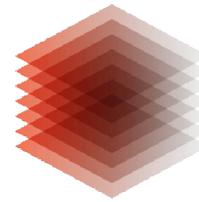


---

LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



TIB



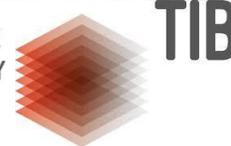
## From papers to knowledge: Representing scientific contributions in the Open Research Knowledge Graph

**Prof. Dr. Sören Auer**  
Leibniz University of Hannover  
TIB Technische Informationsbibliothek

11  
102  
1004

Leibniz  
Universität  
Hannover

LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



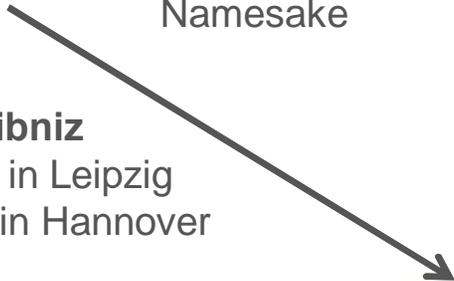
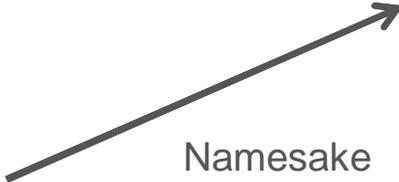
TIB



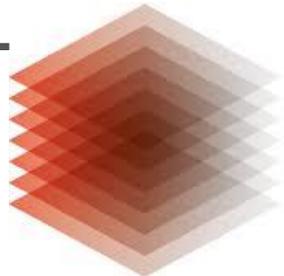
**Gottfried Wilhelm Leibniz**

\* 21. Juni/ 1. Juli 1646 in Leipzig

« 14. November 1716 in Hannover



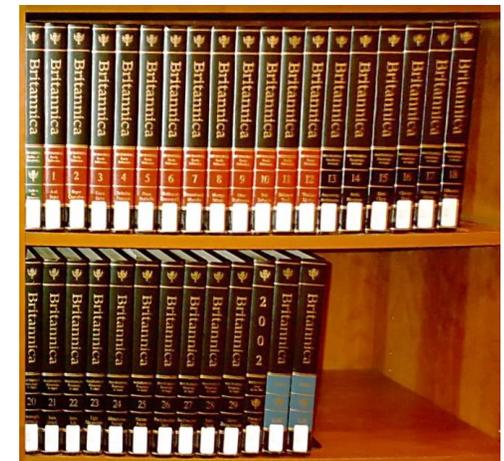
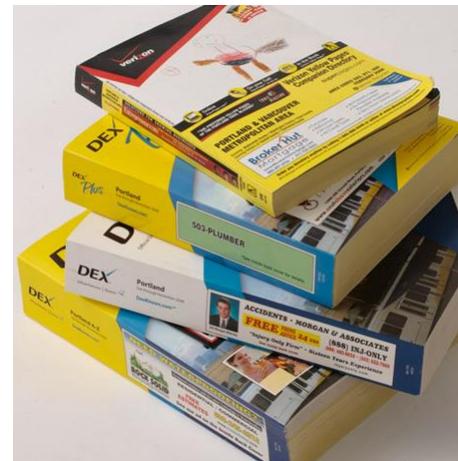
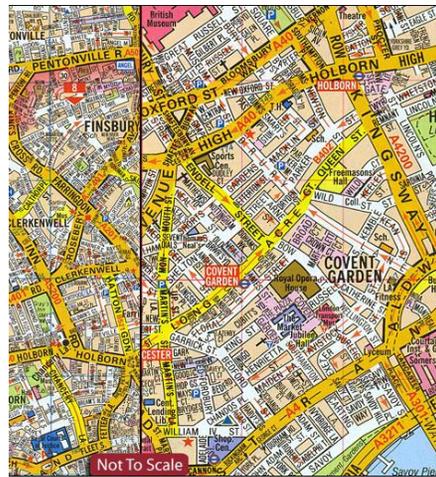
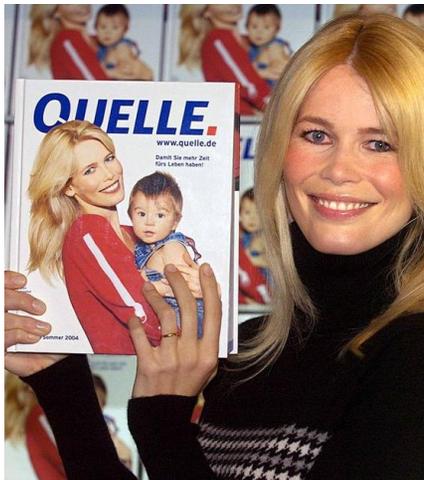
Member of



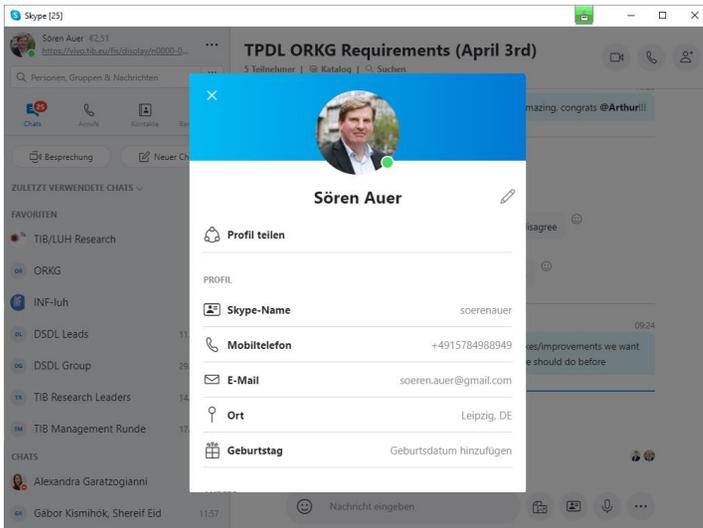
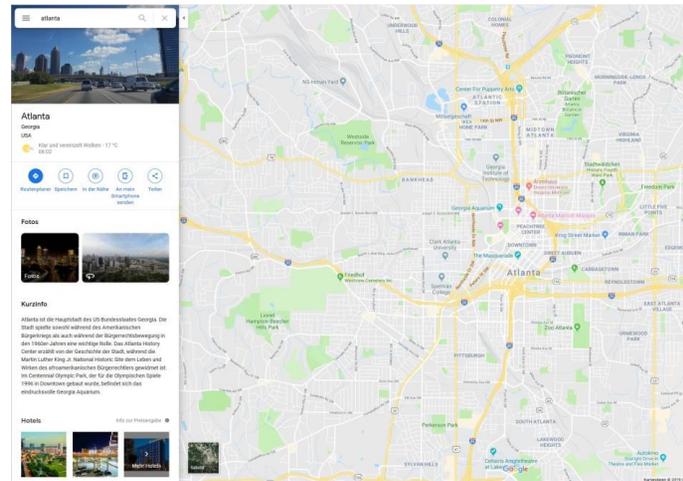
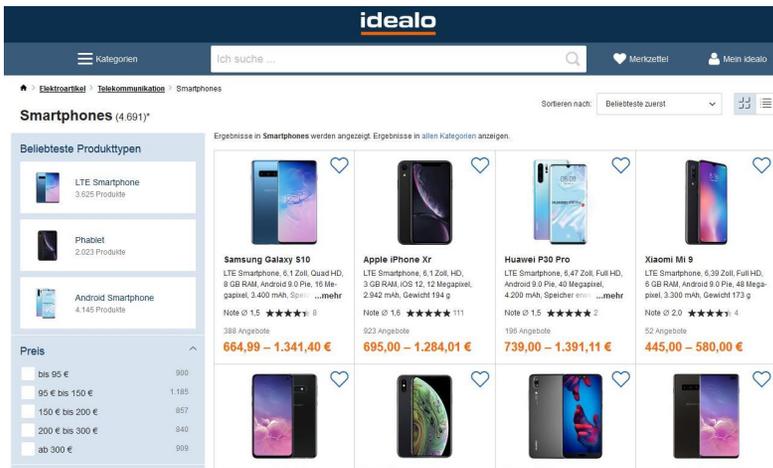
**TIB**



# How did information flows change in the digital era?



# How does it work today?



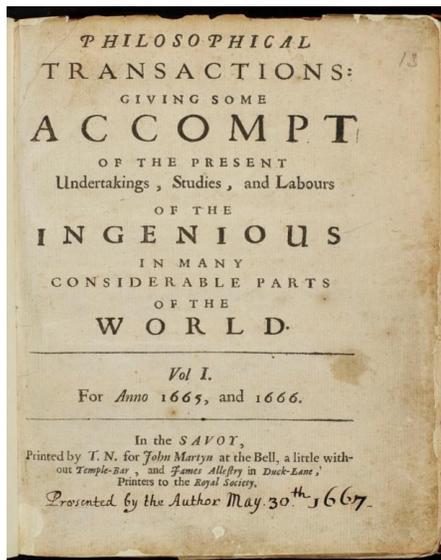
## The World of Publishing & Communication has profoundly changed

- “ New means adapted to the new possibilities were developed, e.g. zooming, dynamics
- “ **Business models** changed completely
- “ More focus on data, interlinking of **data / services and search** in the data
- “ Integration, **crowdsourcing, data curation** play an important role

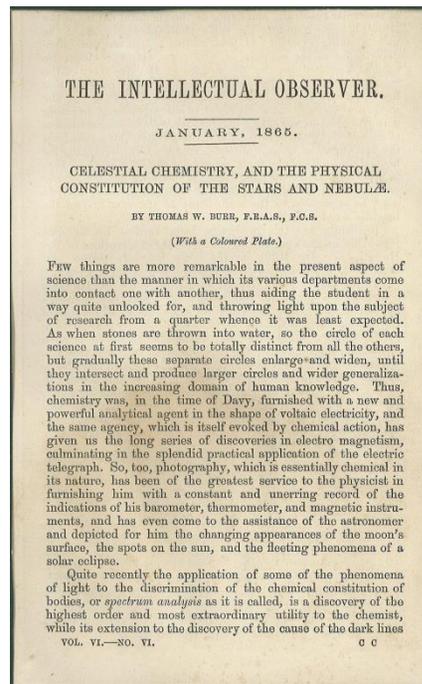
# What about Scholarly Communication?

# Scholarly Communication has not changed (much)

## 17<sup>th</sup> century



## 19<sup>th</sup> century



## 20<sup>th</sup> century



## 21<sup>st</sup> century



Meanwhile other information intense domains were completely disrupted: mail order catalogs, street maps, phone books, ã

## We need to rethink the way how research is represented and communicated

### Challenges we are facing:

Digitalisation of Science	Monopolisation by commercial actors	Reproducibility Crisis	Proliferation of publications	Deficiency of Peer Review
<ul style="list-style-type: none"><li>▪ Data integration and analysis</li><li>▪ Digital collaboration</li></ul>	<ul style="list-style-type: none"><li>▪ Publisher look-in effects</li><li>▪ Maximization of profits <sup>[1]</sup></li></ul>	<ul style="list-style-type: none"><li>▪ Majority of experiments are hard or not reproducible <sup>[2]</sup></li></ul>	<ul style="list-style-type: none"><li>▪ Publication output doubled within a decade</li><li>▪ continues to rise <sup>[3]</sup></li></ul>	<ul style="list-style-type: none"><li>▪ Deteriorating quality <sup>[4]</sup></li><li>▪ Predatory publishing</li></ul>

[1] <http://thecostofknowledge.com>, <https://www.projekt-deal.de>

[2] M. Baker: *1,500 scientists lift the lid on reproducibility*, *Nature*, 2016.

[3] *Science and Engineering Publication Output Trends*, National Science Foundation, 2018.

[4] J. Couzin-Frankel: *Secretive and Subjective, Peer Review Proves Resistant to Study*, *Science*, 2013.

## Proliferation of scientific literature

### Science and engineering articles by region, country: 2004 and 2014

Rank	Region, country, or economy	2004	2014	Average annual growth rate (%)	2014 world total (%)	2014 cumulative world total (%)
na	World	1,272,362	2,290,294	6.1	100.0	na
1	United States	336,194	431,623	2.5	18.8	18.8
2	China	110,388	395,588	13.6	17.3	36.1
3	Germany	72,177	107,747	4.1	4.7	40.8
4	India	28,752	106,574	14.0	4.7	45.5
5	Japan	95,999	103,793	0.8	4.5	50.0
6	United Kingdom	75,119	101,536	3.1	4.4	54.4
7	France	53,375	74,269	3.4	3.2	57.7
8	Italy	42,647	70,453	5.1	3.1	60.8
9	South Korea	27,029	63,748	9.0	2.8	63.5
10	Canada	40,624	60,916	4.1	2.7	66.2
11	Spain	30,977	56,604	6.2	2.5	68.7
12	Brazil	18,814	53,152	10.9	2.3	71.0
13	Australia	26,277	52,269	7.1	2.3	73.3
14	Russia	26,869	43,487	4.9	1.9	75.2
15	Iran	4,952	36,539	22.1	1.6	76.8

## Reproducibility Crisis

### 1,500 scientists lift the lid on reproducibility

Monya Baker in *Nature*, 2016. **533** (7604): 452. 454.  
[doi:10.1038/533452a](https://doi.org/10.1038/533452a):

- 70% failed to reproduce at least one other scientist's experiment
- 50% failed to reproduce one of their own experiments

### Failure to reproduce results among disciplines

(in brackets own results)

chemistry	87%	(64%)
biology	77%	(60%)
physics and engineering	69%	(51%)
Earth sciences	64%	(41%)

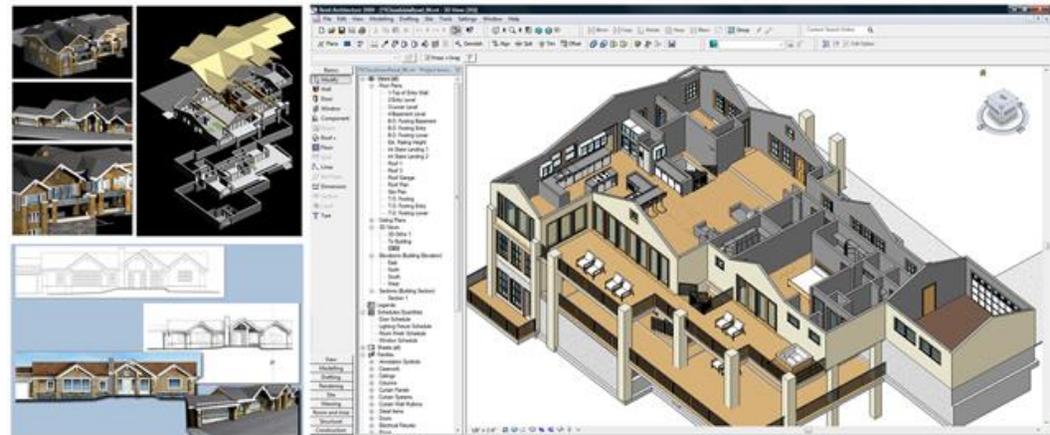


Source: © Stanford Medicine - Stanford University

## Duplication and Inefficiency

How can we avoid duplication if the terminology, research problems, approaches, methods, characteristics, evaluations, etc. are not properly defined and identified?

How would you build an engine / building without properly defining their parts, relationships, materials, characteristics etc. ?



Source: [https://thumbs.worthpoint.com/zoom/images2/1/0316/22/revell-4-visible-8-engine-plastic\\_1\\_d2162f52c3fa3a6f72d2722f6c50b7b2.jpg](https://thumbs.worthpoint.com/zoom/images2/1/0316/22/revell-4-visible-8-engine-plastic_1_d2162f52c3fa3a6f72d2722f6c50b7b2.jpg)

Source: <http://xnewlook.com/cad-and-revit-3d-design.html/bill-ferguson-portfolio-computer-graphics-games-cad-related-3d-models-cad-and-revit-design>

## Root Cause – Deficiency of Scholarly Communication?

### Lack of...

#### Transparency

information is hidden  
in text

#### Integratability

fitting different  
research results  
together

#### Machine assistance

unstructured content  
is hard to process

#### Identifiability

of concepts beyond  
metadata

#### Collaboration

one brain barrier

#### Overview

Scientists look for the  
needle in the haystack

CRISPR

Nur im Bibliothekskatalog der TIB suchen



**Search for CRISPR:  
> 9.000 Results**

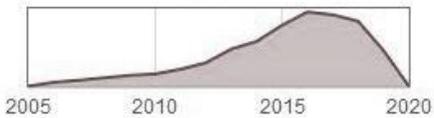
Hier geht es zum klassischen Katalog [↗](#)

Seite 1 von 9.881 Ergebnissen

Sortieren nach: Relevanz | [Aktualität](#) | [Titel](#)

### Treffer filtern

#### Erscheinungsjahr


 -  

#### Medientyp

- Aufsatz (Zeitschrift) (6.870)
- Patent (2.619)
- Hochschulschrift (127)
- Sonstige (82)
- Aufsatz (Konferenz) (58)

+ Mehr anzeigen

Source: [https://www.tib.eu/de/suchen/?id=198&tx\\_tibsearch\\_search%5Bquery%5D=CRISPR&tx\\_tibsearch\\_search%5Bsr%5D=rk&tx\\_tibsearch\\_search%5Bcnt%5D=20,04.2019](https://www.tib.eu/de/suchen/?id=198&tx_tibsearch_search%5Bquery%5D=CRISPR&tx_tibsearch_search%5Bsr%5D=rk&tx_tibsearch_search%5Bcnt%5D=20,04.2019)

Für folgende Begriffe wurden Synonyme verwendet: [CRISPR](#)  
Suche ohne Synonyme: [CRISPR](#)

« < 1 2 3 4 5 6 7 > »



### CRISPR human trial.(clustered regularly interspaced short palindromic repeats)

Online Contents | 2016

cited: 0 | score: 310.36673



### Advances in clustered regularly interspaced short palindromic repeats - A review

Wang, L. / He, J. / Wang, J. | British Library Online Contents | 2011

cited: 0 | score: 297.1446

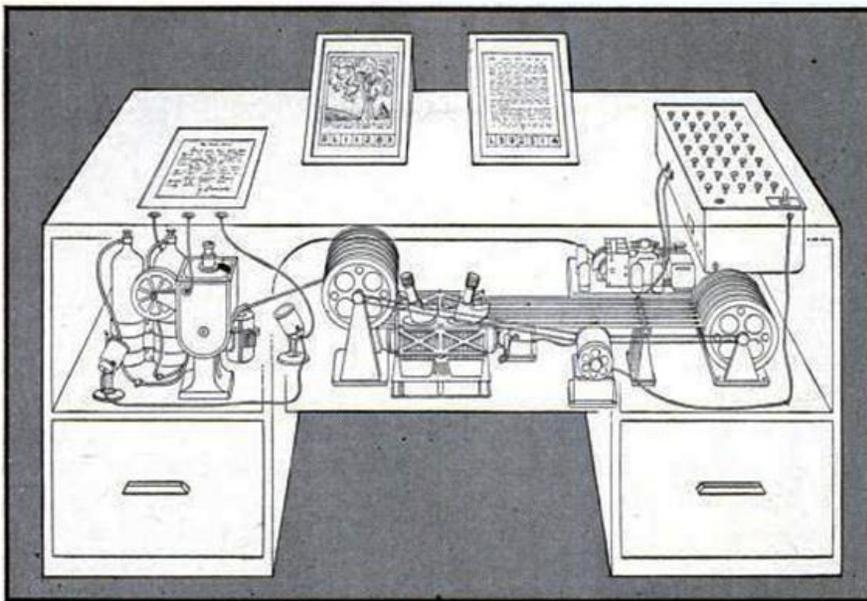


### regularly interspaced short palindromic repeats

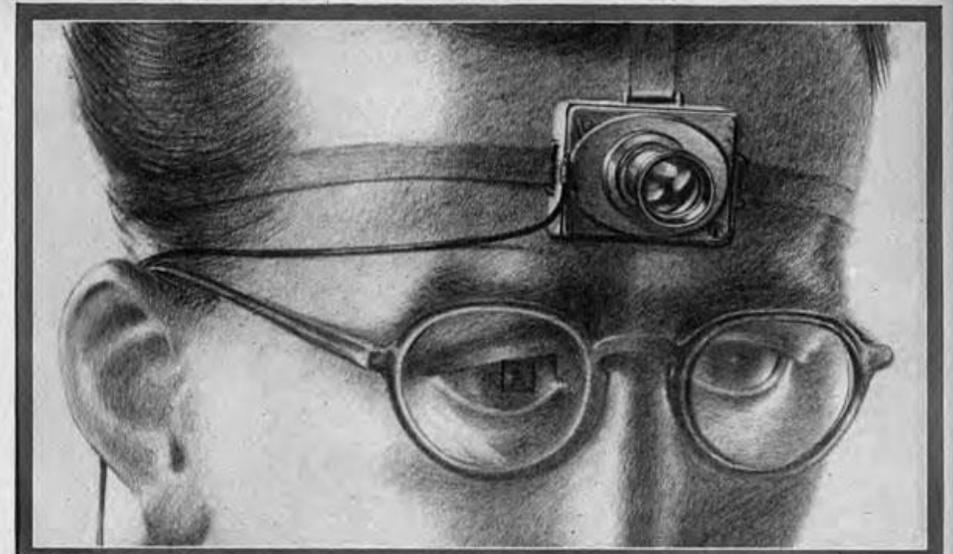


**How can  
we fix it?**

## Realizing Vannevar Bush's vision of Memex



Source: <http://photos1.blogger.com/blogger/5874/1071/1600/Memex.jpg>



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBJECT

# AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD  
IN WHICH MAN-MADE MACHINES WILL START TO THINK

by VANNEVAR BUSH

DIRECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT  
Condensed from the Atlantic Monthly, July 1945

This has not been a scientists' war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indecision, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who have had to devise new methods for their unanticipated assignments. They have done their best on the devices that made it possible to turn back

rews, and the effort to bridge between disciplines is correspondingly superficial.

Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. If the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time might well be startling. Those who conscientiously attempt to keep abreast of current thought, even in restricted fields, by close and continuous reading might well shy away from an examination calculated to show how much of the previous month's efforts could be produced on call.

Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of reading and understanding it. This sort of catastrophe is undoubtedly being repeated in many fields. Significant attainments become lost in the mass

been part of a great team. Now one asks where they will find objectives

Publication has been extended far beyond our present ability to make real

Source: <http://tntindex.blogspot.com/2014/10/tabletalk-vannevar-bushs-memex.html>

## Concepts

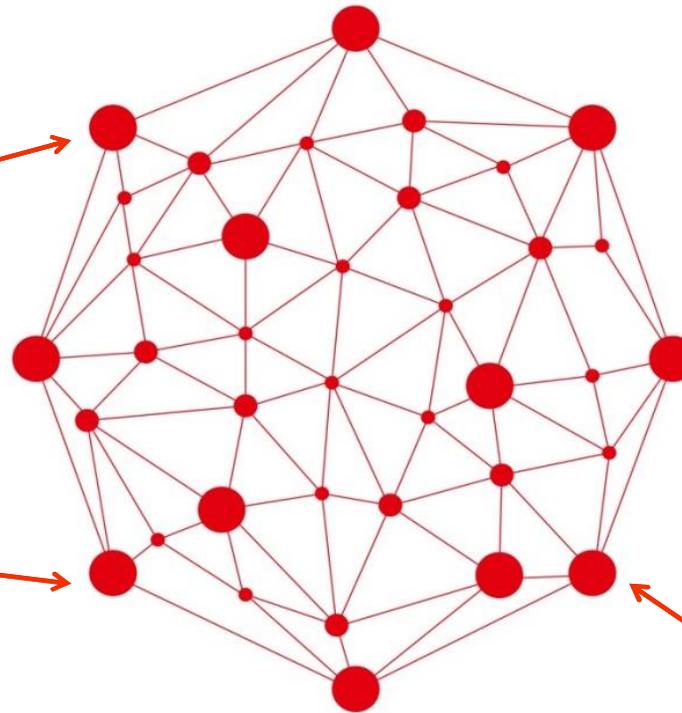
### Overarching Concepts

- Research problems
- Definitions
- Research approaches
- Methods

### Artefacts

- Publications
- Data
- Software
- Image/Audio/Video
- Knowledge Graphs / Ontologies

### Domain specific Concepts



Mathematics	Physics	Chemistry	Computer Science	Technology	Architecture
<ul style="list-style-type: none"> <li>" Definitions</li> <li>" Theorems</li> <li>" Proofs</li> <li>" Methods</li> <li>" ð</li> </ul>	<ul style="list-style-type: none"> <li>" Experiments</li> <li>" Data</li> <li>" Models</li> <li>" ð</li> </ul>	<ul style="list-style-type: none"> <li>" Substances</li> <li>" Structures</li> <li>" Reactions</li> <li>" ð</li> </ul>	<ul style="list-style-type: none"> <li>" Concepts</li> <li>" Implementations</li> <li>" Evaluations</li> <li>" ð</li> </ul>	<ul style="list-style-type: none"> <li>" Standards</li> <li>" Processes</li> <li>" Elements</li> <li>" Units, Sensor data</li> </ul>	<ul style="list-style-type: none"> <li>" Regulations</li> <li>" Elements</li> <li>" Models</li> <li>" ð</li> </ul>

## Linked Data Principles



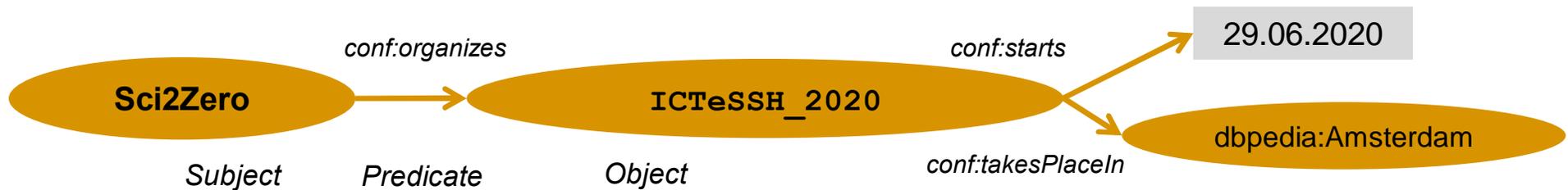
### Addressing the neglected third V (Variety)

1. Use **URIs** to **identify** the things in your data
2. Use **http:// URIs** so people (and machines) can **look** them **up** on the web
3. When a URI is looked up, **return a description** of the thing **in** the W3C **Resource Description Format (RDF)**
4. Include **links to related things**

<http://www.w3.org/DesignIssues/LinkedData.html>

## RDF & Linked Data in a Nutshell

1. Graph based RDF data model consisting of S-P-O statements (facts)



2. Serialised as RDF Triples:

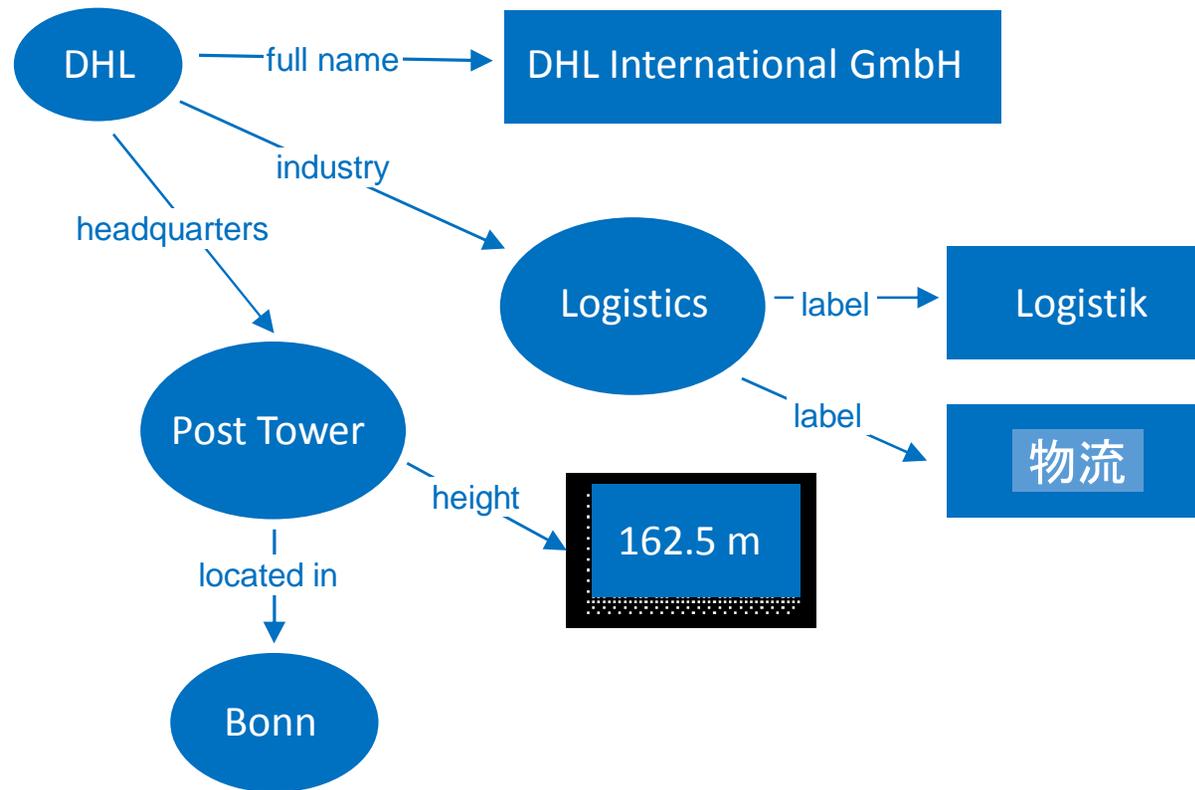
```

Sci2Zero          conf:organizes      ICTeSSH_2020 .
ICTeSSH_2020     conf:starts          "2020-06-29"^^xsd:date .
ICTeSSH_2020     conf:takesPlaceAt    dbpedia:Amsterdam .
  
```

3. Publication under URL in Web, Intranet, Extranet

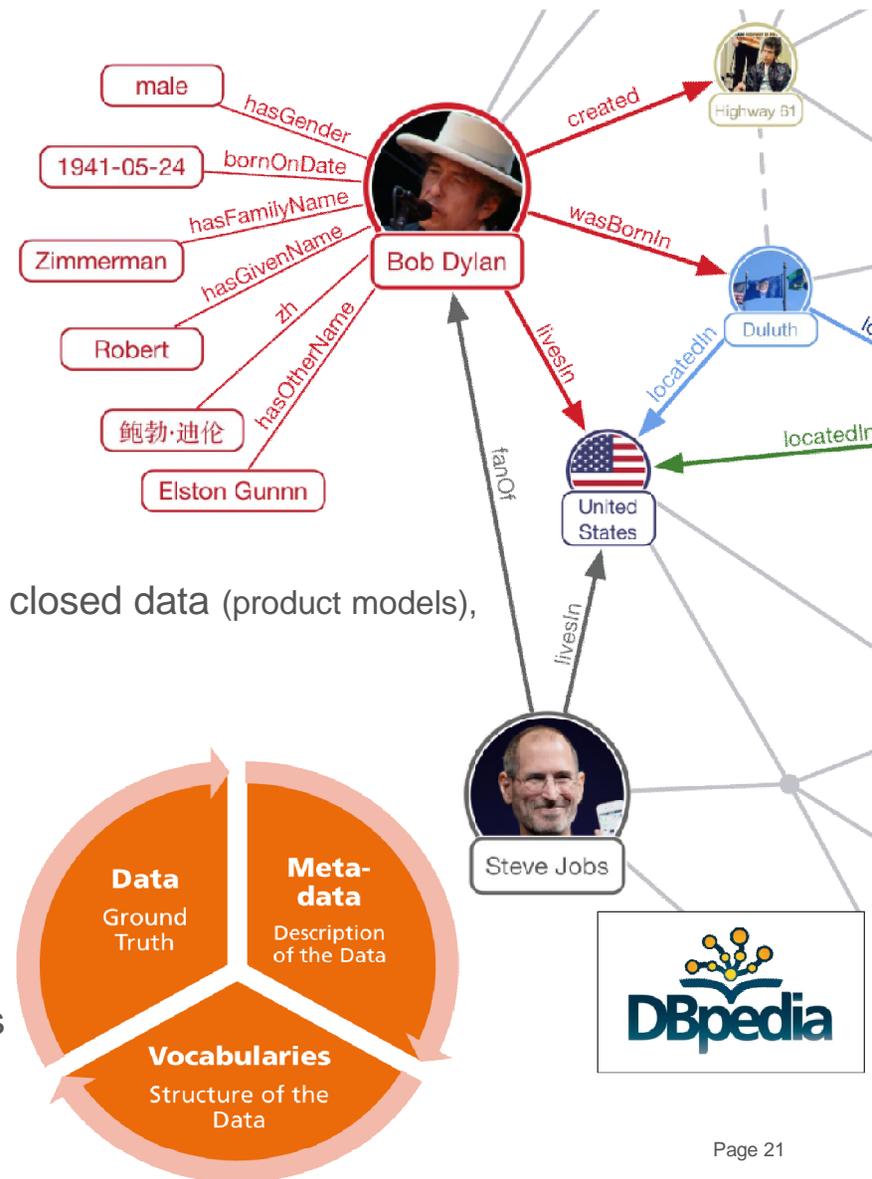
# Linked Data

## Creating Knowledge Graphs with RDF



## Knowledge Graphs – A definition

- Fabric of concept, class, property, relationships, entity desc.
- Uses a knowledge representation formalism (RDF, OWL)
- Holistic knowledge (multi-domain, source, granularity):
  - **instance data** (ground truth),
    - open (e.g. DBpedia, WikiData), private (e.g. supply chain data), closed data (product models),
  - derived, aggregated data,
  - **schema data** (vocabularies, ontologies)
  - **meta-data** (e.g. provenance, versioning, documentation licensing)
  - comprehensive **taxonomies** to categorize entities
  - **links** between internal and external data
  - **mappings** to data stored in other systems and databases





# Industry Knowledge Graph Adoption

<https://www.slideshare.net/Frank.van.Harmelen/adoption-of-knowledge-graphs-late-2019>

# Chemistry Example: CRISPR Genome Editing





**bioRxiv**  
THE PREPRINT SERVER FOR BIOLOGY

[HOME](#) | [ABOUT](#) | [SUBMIT](#) | [ALERTS / RSS](#)  
[CHANNELS](#)

  
[Advanced Search](#)

New Results

**A practical guide to CRISPR/Cas9 genome editing in Lepidoptera**

Linlin Zhang,  Robert Reed  
**doi:** <https://doi.org/10.1101/130344>  
Now published in *Diversity and Evolution of Butterfly Wing Patterns* doi: 10.1007/978-981-10-4956-9\_8

[Abstract](#) | [Info/History](#) | [Metrics](#) | [Preview PDF](#)

**Abstract**

CRISPR/Cas9 genome editing has revolutionized functional genetic work in many organisms and is having an especially strong impact in emerging model systems. Here we summarize recent advances in applying CRISPR/Cas9 methods in Lepidoptera, with a focus on providing practical advice on the entire process of genome editing from experimental design through to genotyping. We also describe successful targeted GFP

[Previous](#) [Next](#)

Posted June 22, 2017.

[Download PDF](#) | [Share](#)  
[Email](#) | [Citation Tools](#)

[Tweet](#) | [Like 0](#) | [G+](#)

**Subject Area**

[Genetics](#)

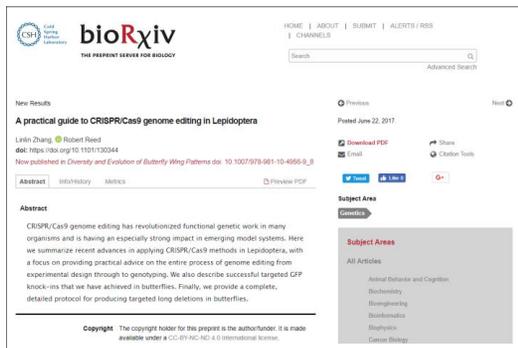
**Subject Areas**

[All Articles](#)

Source: [https://cacm.acm.org/system/assets/0002/2618/021116\\_Google\\_KnowledgeGraph.large.jpg?1476779500&1455222197](https://cacm.acm.org/system/assets/0002/2618/021116_Google_KnowledgeGraph.large.jpg?1476779500&1455222197)

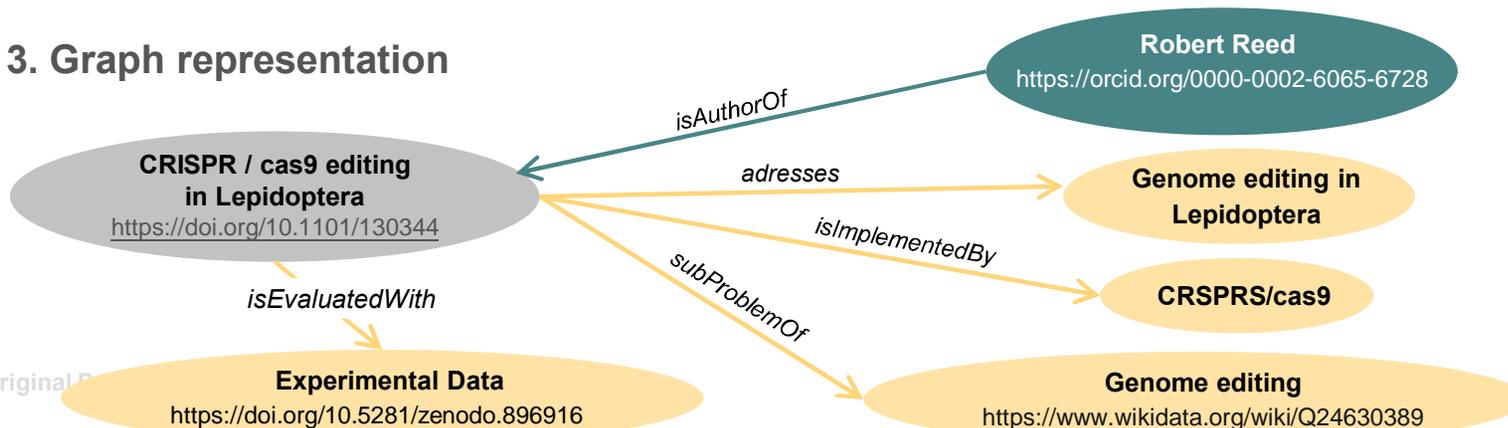
# Chemistry Example: Populating the Graph

## 2. Adaptive Graph Curation & Completion



<b>Author</b>	Robert Reed
<b>Research Problem</b>	Genome editing in Lepidoptera
<b>Methods</b>	CRISPR / cas9
<b>Applied on</b>	Lepidoptera
<b>Experimental Data</b>	<a href="https://doi.org/10.5281/zenodo.896916">https://doi.org/10.5281/zenodo.896916</a>

## 3. Graph representation



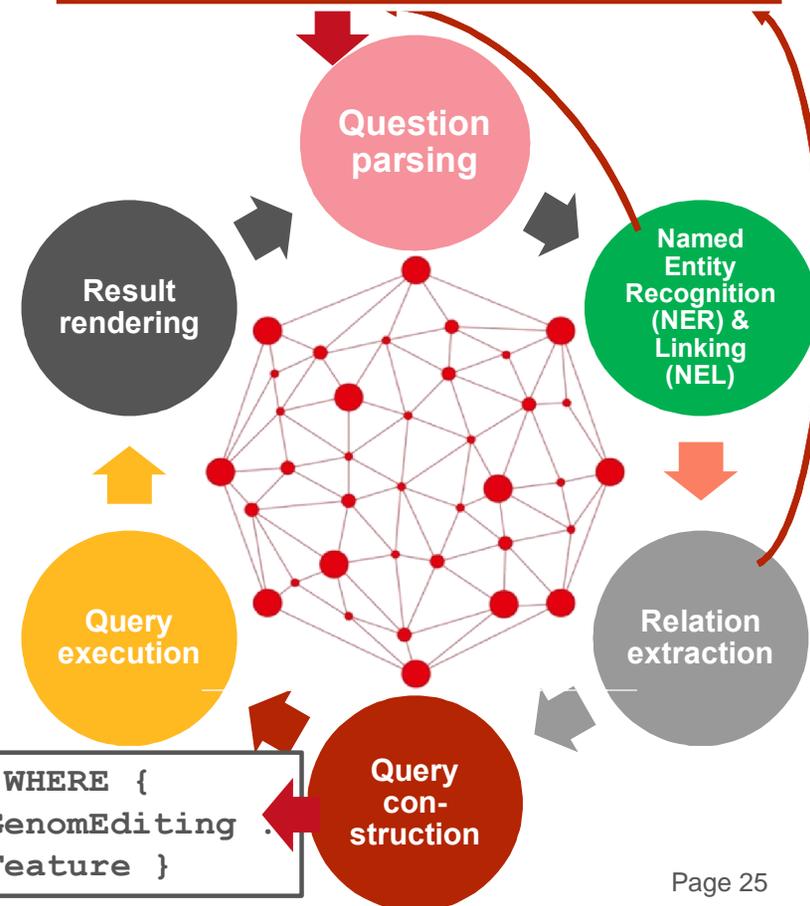
1. Original P

# Exploration and Question Answering

## Research Challenge:

- “ Intuitive exploration leveraging the rich semantic representations
- “ Answer natural language questions

Q: How do different genome editing techniques compare?



```

SELECT Approach, Feature WHERE {
  Approach addresses GenomEditing .
  Approach hasFeature Feature }
  
```

[1] K. Singh, S. Auer et al: *Why Reinvent the Wheel? Let's Build Question Answering Systems Together.* The Web Conference (WWW 2018).

# Result:

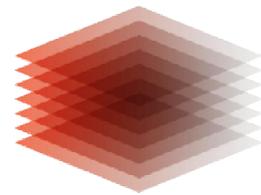
## Automatic Generation of Comparisons / Surveys

Q: How do different **genome editing** techniques **compare**?

Engineered Nucleases	Site-specificity	Safety	Ease-of-use / costs/ speed
zinc finger nucleases (ZFN)	++ 9-18nt	+	-- \$\$\$: screening, testing to define efficiency
transcription activator-like effector nucleases (TALENs)	+++ 9-16nt	++	++ Easy to engineer 1 week / few hundred dollar
engineered meganucleases	+++ 12-40 nt	0	-- \$\$\$ Protein engineering, high-throughput screening
CRISPR system/cas9	++ 5-12 nt	-	+++ Easy to engineer few days / less 200 dollar

---

LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



**TIB**

# **Demo: Open Research Knowledge Graph**

Properties	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 1 - 2020	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 2 - 2020	Estimation of the epidemic properties of the 2019 novel coronavirus: A mathematical modeling study Contribution 3 - 2020	Transmission potential of COVID-19 in Iran Contribution 1 - 2020	Transmission potential of COVID-19 in Iran Contribution 2 - 2020	Estimating the generation interval for COVID-19 based on symptom onset data Contribution 1 - 2020
Has research problem	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number	COVID-19 reproductive number
Location	Wuhan City, China	Wuhan City, China	Wuhan City, China	Iran	Iran	Singapore
Study date	2020-01-10/2020-01-23	2020-01-23/2020-02-08	2020-01-10/2020-02-08	2020-02-19/2020-02-29	2020-02-19/2020-02-29	2020-01-21/2020-02-26
R0 estimates (average)	4.38	3.41	3.39	3.6	3.58	1.27
95% confidence interval	3.63-5.13	3.16-3.65	3.09-3.70	3.2-4.2	1.29-8.46	1.19-1.36
Method	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	a weighted average of Exponential growth, Maximum likelihood, Sequential Bayesian, Time-dependent reproduction numbers, and SEIR model basic reproduction numbers by calculating weights from a Poisson loss function	generalized growth model	based on the calculation of the epidemic's doubling times: estimated epidemic doubling time of 1.20 (95% CI, 1.05, 1.44) days	generation interval



## View paper

View paper

Graph view

Edit

## Transmission interval estimates suggest pre-symptomatic spread of COVID-19

March 2020

Virology

Lauren Tindale

Michelle Coombe

Jessica Stockdale

Emma Garlock

Wing Yin Venus Lau

Manu Saraswat

Yen-Hsiang Brian Lee

Louxin Zhang

Dongxuan Chen

Jacco Wallinga

Caroline Colijn

DOI: 10.1101/2020.03.03.20029983

Share this paper:  

Contribution 1

Contribution 2

## Research problems

 Add to comparison

COVID-19 reproductive number

## Contribution data

95% Confidence interval	1.45-2.48
Location	Singapore
Mean incubation period	7.1 (6.13, 8.25) days
Mean serial interval	4.56 (2.69, 6.42) days
R0 estimates (average)	1.97
Study date	2020-01-19/2020-02-26

Edit mode Every change you make is automatically saved

[View paper](#)[Graph view](#)[Stop editing](#)

-  March 2020
-  Virology
-  Lauren Tindale
-  Michelle Coombe
-  Jessica Stockdale
-  Emma Garlock
-  Wing Yin Venus Lau
-  Manu Saraswat
-  Yen-Hsiang Brian Lee
-  Louxin Zhang
-  Dongxuan Chen
-  Jacco Wallinga
-  Caroline Colijn

DOI: 10.1101/2020.03.03.20029983

[Edit data](#)Share this paper:  Contribution 1  

Contribution 2

[+](#)**Research problems** Add to comparisonCOVID-19 reproductive number **Contribution data**

95% Confidence interval 1.45-2.48

[+](#)

Location

Singapore

[+](#)

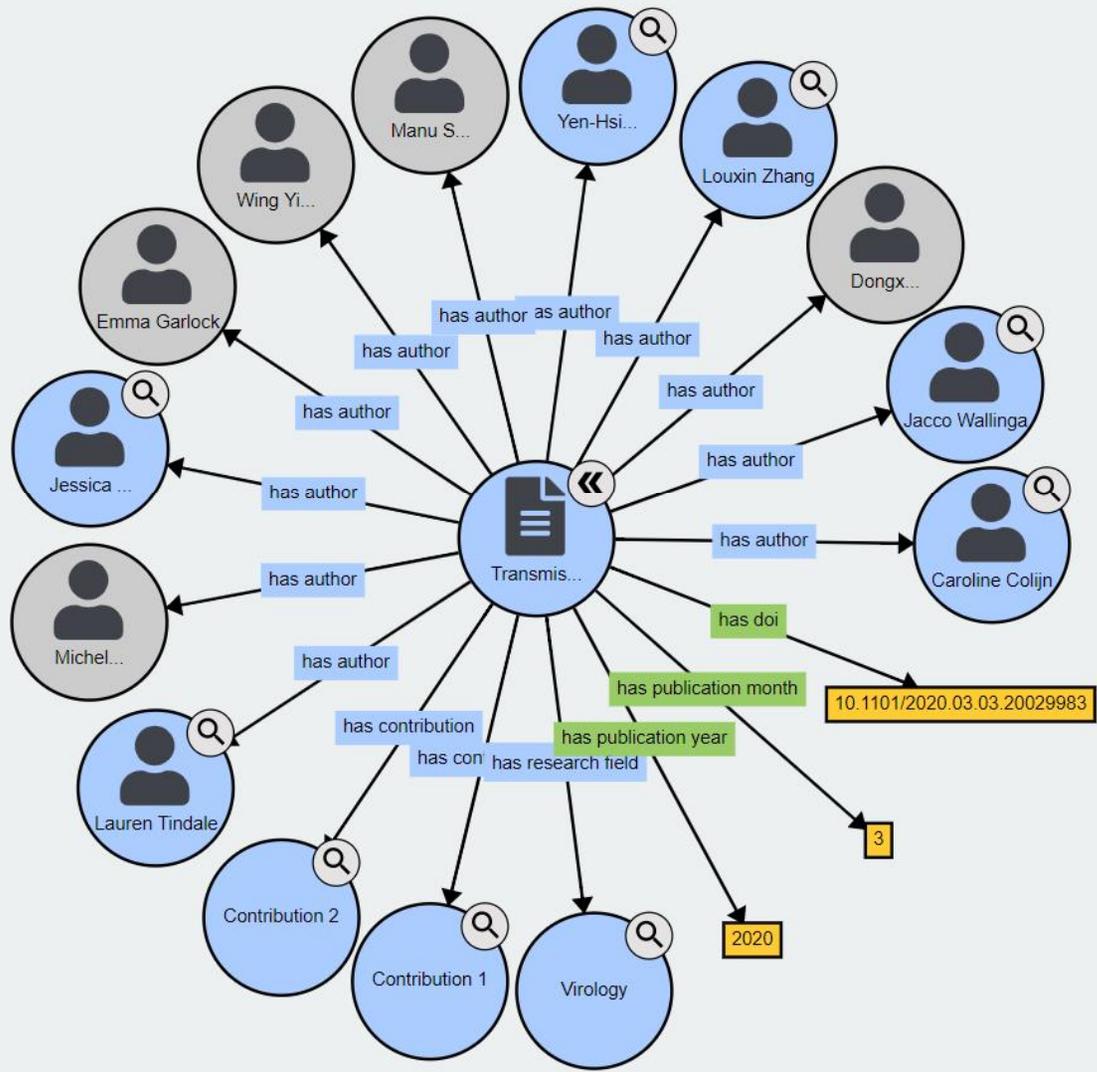
Mean incubation period

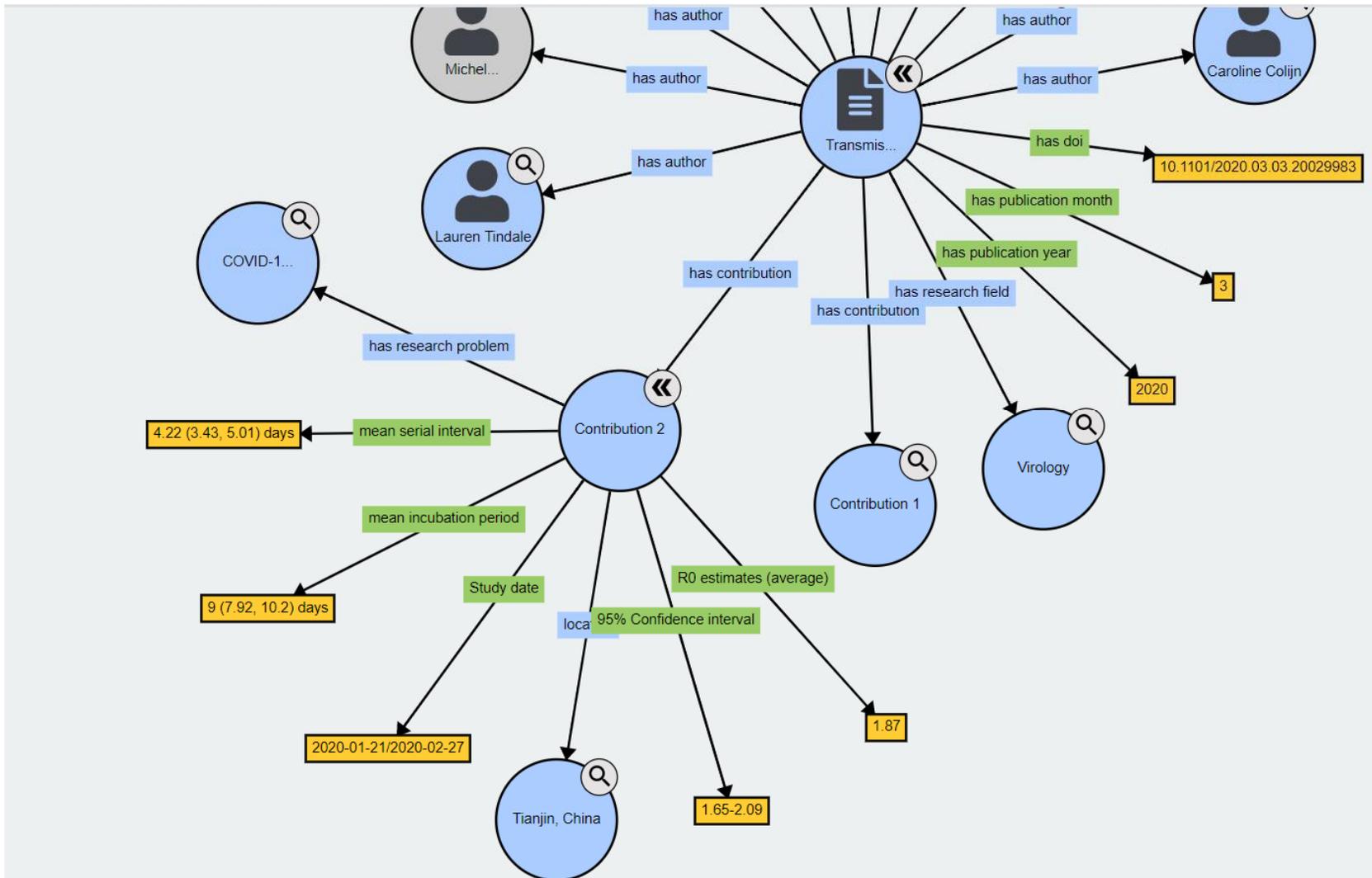
7.1 (6.13, 8.25) days

[+](#)[Edit value](#)

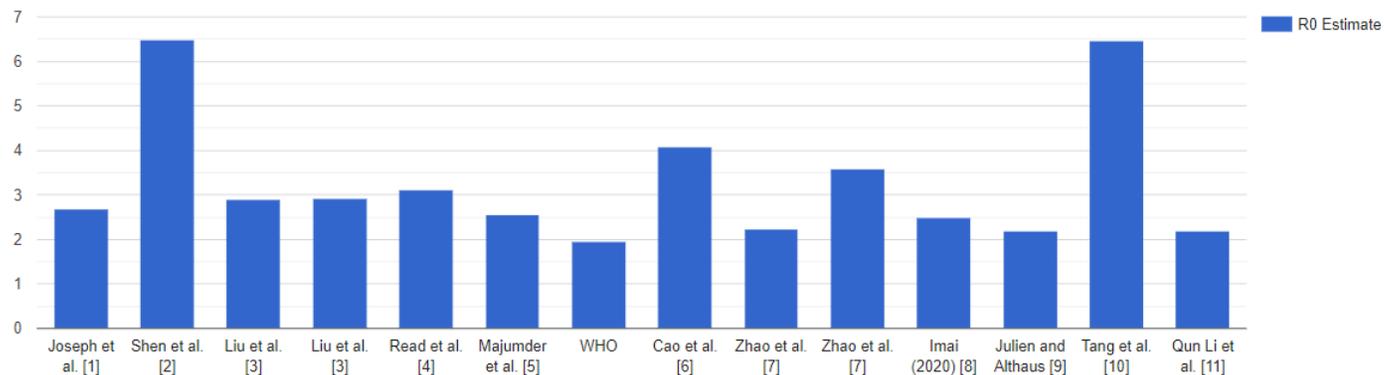
Mean serial interval

4.56 (2.69, 6.42) days  [+](#)

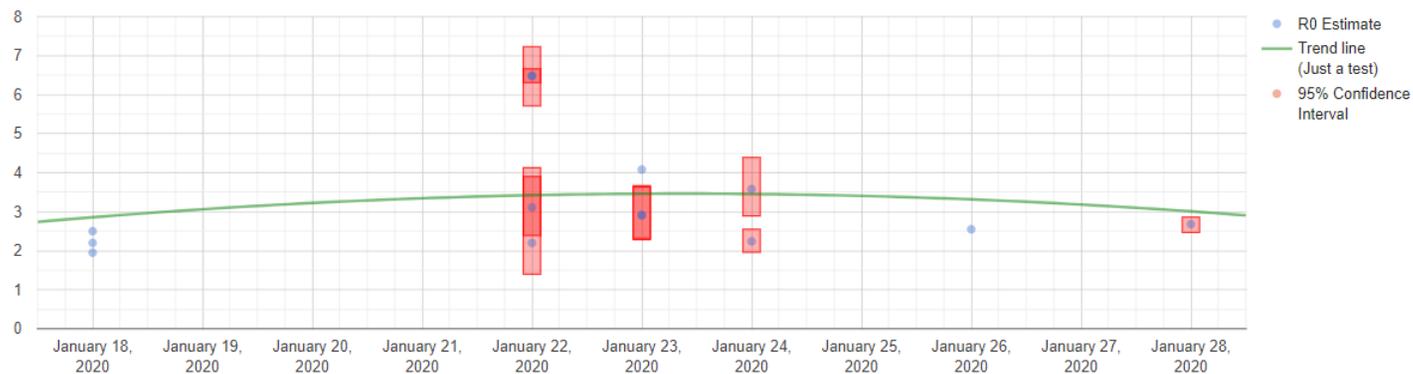




Bar Chart

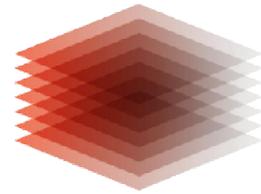


Combo Chart



---

LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



**TIB**

## **Social Science Examples**

# The Team



## Group Leaders



Dr. Markus Stocker



Dr. Gábor Kismihók

## PostDocs



Dr. Javad Chamanara



Dr. Jennifer D'Souza

## Software Development



Manuel Prinz



Kheir Eddine Farfar

## Doctoral Researchers



Olga Lezhnina



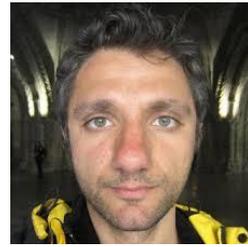
Allard Oelen



Yaser Jaradeh



Shereif Eid



Sarven Capadisli



Vitalis Wiens



## Collaborators TIB/L3S Scientific Data Management



Prof. (Univ. S. Bolivar)  
Dr. Maria Esther Vidal



Kemele Endris



Farah Karim

## Collaborators InfAI Leipzig / AKSW



Dr. Michael Martin



Natanael Arndt

## Project Management



Alex Garatzogianni

## Conclusions

- We need to **reinvent scholarly communication**
- **Knowledge Graphs** are perfectly suited to capture research contributions in a structured and semantic way making them human and machine interpretable
- With our **Open Research Knowledge Graph initiative** we aim to establish a registry for research contributions
- Curation and **synergistic combination of human, expert and machine intelligence** is a challenge

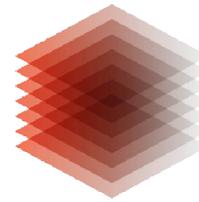
### Stay tuned

- <https://tib.eu>
- Consider creating an ORKG observatory for your domain
- Mailinglist/group:  
<https://groups.google.com/forum/#!forum/orkg>
- Open Research Knowledge Graph:  
<https://orkg.org>
- ERC Consolidator Grant ScienceGRAPH on the topic



European  
Research  
Council

LEIBNIZ INFORMATION CENTRE  
FOR SCIENCE AND TECHNOLOGY  
UNIVERSITY LIBRARY



TIB



<https://de.linkedin.com/in/soerenauer>



<https://twitter.com/soerenauer>



[https://www.xing.com/profile/Soeren\\_Auer](https://www.xing.com/profile/Soeren_Auer)



[http://www.researchgate.net/profile/Soeren\\_Auer](http://www.researchgate.net/profile/Soeren_Auer)

### Contact

Prof. Dr. Sören Auer

TIB & Leibniz University of Hannover

[auer@tib.eu](mailto:auer@tib.eu)



Leibniz  
Universität  
Hannover

Leibniz  
Association