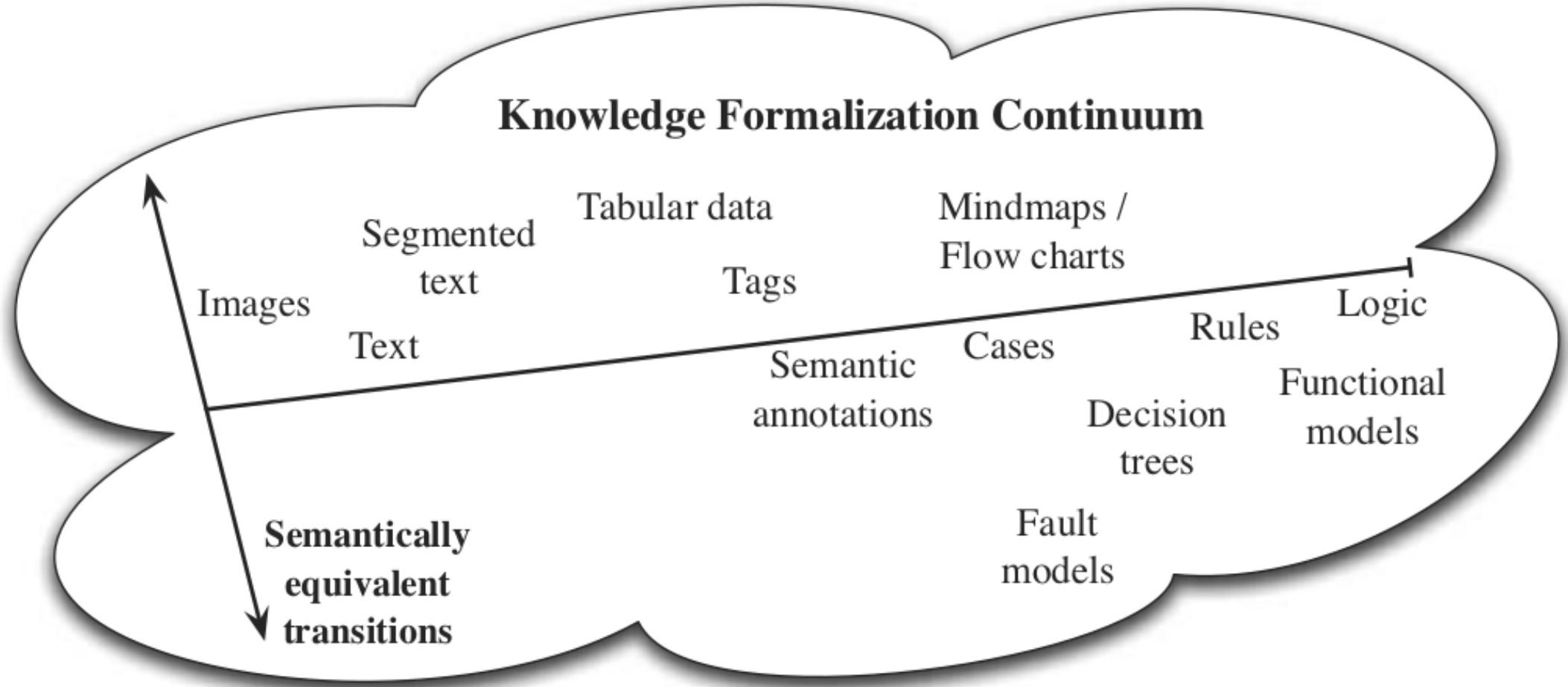


(according to Guarino: Formal Ontology in Information Systems, 1998)

(according to Lassila and McGuinness: The Role of Frame-Based Representation on the Semantic Web, 2001)

Knowledge Graphs 2020, Prof. Dr. Harald Sack & Dr. Mehwish Alam, FIZ Karlsruhe - Leibniz Institute for Information Infrastructure & Karlsruhe Institute of Technology

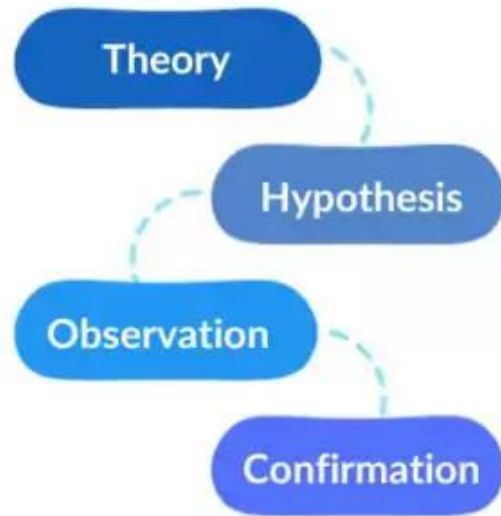
# KFC



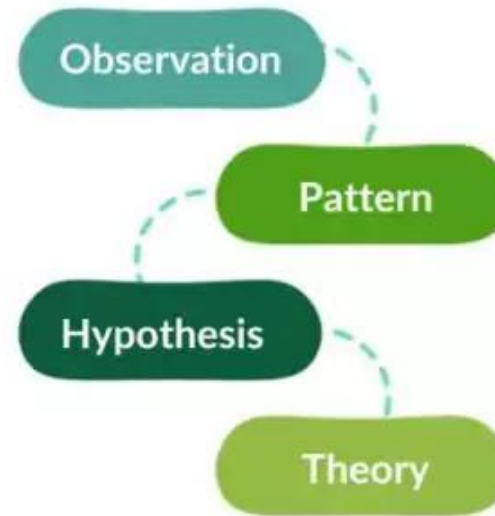


# REASONING

## Deductive Reasoning



## Inductive Reasoning



## Abductive Reasoning



Deductive R.:

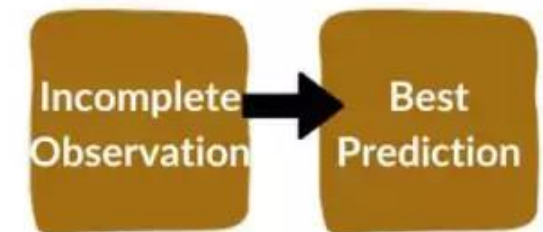
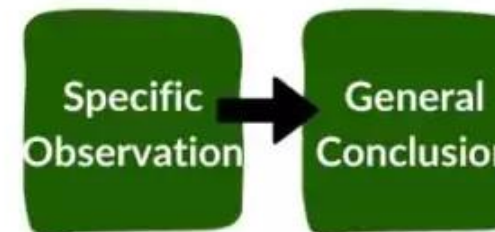
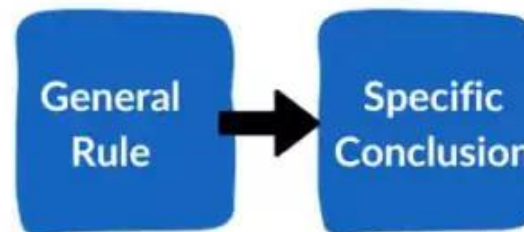
- Rule-based systems, logic

Inductive R.:

- Machine learning

Abductive R.:

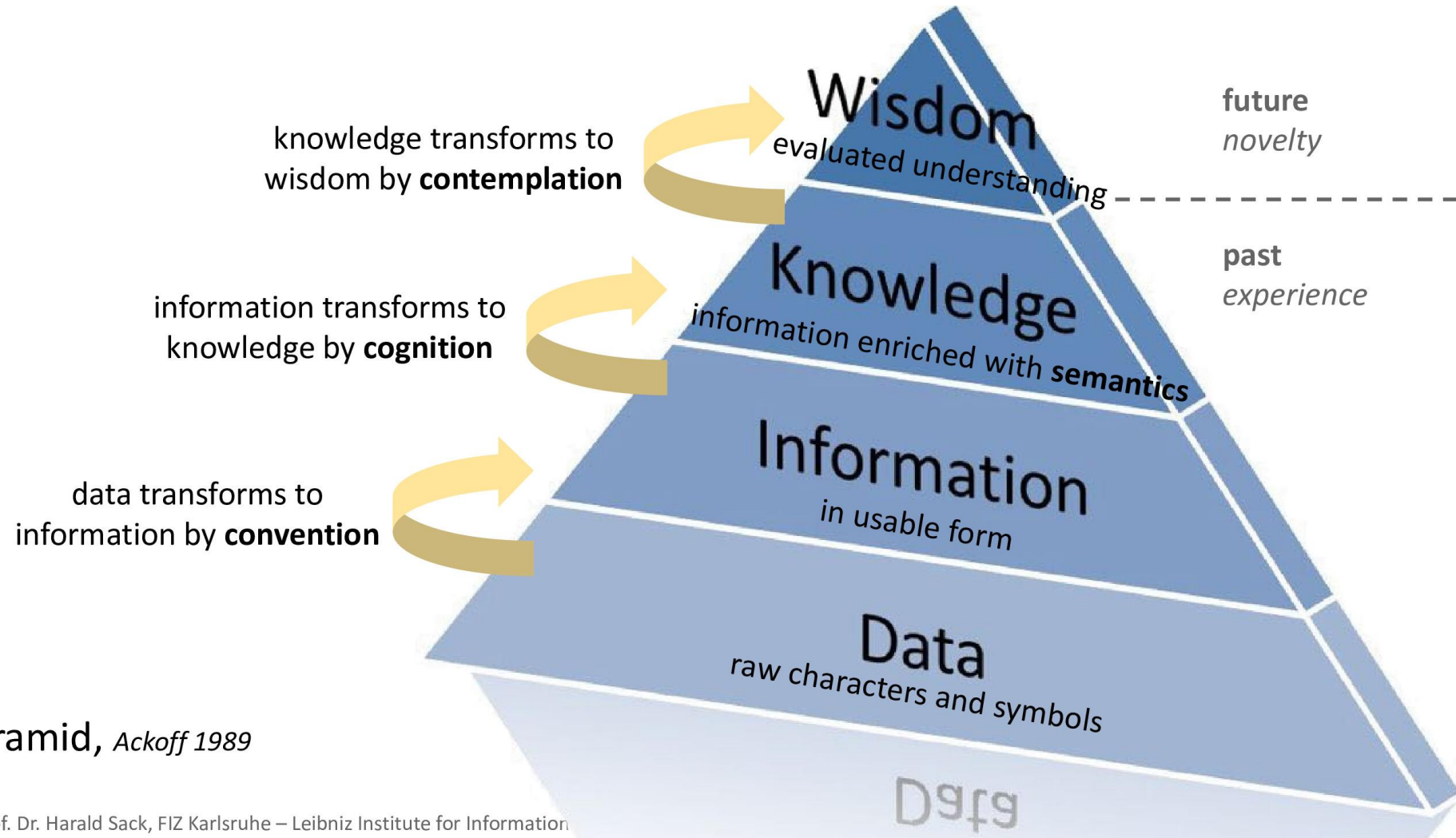
- Problem-solving, diagnosis



# DATA MINING & MACHINE LEARNING

It can be done (semi-  
)automatically

# Data, Information, Knowledge



DIKW Pyramid, Ackoff 1989

# MACHINE LEARNING

## KG embeddings for learning relationships

- Predict missing links, e.g. if two entities should be connected
- Identify new relationships, e.g., a drug's potential use in medicine
- Classify entities, e.g., grouping similar concepts together

## Deep Learning for learning abstract representations

- Natural Language Processing, e.g., for text summarization
- Forecasting in time series, e.g., for learning economic patterns

# MACHINE LEARNING

Reinforcement learning  
for discovering strategies

Learning optimal strategies through trial  
and error approach

- AlphaGo Zero discovered new strategies in Go
- Physics simulations

Generative AI  
for creating new knowledge

Use with special care!

# DATA MINING

## Sequential Pattern Mining

- Finds patterns in sequences of events or actions over time
- *Common techniques:* PrefixSpan, GSP (Generalized Sequential Pattern)
- *Used in:* stock market prediction, clickstream analysis

## Process Mining

- Analyzes event logs to optimize business workflows
- *Common techniques:* Alpha Miner, Heuristic Miner
- *Used in:* business process automation



# DATA MINING

## Text Mining

- Extracts insights from unstructured text data
- *Common techniques:* TF-IDF, topic modeling (e.g., LDA, BERT), Named Entity Recognition
- *Used in:* fake news detection, chatbots, search engines

## Sentiment Mining

- Analyzes emotions and opinions from text data
- *Common techniques:* Transformers (e.g., BERT), lexicon-based techniques (e.g., SentiWordNet)
- *Used in:* social media monitoring, brand reputation management

# DATA MINING

## Cluster Mining

- Groups similar data points together
- *Common techniques:* K-Means, Hierarchical Clustering, DBSCAN
- *Used in:* customer segmentation, anomaly detection

## Graph Mining

- Extracts patterns and structures from graphs
- *Common techniques:* Node2Vec, PageRank
- *Used in:* social network analysis, knowledge graphs, fraud detection

# DATA MINING

## Semantic Data Mining

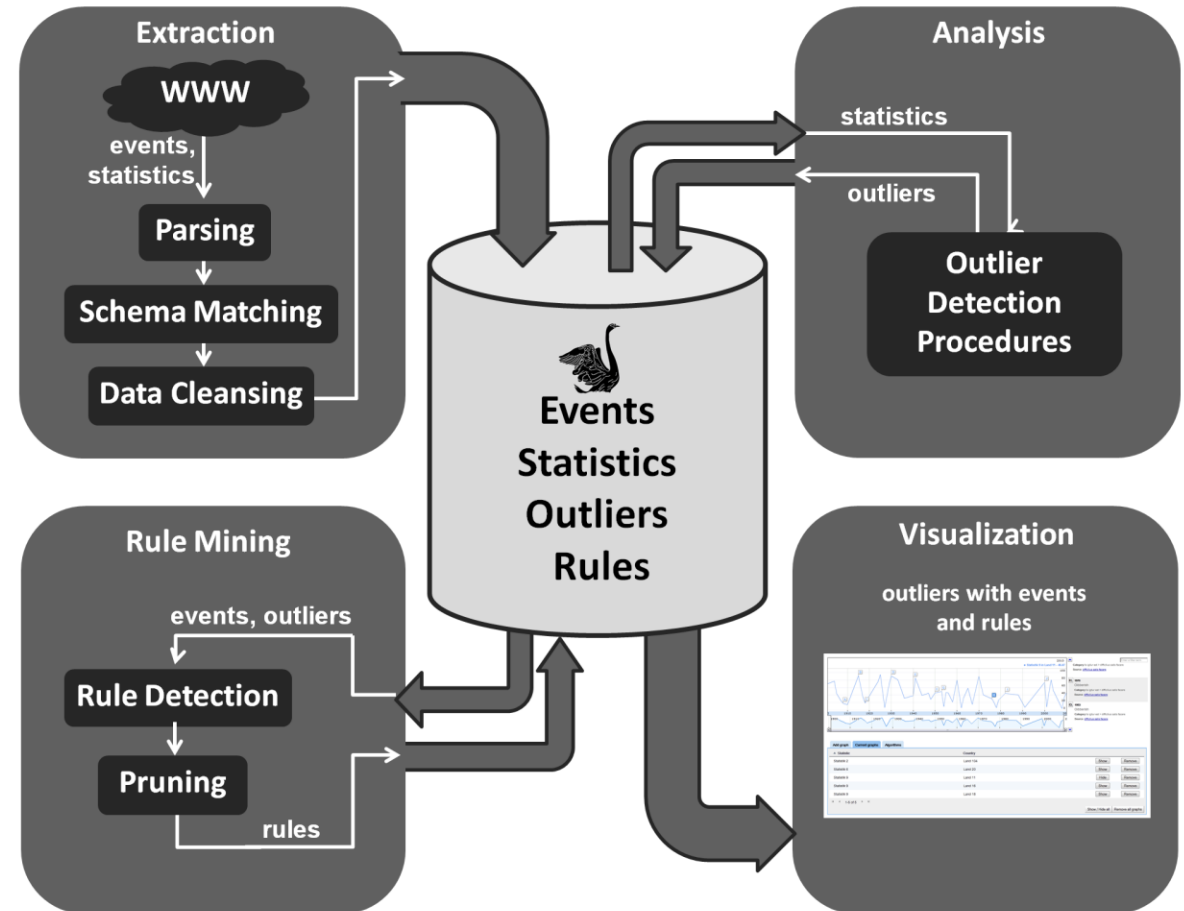
- Integrates domain knowledge, ontologies and structured data to exact meaningful insights
- **Understands the context** instead of just finding statistical patterns (like other mining techniques)
- *Common techniques:* ontology-based mining, semantic similarity measures, knowledge graph embeddings

## Rule Mining

- Extracts if-then rules leading to discovering hidden correlations and causality
- *Common techniques:* Apriori algorithm, FP-growth algorithm
- Can be combined with Semantic Data Mining to verify rules against domain knowledge or to provide better explanations (based on domain knowledge)

# BLACK SWAN

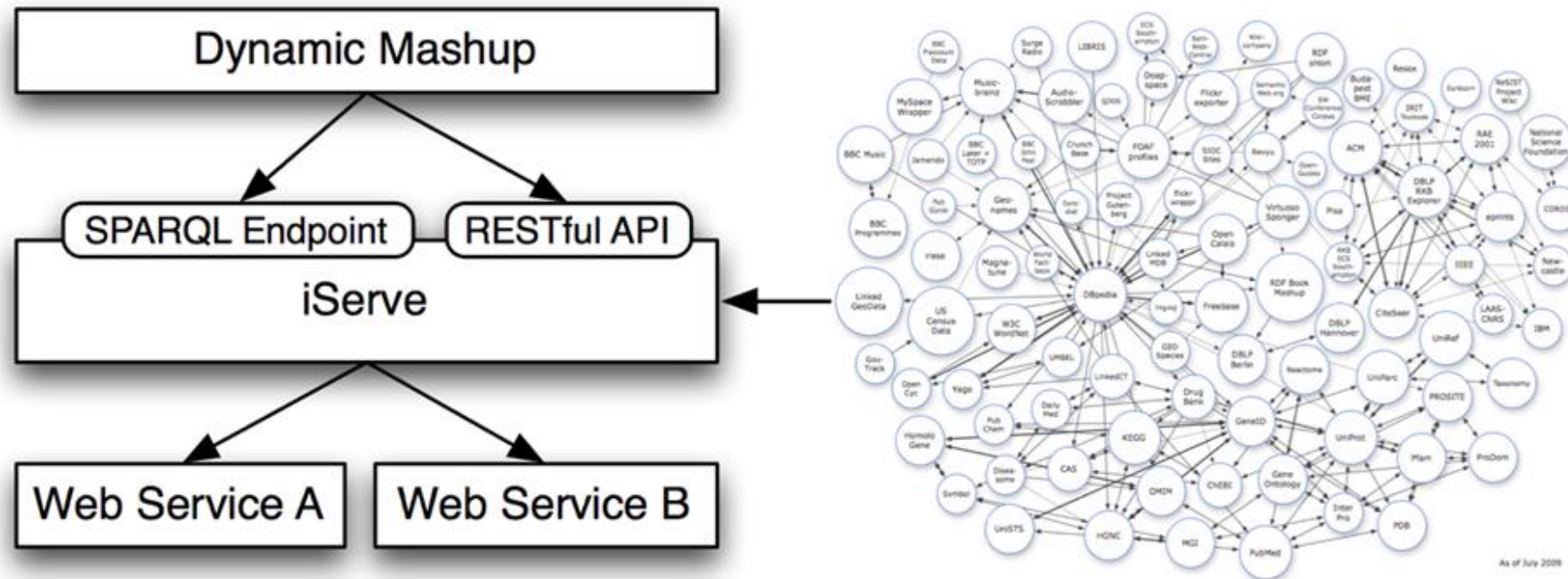
- Extracts raw data and creates knowledge by, e.g., outlier detection and rule mining



# DATA ENRICHMENT

Enhance, improve or augment the data with additional information

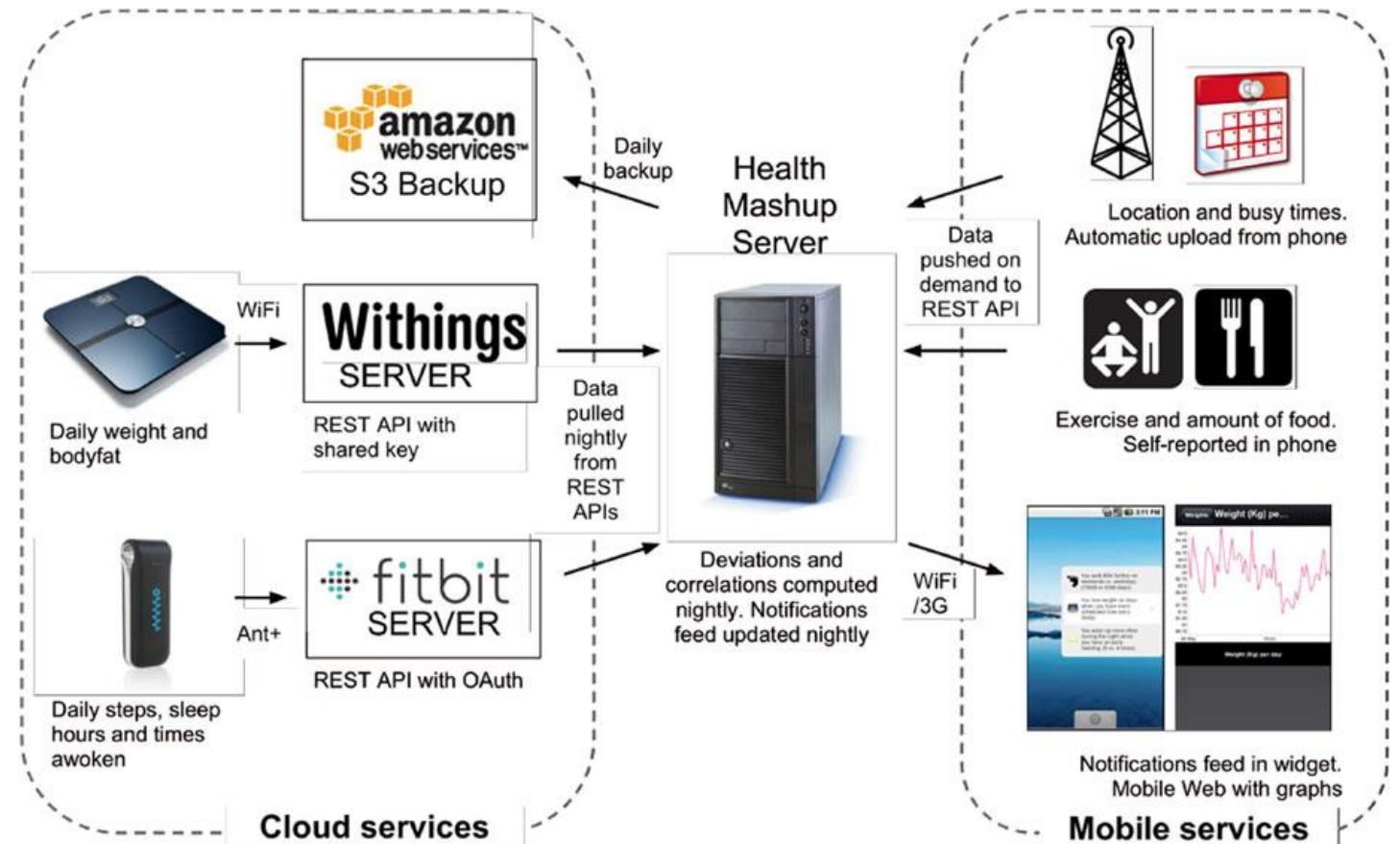
# DATA INTEGRATION: MASHUPS





# PERSONAL HEALTH MASHUP

- Sensory data (from wearables) integrated with self-reports and automatically collected data about location
- Pre-processed and then displayed to the user as a set of widgets



# DATA ENRICHMENT

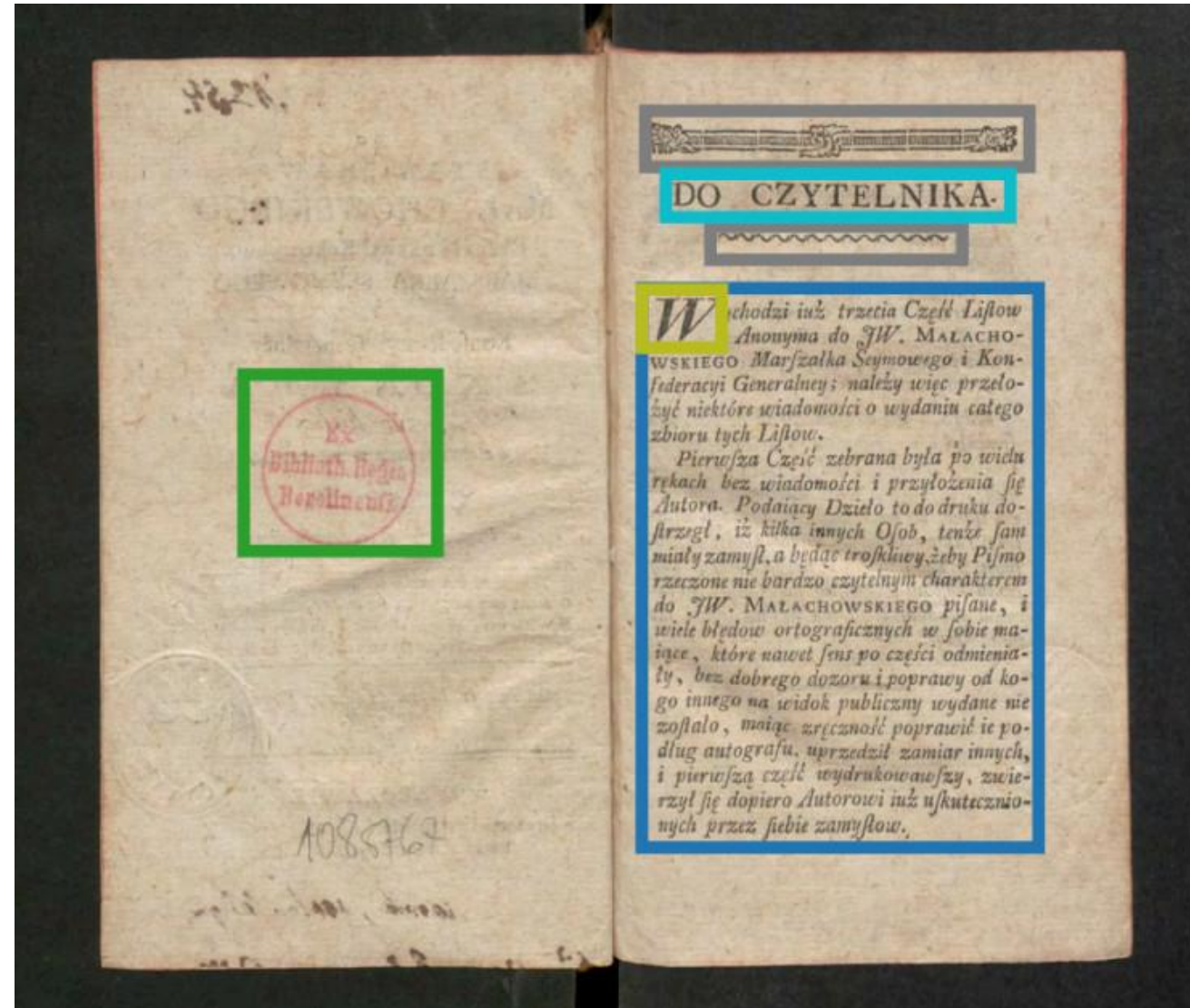
- Data cleaning and standardization: removes errors, duplicates, etc
- Data imputation: fills missing data
- Data augmentation: creates additional data, e.g. by rotating images in CV

# CHEXRISH

Lack of metadata in digital libraries = it is difficult to find specific documents

Our approach:

- Step 1: Use semantic segmentation to identify various parts of the document
- Step 2: Use segment-specific models to generate more detailed metadata



# KNOWLEDGE IS EVERYWHERE

- **Experts:** domain experts, crowdsourcing
- **Knowledge bases & reasoning**
- **Machine learning and data mining**  
techniques for discovering new knowledge from structured and unstructured data and information
- Manual/semi-automatic **data enrichment** with contextual information



**KEEP  
CALM  
AND  
CARRY  
ON**

THANK YOU FOR  
YOUR ATTENTION!

GEIST Research Group: <https://geist.re/>

Krzysztof Kutt: <https://krzysztof.kutt.pl/>



*This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/).*



**KEEP  
CALM  
AND  
ASK  
QUESTIONS!**