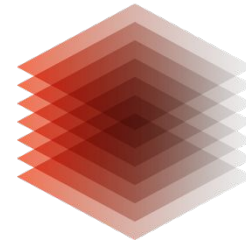


---

LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK

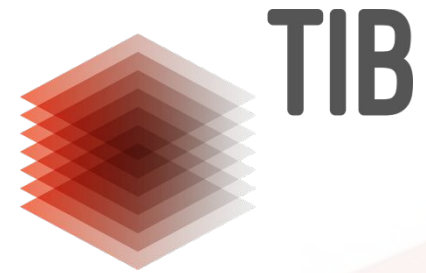


**TIB**

# Open Research Knowledge Graph

Markus Stocker & *The Team*  
Porto, September 17, 2019  
Open Science Fair 2019

LEIBNIZ-INFORMATIONSZENTRUM  
TECHNIK UND NATURWISSENSCHAFTEN  
UNIVERSITÄTSBIBLIOTHEK



Jennifer D'Souza  
Researcher



Mohamad Yaser Jaradeh  
Developer



Manuel Prinz  
Developer



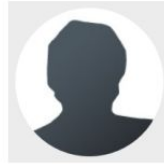
Allard Oelen  
Developer



Arthur Brack  
PhD Student



Kheir Farfar  
Developer



Markus Stocker  
Co-Lead



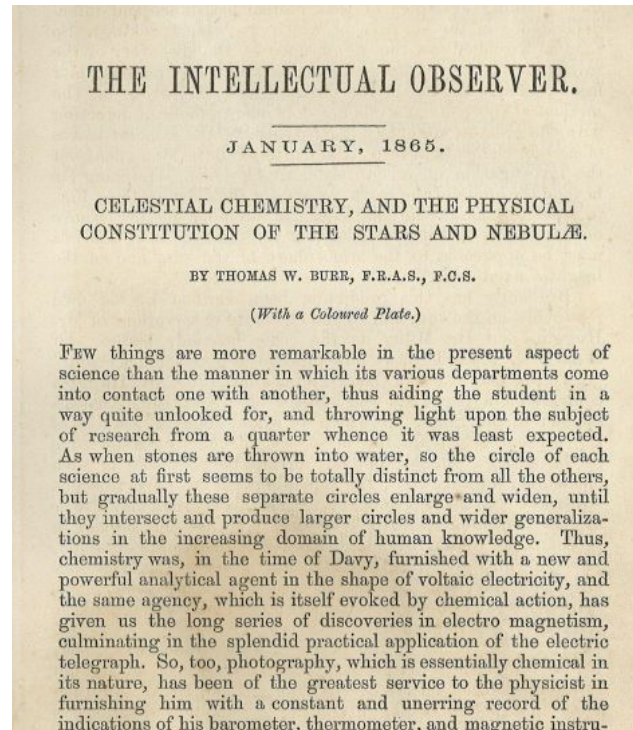
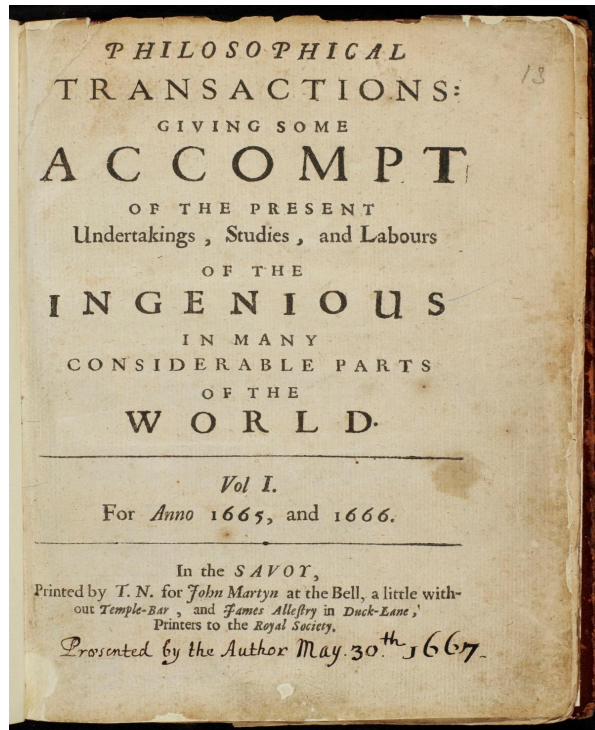
Sören Auer  
Lead



## What if ...

- The global scientific knowledge base would be more than a document repository
- Scientific information and knowledge would be FAIR also for machines
- Currently
  - Findability could be better
  - Assuming OA, accessibility is OK
  - Interoperability and Reusability is non-existent
- The scholarly communications infrastructure is stuck in the last century
- We have managed to *digitize* documents that used to be in print
- While other areas have seen a transformative *digitalization*

# Digitization of scholarly communications



### AGDISTIS - Graph-Based Disambiguation of Named Entities using Linked Data

Ricardo Usbeck<sup>1,2</sup>, Axel-Cyrille Ngonga Ngomo<sup>1</sup>, Michael Röder<sup>1,2</sup>, Daniel Gerber<sup>1</sup>, Sandro Athaide Coelho<sup>3</sup>, Sören Auer<sup>4</sup>, and Andreas Both<sup>2</sup>

<sup>1</sup> University of Leipzig, Germany, <sup>2</sup> R&D, Unister GmbH, Germany, <sup>3</sup> Federal University of Juiz de Fora, Brazil, <sup>4</sup> University of Bonn & Fraunhofer IAIS, Germany  
email: {usbeck|ngonga}@informatik.uni-leipzig.de

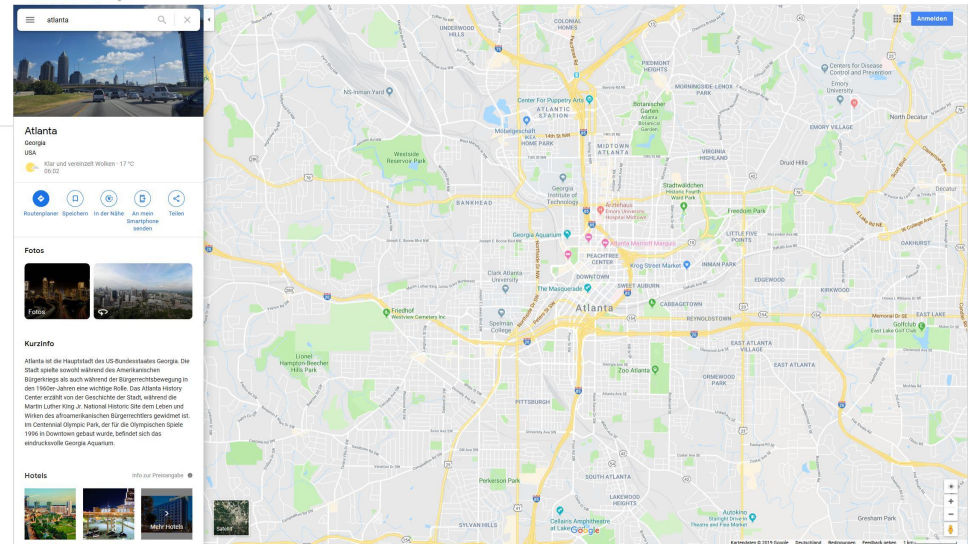
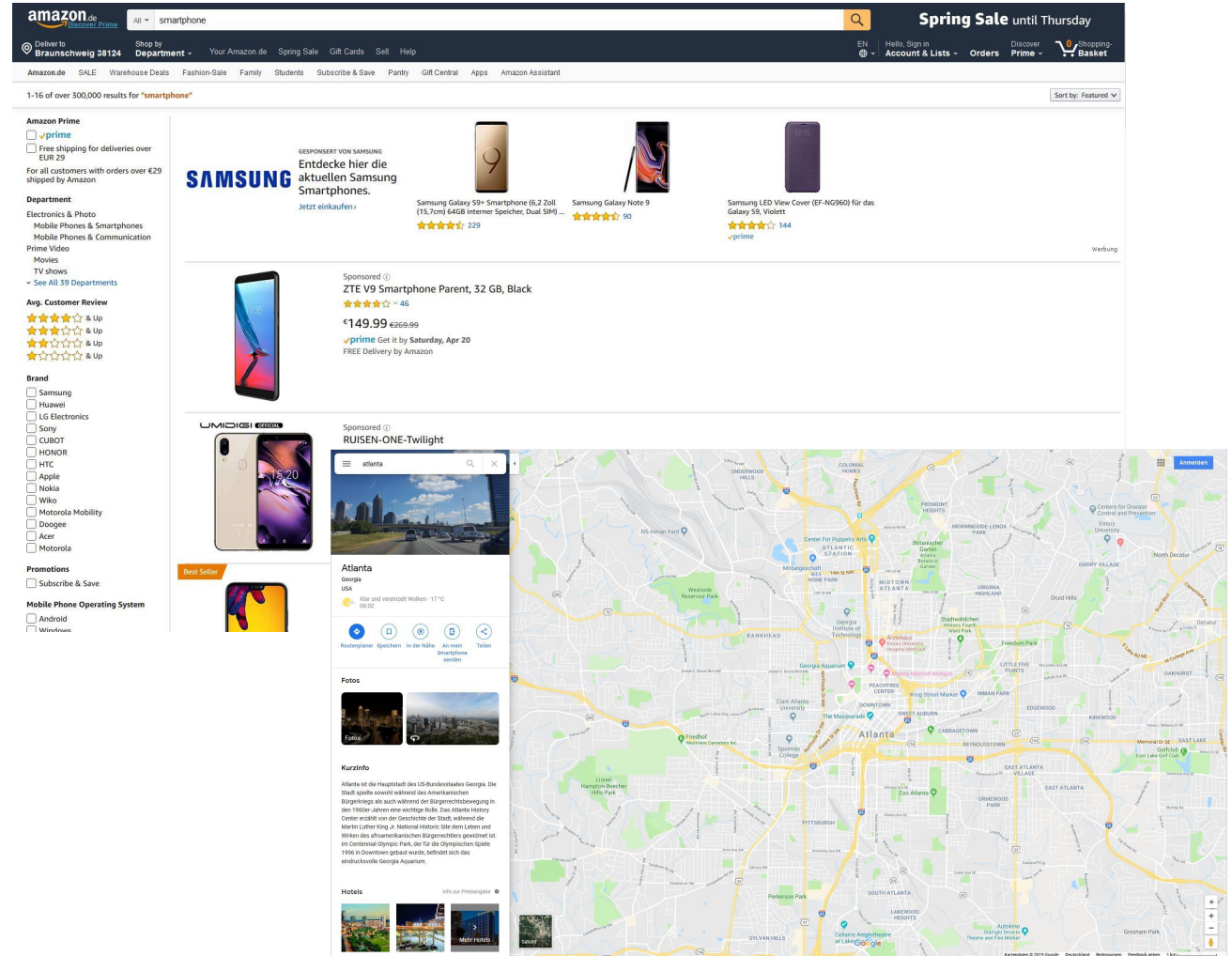
**Abstract.** Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Web of Data yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework) from these websites. One of the key steps towards extracting RDF from text is the disambiguation of named entities. While several approaches aim to tackle this problem, they still achieve poor accuracy. We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation. Our approach combines the Hypertext-Induced Topic Search (HITS) algorithm with label expansion strategies and string similarity measures. Based on this combination, AGDISTIS can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on eight different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 29% F-measure.

#### 1 Introduction

The vision behind the Web of Data is to provide a new machine-readable layer to the Web where the content of Web pages is annotated with structured data (e.g., RDFa [1]). However, the Web in its current form is made up of at least 15 billion Web pages.<sup>1</sup> Most of these websites are unstructured in nature. Realizing the vision of a usable and up-to-date Web of Data thus requires scalable and accurate natural-language-processing approaches that allow extracting RDF from such unstructured data. Three tasks play a central role when extracting RDF from unstructured data: named entity recognition (NER), named entity disambiguation (NED), also known as entity linking [16], and relation extraction (RE). For the first sentence of Example 1, an accurate named entity recognition approach would return the strings *Barack Obama* and *Washington, D.C.*. A high-quality DBpedia-based named entity disambiguation (NED) approach would use these already recognized named entities and map the strings

<sup>1</sup> Data gathered from <http://www.worldwidewebsitesize.com/> on January 4th, 2014.

# Digitalization elsewhere



## Open Research Knowledge Graph

- Digital library for machine actionable scientific knowledge communicated in scholarly literature
- “Deep Content Analysis”: Not just bibliographic metadata and beyond keywords
- Multimodality with crowdsourcing, text mining, and virtual research environments
- At time of creating knowledge as well as writing, submitting, publishing, reading articles
- There is a public alpha version at <https://labs.tib.eu/orkg/>
- API documentation at <https://labs.tib.eu/orkg/doc/api/>
- Code is Open Source and available at <https://gitlab.com/TIBHannover/orkg>

# Example



European Heart Journal (2017) 38, 362–372  
doi:10.1093/eurheartj/ehw333

## Iron-regulatory proteins secure iron availability in cardiomyocytes to prevent heart failure

Saba Haddad<sup>1,2</sup>, Yong Wang<sup>1,2</sup>, Bruno Galy<sup>3,4</sup>, Mortimer Korf-Klingebiel<sup>1</sup>, Valentin Hirsch<sup>1,2</sup>, Abdul M. Baru<sup>1,2</sup>, Fatemeh Rostami<sup>1,2</sup>, Marc R. Reboldi<sup>1</sup>, Jörg Heineke<sup>2</sup>, Ulrich Flögel<sup>5</sup>, Stephanie Groos<sup>6</sup>, André Renner<sup>7</sup>, Karl Tölg<sup>8</sup>, Fabian Zimmermann<sup>9</sup>, Stefan Engeli<sup>10</sup>, Jens Jordan<sup>10</sup>, Johann Bauersachs<sup>2</sup>, Matthias W. Hentze<sup>3</sup>, Kai C. Wollert<sup>1,2\*</sup>, and Tibor Kempf<sup>1,2\*</sup>

<sup>1</sup>Division of Molecular and Translational Cardiology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; <sup>2</sup>Department of Cardiology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; <sup>3</sup>European Molecular Biology Laboratory, Meyerhofstraße 1, 69117 Heidelberg, Germany; <sup>4</sup>Division of Virus-associated Carcinogenesis, German Cancer Research Centre, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; <sup>5</sup>Department of Cardiology, University of Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany; <sup>6</sup>Institute of Cell Biology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany; <sup>7</sup>Department of Thoracic and Cardiovascular Surgery, University of Bochum, Georgstraße 11, 32545 Bad Oeynhausen, Germany; <sup>8</sup>Department of Cardiology, University of Göttingen, Robert-Koch-Straße 40, 37075 Göttingen, Germany; <sup>9</sup>Department of Analytical Chemistry, Leibniz University Hannover, 30167 Hannover, Germany; and <sup>10</sup>Institute of Clinical Pharmacology, Hannover Medical School, Carl-Neuberg-Straße 1, 30625 Hannover, Germany

Received 30 November 2015; revised 27 June 2016; accepted 12 July 2016; online publish-ahead-of-print 21 August 2016

See page 373 for the editorial comment on this article (doi: 10.1093/eurheartj/ehw386)

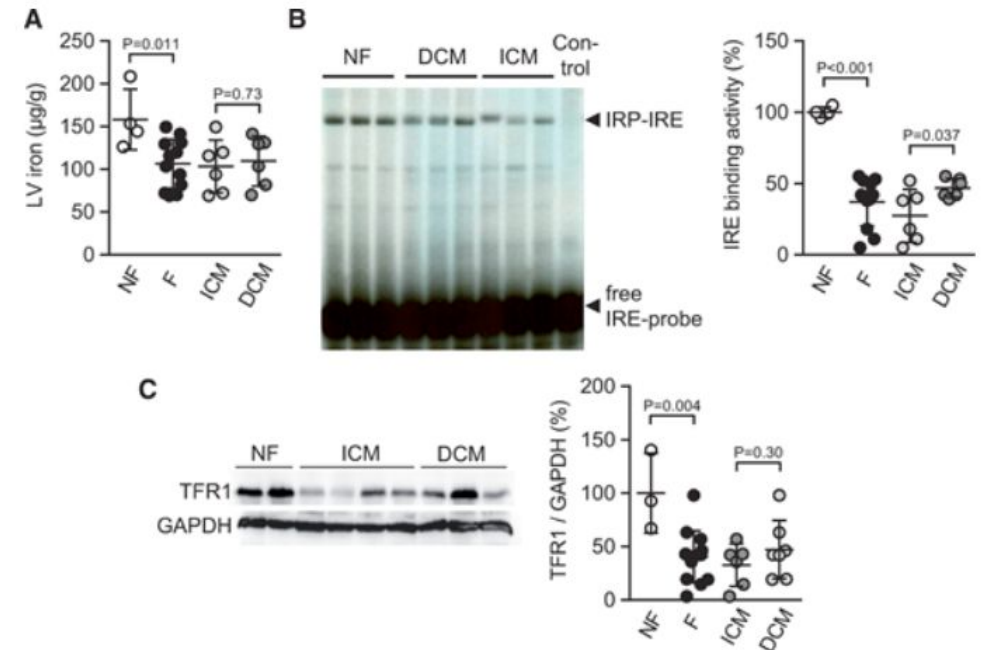
BASIC

## Results

### Reduced iron content, IRE binding activity, and transferrin receptor expression in the failing human heart

Consistent with previous reports,<sup>5,6</sup> iron concentration was significantly lower in LV tissue samples from patients with advanced heart failure than in LV tissue samples from unused donor hearts (Figure 1A). As shown by electrophoretic mobility shift assays, IRE binding activity was significantly reduced in failing hearts (most pronounced in patients with ischemic cardiomyopathy) (Figure 1B). Protein expression levels of the transferrin receptor were significantly lower in failing hearts than in the controls (Figure 1C).

Figure 1





orkg.ipynb



Code



Python 3



```
In [2]: labels = ['non-failing heart', 'failing heart']
data = [(99, 52),
        (96, 40),
        (100, 38),
        (105, 18),
        (np.nan, 11),
        (np.nan, 5),
        (np.nan, 42),
        (np.nan, 55),
        (np.nan, 53),
        (np.nan, 39),
        (np.nan, 42),
        (np.nan, 50)]

d = pd.DataFrame.from_records(data, columns=labels)
t = ttest_ind(d['non-failing heart'],
             d['failing heart'],
             equal_var=False, nan_policy='omit')

store(represent(d, t))

t.pvalue
```

```
Out[2]: 1.3111247517411591e-08
```



## Add paper

1

General

2

Research field

3

Contributions

4

Finish

### General paper data

By DOI

Manually

Paper DOI 

10.1093/eurheartj/ehw333

Lookup

Next step

1

General

2

Research field

3

Contributions

4

Finish

## General paper data

By DOI

Manually

Paper DOI 

10.1093/eurheartj/ehw333

Lookup

## Lookup result

**Paper title:** Iron-regulatory proteins secure iron availability in cardiomyocytes to prevent heart failure

**Authors:** Saba Haddad, Yong Wang, Bruno Galy, Mortimer Korf-Klingebiel, Valentin Hirsch, Abdul M. Baru, Fatemeh Rostami, Marc R. Reboll, Jörg Heineke, Ulrich Flögel, Stephanie Groos, André Renner, Karl Toischer, Fabian Zimmermann, Stefan Engeli, Jens Jordan, Johann Bauersachs, Matthias W. Hentze, Kai C. Wollert, Tibor Kempf

**Publication date:** August 2016

Next step



## Select the research field

Arts and Humanities	Nutrition	Systems and Integrative Physiology
Social and Behavioral Sciences	Forestry and Forest Sciences	Exercise Physiology
Engineering	Entomology Food Science	Endocrinology
Physical Sciences & Mathematics	Animal Sciences	Comparative and Evolutionary Physiology
<b>Life Sciences</b>	<b>Physiology</b>	<b>Cellular and Molecular Physiology</b>
	Nursing Pharmacology, Toxicology and Environmental Health	

Previous step

Next step



## Specify research contributions

Contribution 1



+ Add another contribution

Research problems

Iron deficiency in heart failure patients

Contribution data

No values

+ Add property

Previous step

Next step

Contribution 1 


+ Add another contribution

Research problems 

Iron deficiency in heart failure patients 

Contribution data 

Yields  Delete 

Object  IRE Cancel Done

+ Add property

Statistically significant hypothesis test with IRE binding dependent variable on failing and non-failing hearts

Previous step

Next step




Paper has been added successfully

[View paper](#)



## Homepage

**i** The **Open Research Knowledge Graph** - or - ORKG aims to describe research papers and contributions in a structured manner. With ORKG research contributions become findable and comparable. In order to add your own research, or to contribute, [learn more](#) 

### ★ Browse by research field

Main » Life Sciences » Physiology » Cellular and Molecular Physiology

---

#### Cellular and Molecular Physiology papers

- [Iron-regulatory proteins secure iron availability in cardiomyocytes to prevent heart failure](#)

### + Recently added papers


Iron-regulatory proteins secure iron availability in cardiomyocytes to prevent heart failure

Saba Haddad

[More papers](#)


## View paper

# Iron-regulatory proteins secure iron availability in cardiomyocytes to prevent heart failure

 August 2016

 Cellular and Molecular Physiology


 Saba Haddad

 Yong Wang

 Bruno Galy


 Mortimer Korf-Klingebliel

 Valentin Hirsch

 Abdul M. Baru


 Fatemeh Rostami

 Marc R. Reboll


 Jörg Heineke

 Ulrich Flögel

 Stephanie Groos

 André Renner

 Karl Toischer


 Fabian Zimmermann


 Stefan Engell

 Jens Jordan

 Johann Bauersachs

 Matthias W. Hentze

 Kai C. Wollert


 Tibor Kempf

### Contribution 1

#### Research problems

Iron deficiency in heart failure patients

#### Contribution data

Yields: Statistically significant hypothesis test with IRE binding dependent variable on failing and non-failing hearts 

#### Similar contributions [Show full comparison](#)

80% Wiles's proof of Fermat's last theorem

54% Gruber's design of ontologies

14% Design criteria for ontologies



## Research problems

Add to comparison

Iron deficiency in heart failure patients

## Contribution data

[← Back](#)

Ma Statistically significant hypothesis test with IRE binding dependent variable on failin

Has specified output: *the p-value of the statistical hypothesis test*



Type: *two sample t-test with unequal variance*



Has specified input: *LSUC Dataset*



Label: *Statistically significant hypothesis test with IRE binding dependent variable on failing and non-failing hearts*



## View dataset: LSUC Dataset



Showing 24 observations :

Options

left ventricular tissue sample	iron-responsive element binding activity
Search...	Search...
non-failing heart	nan
non-failing heart	99.0
non-failing heart	nan
failing heart	11.0
non-failing heart	nan
non-failing heart	nan
failing heart	42.0
non-failing heart	nan
non-failing heart	96.0
non-failing heart	nan

Previous Page 1 of 3 10 rows Next

Has specified output: *the p-value of the statistical hypothesis test* Type: *two sample t-test with unequal variance*

Has specified input

[LSUC Dataset](#) Label: *Statistically significant hypothesis test with IRE binding dependent variable on failing and non-failing hearts*

## Similar contributions

Failed to connect to the similarity service, please try again later

# Data Structure Definition

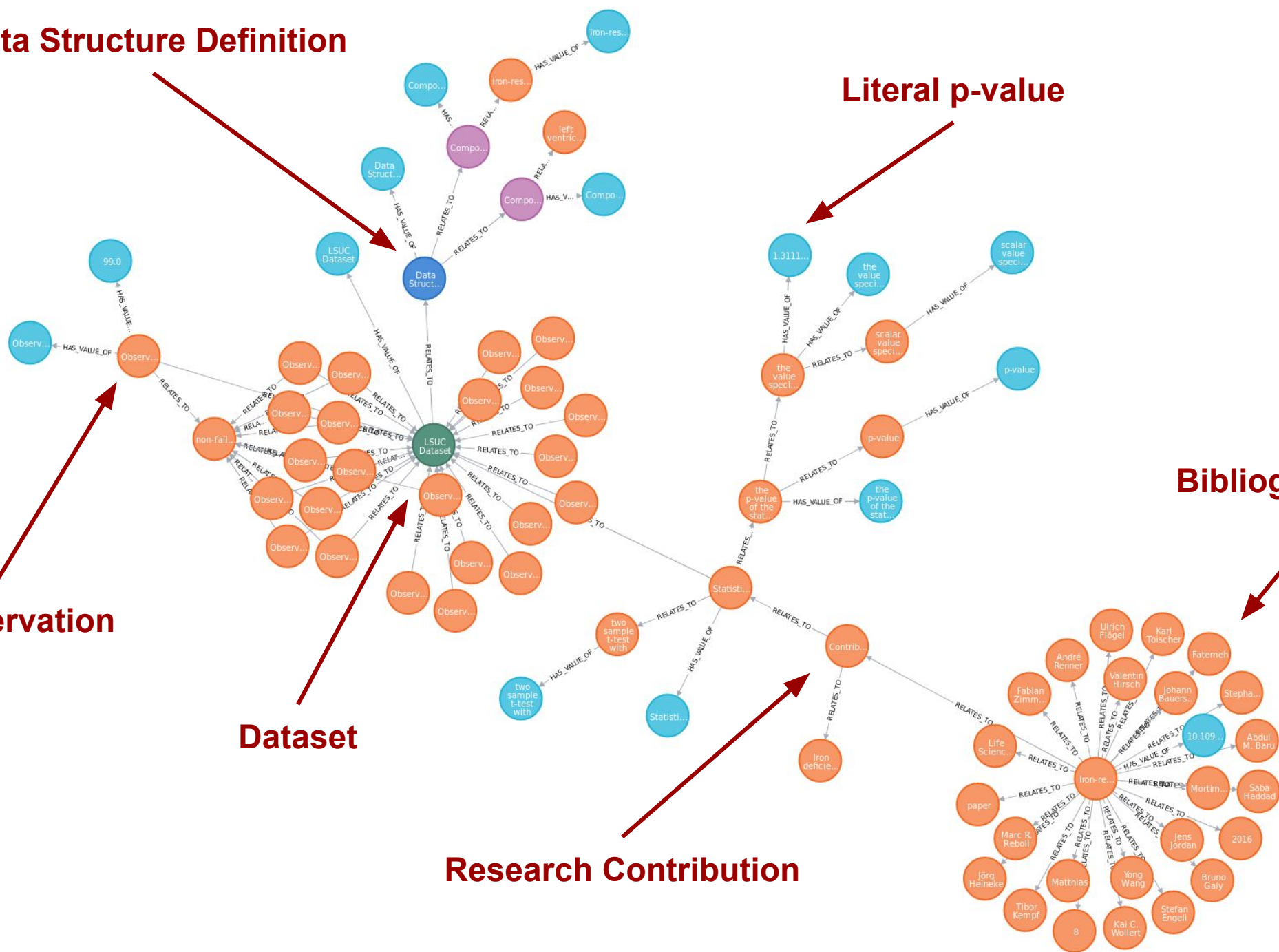
# Literal p-value

# Bibliographic Data

# An Observation

# Dataset

# Research Contribution



## Discussion

- Numerous challenges
  - What's the structure of scholarly knowledge
  - How granular can the knowledge model be to remain tractable
  - Implementing a multimodal infrastructure
  - Crowdsourcing relies on the “1-9-90 rule”: Does the rule apply to scholarly communication?
- Interlinking with other research/science graphs
  - PID Graph
  - OpenAIRE Explore
  - Research Graph
  - Springer Nature SciGraph
  - Microsoft Academic Graph
  - Others