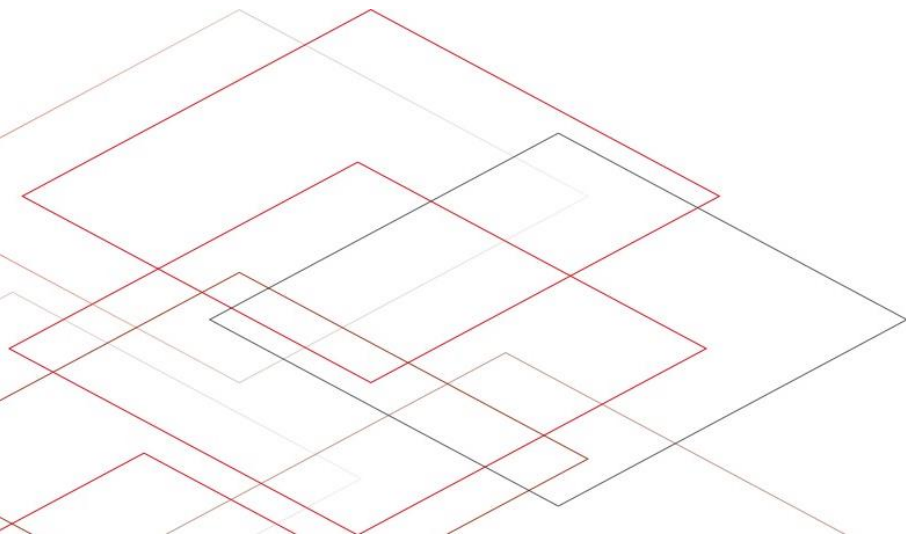


# Rethinking Scholarly Communication with the Open Research Knowledge Graph

Vinodh Ilangovan  
29.01.2024

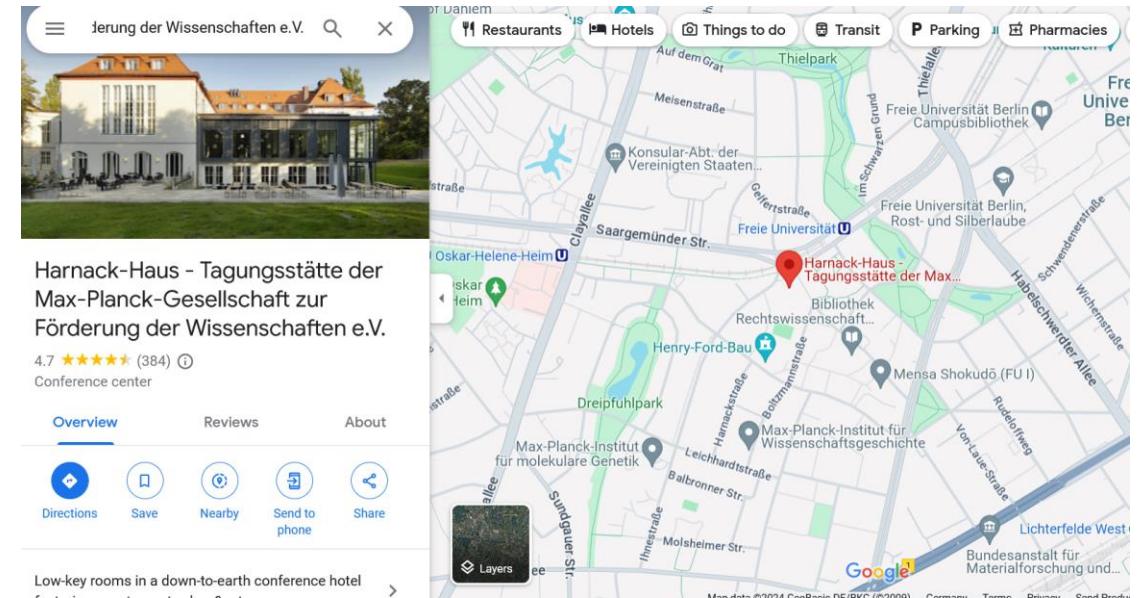
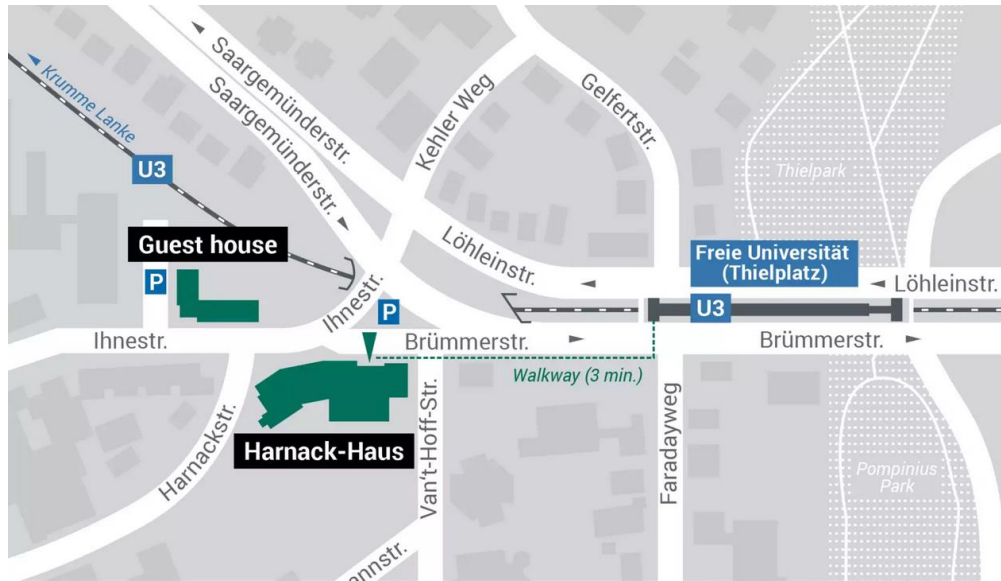
Max Planck Open Science Days 2024



# Outline

- Digitalization in 21st century
- Problems of document centric information flows
- Open Research Knowledge Graph (ORKG)
- Features of the ORKG
- Data curation in the ORKG

# Digitalization in everyday life



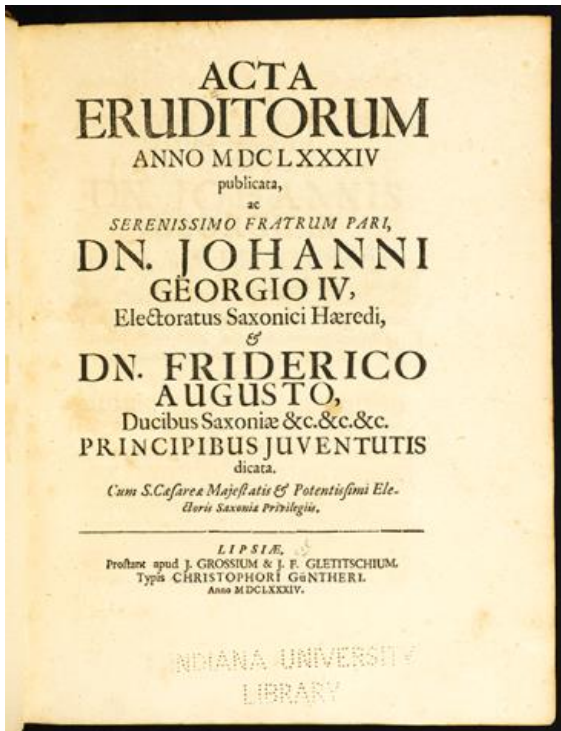
Credits: Harnack-Haus, Max Planck Society

+ New Features:

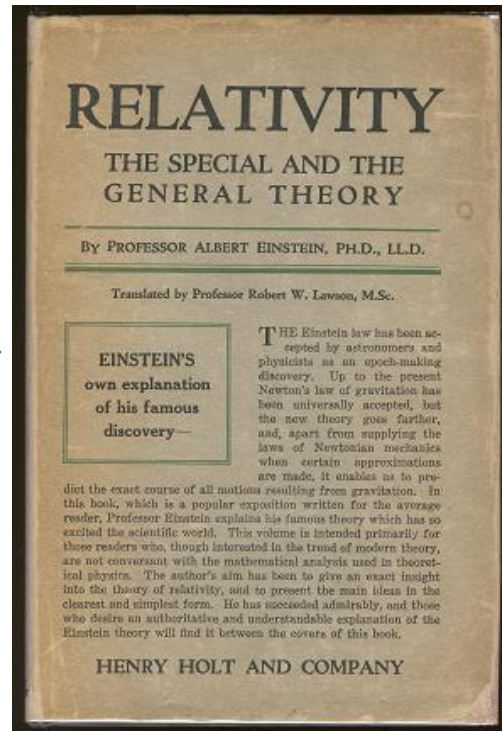
- Zoom in
- Traffic jam warning
- Opening hours
- Interesting places around

# What about Science/ Scholarly communication?

Over 300 years ago



100 years ago



20 years ago

Information Retrieval P. BAXENDALE, Editor

### A Relational Model of Data for Large Shared Data Banks

E. F. COOD  
IBM Research Laboratory, San Jose, California

Future users of large data banks must be protected from having to know how the data is organized in the machine (the internal representation). A prompting service which supplies such information is not a satisfactory solution. Activities of users of terminals and most application programs should remain unaffected when the internal representation of data is changed and even when some aspects of the external representation are changed. Changes in data representation will often be needed as a result of changes in query, update, and report traffic and natural growth in the types of stored information.

Existing noninferential, formatted data systems provide users with tree-structured files or slightly more general network models of the data. In Section 1, inadequacies of these models are discussed. A model based on n-ary relations, a normal form for data base relations, and the concept of a universal data sublanguage are introduced. In Section 2, certain operations on relations (other than logical inference) are discussed and applied to the problems of redundancy and consistency in the user's model.

KEY WORDS AND PHRASES: data bank, data base, data structure, data organization, hierarchies of data, networks of data, relations, derivability, redundancy, consistency, composition, join, retrieval language, predicate calculus, security, data integrity

CC CATEGORIES: 3.70, 3.73, 3.75, 4.20, 4.22, 4.29

1. Relational Model and Normal Form

1.1. INTRODUCTION

This paper is concerned with the application of elementary relation theory to systems which provide shared access to large banks of formatted data. Except for a paper by Childs [1], the principal application of relations to data systems has been to deductive question-answering systems. Levin and Maron [2] provide numerous references to work in this area.

In contrast, the problems treated here are those of data independence—the independence of application programs and terminal activities from growth in data types and changes in data representation—and certain kinds of data inconsistency which are expected to become troublesome even in nondeductive systems.

The relational view (or model) of data described in Section 1 appears to be superior in several respects to the graph or network model [3, 4] presently in vogue for non-inferential systems. It provides a means of describing data with its natural structure only—that is, without superimposing any additional structure for machine representation purposes. Accordingly, it provides a basis for a high level data language which will yield maximal independence between programs on the one hand and machine representation and organization of data on the other.

A further advantage of the relational view is that it forms a sound basis for treating derivability, redundancy, and consistency of relations—these are discussed in Section 2. The network model, on the other hand, has spawned a number of confusions, not the least of which is mistaking the derivation of connections for the derivation of relations (see remarks in Section 2 on the "connection trap").

Finally, the relational view permits a clearer evaluation of the scope and logical limitations of present formatted data systems, and also the relative merits (from a logical standpoint) of competing representations of data within a single system. Examples of this clearer perspective are cited in various parts of this paper. Implementations of systems to support the relational model are not discussed.

1.2. DATA DEPENDENCIES IN PRESENT SYSTEMS

The provision of data description tables in recently developed information systems represents a major advance toward the goal of data independence [5, 6, 7]. Such tables facilitate changing certain characteristics of the data representation stored in a data bank. However, the variety of data representation characteristics which can be changed without logically impairing some application programs is still quite limited. Further, the model of data with which users interact is still cluttered with representational properties, particularly in regard to the representation of collections of data (as opposed to individual items). Three of the principal kinds of data dependencies which still need to be removed are: ordering dependence, indexing dependence, and access path dependence. In some systems these dependencies are not clearly separable from one another.

1.2.1. Ordering Dependence. Elements of data in a data bank may be stored in a variety of ways, some involving no concern for ordering, some permitting each element to participate in one ordering only, others permitting each element to participate in several orderings. Let us consider those existing systems which either require or permit data elements to be stored in at least one total ordering which is closely associated with the hardware-determined ordering of addresses. For example, the records of a file concerning parts might be stored in ascending order by part serial number. Such systems normally permit application programs to assume that the order of presentation of records from such a file is identical to (or is a subordering of) the

Volume 13 / Number 6 / June, 1970 Communications of the ACM 377

Today

BIBLIOTHEK - Forschung und Praxis 2020; 44(3): 516-529 DE GRUYTER

### Textmining

Sören Auer\*, Allard Oelen, Muhammad Haris, Markus Stocker, Jennifer D'Souza, Khair Eddine Farfar, Lars Vogt, Manuel Prinz, Vitalis Wiens and Mohamed Yaser Jaradeh

### Improving Access to Scientific Literature with Knowledge Graphs

<https://doi.org/10.5151/fbp-2020-0942>

**Keywords:** Subject classification; knowledge graph; semantic web; crowdsourcing; text mining

**Abstract:** The transfer of knowledge has not changed fundamentally for many hundreds of years: it is usually documented—formerly printed on paper as a classic essay and nowadays as PDF. With around 2.5 million new research contributions every year, researchers drown in a flood of pseudo-digitized PDF publications. As a result research is seriously weakened. In this article, we argue for representing scholarly contributions in a structured and semantic way as a knowledge graph. The advantage is that information represented in a knowledge graph is readable by machines and humans. As an example, we give an overview on the Open Research Knowledge Graph (ORKG), a service implementing this approach. For creating the knowledge graph representation, we rely on a mixture of manual (crowd/expert sourcing) and (semi-)automated techniques. Only with such a combination of human and machine intelligence, we can achieve the required quality of the representation to allow for novel exploration and assistance services for researchers. As a result, a scholarly knowledge graph such as the ORKG can be used to give a condensed overview on the state-of-the-art addressing a particular research quest, for example as a tabular comparison of contributions according to various characteristics of the approaches. Further possible intuitive access interfaces to such scholarly knowledge graphs include domain-specific (chart) visualizations or answering of natural language questions.

**Verbesserten Zugang zu wissenschaftlicher Literatur mit Wissensgraphen**

**Zusammenfassung:** Der Verbleib wissenschaftlicher Erkenntnisse hat sich seit vielen hundert Jahren nicht grundlegend verändert: Er erfolgt in der Regel dokumentenbasiert – früher als klassischer Aufsatz auf Papier gedruckt und heute online als PDF. Mit rund 2,5 Millionen neuen Forschungsbeiträgen pro Jahr ertrinken Forscher in einer Flut von pseudo-digitalisierten PDF-Publikationen. Als Folge davon wird die Forschung stark geschwächt. In diesem Artikel plädieren wir dafür, wissenschaftliche Beiträge in strukturierter und semantischer Form als Wissensgraph zu repräsentieren. Der Vorteil ist, dass die in einem Wissensgraph dargestellten Informationen für Maschinen und Menschen lesbar sind. Als Beispiel geben wir einen Überblick über den Open Research Knowledge Graph (ORKG), einen Dienst, der diesen Ansatz umsetzt. Für die Erstellung des Wissensgraph setzen wir eine Mischung aus manuellen (crowd/expert sourcing) und (halb-)automatisierten Techniken ein. Nur mit einer solchen Kombination aus menschlicher und maschineller Intelligenz können wir die erforderliche Qualität der Darstellung erreichen, um neue Explorations- und Unterstützungsdienste für Forscher zu ermöglichen. In Ergebnis kann ein Wissensgraph wie der ORKG verwendet werden, um einen kompakten Überblick über den Stand der Technik in Bezug auf eine bestimmte Forschungsaufgabe zu geben, z. B. als tabellarischer Vergleich der Beiträge nach verschiedenen Merkmalen der Ansätze. Weitere mögliche intuitive Nutzungsschnittstellen zu solchen wissenschaftlichen Wissensgraphen sind domänenspezifische Visualisierungen oder die Beantwortung natürlichsprachlicher Fragen mittels Text Mining.

**Schlüsselwörter:** Sacherschließung; Wissensgraph; Semantische Web; Crowdsourcing; Text Mining

\*Corresponding author: Prof. Dr. Sören Auer, [soeren@fbp.de](mailto:soeren@fbp.de)  
Allard Oelen, [allard.oelen@fbp.de](mailto:allard.oelen@fbp.de)  
Muhammad Haris, [muhhammad.haris@fbp.de](mailto:muhhammad.haris@fbp.de)  
Dr. Markus Stocker, [markus.stocker@fbp.de](mailto:markus.stocker@fbp.de)  
Dr. Jennifer D'Souza, [jennifer.dsouza@fbp.de](mailto:jennifer.dsouza@fbp.de)  
Khair Eddine Farfar, [khair.farfar@fbp.de](mailto:khair.farfar@fbp.de)  
Lars Vogt, [lars.vogt@fbp.de](mailto:lars.vogt@fbp.de)  
Manuel Prinz, [manuel.prinz@fbp.de](mailto:manuel.prinz@fbp.de)  
Vitalis Wiens, [vitalis.wiens@fbp.de](mailto:vitalis.wiens@fbp.de)  
Mohamed Yaser Jaradeh, [yaser.jaradeh@fbp.de](mailto:yaser.jaradeh@fbp.de)

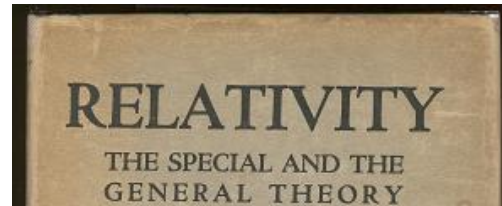


# What about Science/ Scholarly communication?

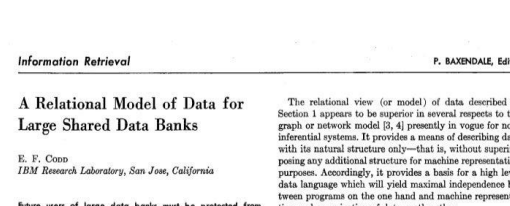
Over 300 years ago



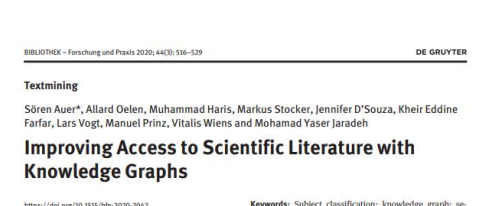
100 years ago



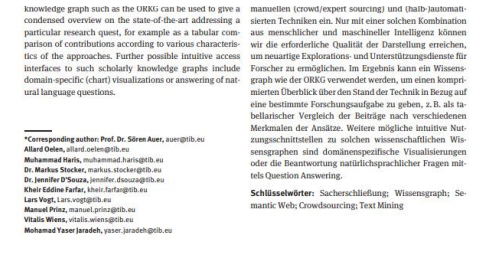
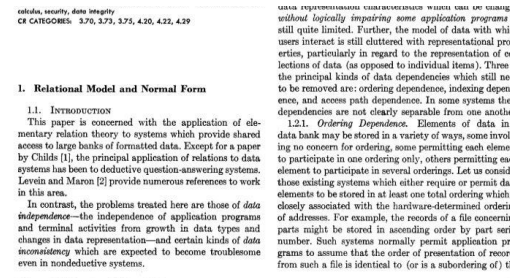
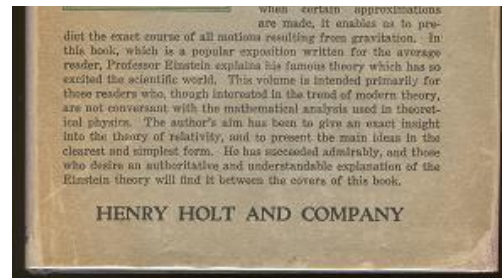
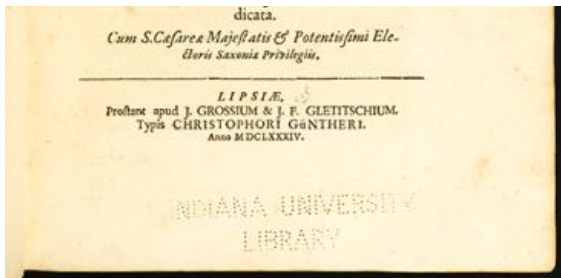
20 years ago



Today



## Science does not harvest the full potential of digitalization



Not much has changed!

# A Consequence of Document Centered Information Flows: The Publication Flood

- ~ 2.5 Mio new publications per year
- Globally ~ \$1.7 trillion spent on research
- Researchers lack overview, even in small fields
- Loss of knowledge
- Answering questions is like looking for a needle in the haystack

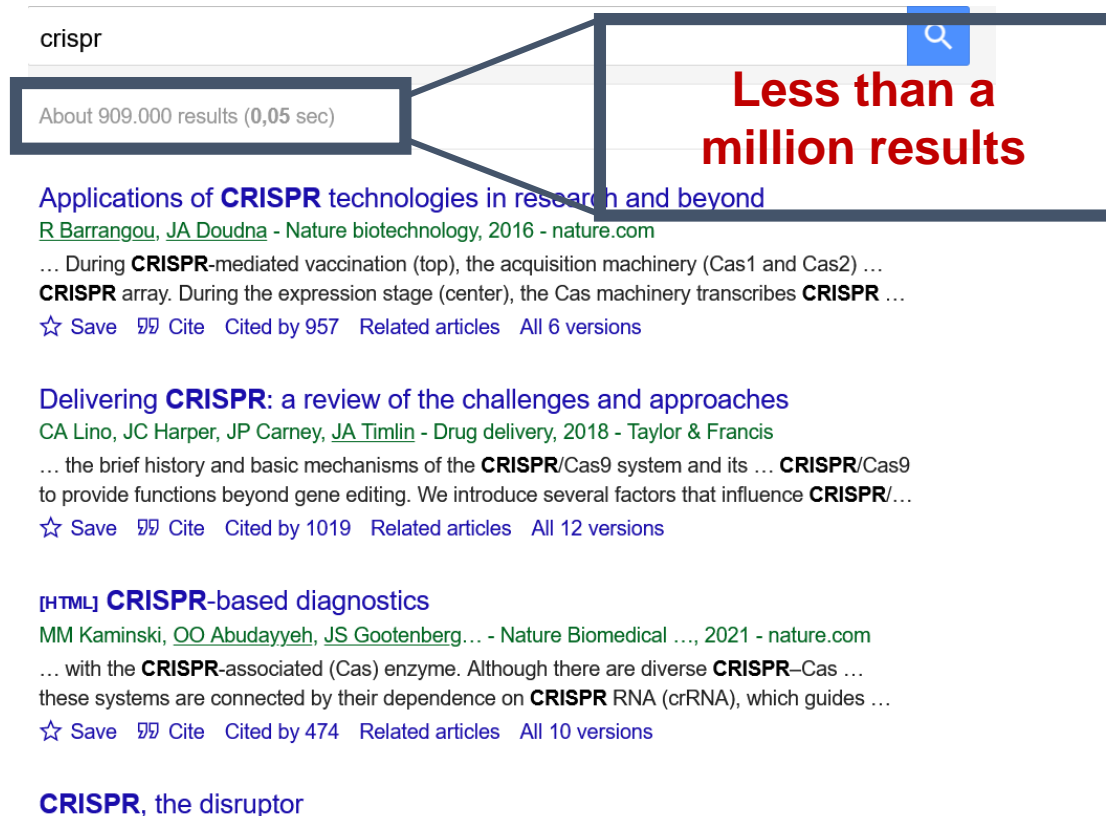


# Further challenges of document-centered information flows

- Reproducibility Crisis
- Monopolization of commercial actors
- Deficiency of Peer-Review
- Predatory Publishing



# An example – CRISPR



The image shows a search engine results page for the query 'crispr'. A search bar at the top contains the text 'crispr' and a magnifying glass icon. Below the search bar, a box indicates 'About 909.000 results (0,05 sec)'. A large callout box with a blue border and a magnifying glass icon contains the text 'Less than a million results' in red. Below the search results, three article snippets are visible, each with a title, author information, a brief description, and interaction options like 'Save', 'Cite', and 'Cited by'.

crispr

About 909.000 results (0,05 sec)

**Less than a million results**

[Applications of \*\*CRISPR\*\* technologies in research and beyond](#)  
R Barrangou, JA Doudna - Nature biotechnology, 2016 - nature.com  
... During **CRISPR**-mediated vaccination (top), the acquisition machinery (Cas1 and Cas2) ...  
**CRISPR** array. During the expression stage (center), the Cas machinery transcribes **CRISPR** ...  
☆ Save 📄 Cite Cited by 957 Related articles All 6 versions

[Delivering \*\*CRISPR\*\*: a review of the challenges and approaches](#)  
CA Lino, JC Harper, JP Carney, JA Timlin - Drug delivery, 2018 - Taylor & Francis  
... the brief history and basic mechanisms of the **CRISPR**/Cas9 system and its ... **CRISPR**/Cas9  
to provide functions beyond gene editing. We introduce several factors that influence **CRISPR**/...  
☆ Save 📄 Cite Cited by 1019 Related articles All 12 versions

[\[HTML\] \*\*CRISPR\*\*-based diagnostics](#)  
MM Kaminski, OO Abudayyeh, JS Gootenberg... - Nature Biomedical ..., 2021 - nature.com  
... with the **CRISPR**-associated (Cas) enzyme. Although there are diverse **CRISPR**-Cas ...  
these systems are connected by their dependence on **CRISPR** RNA (crRNA), which guides ...  
☆ Save 📄 Cite Cited by 474 Related articles All 10 versions

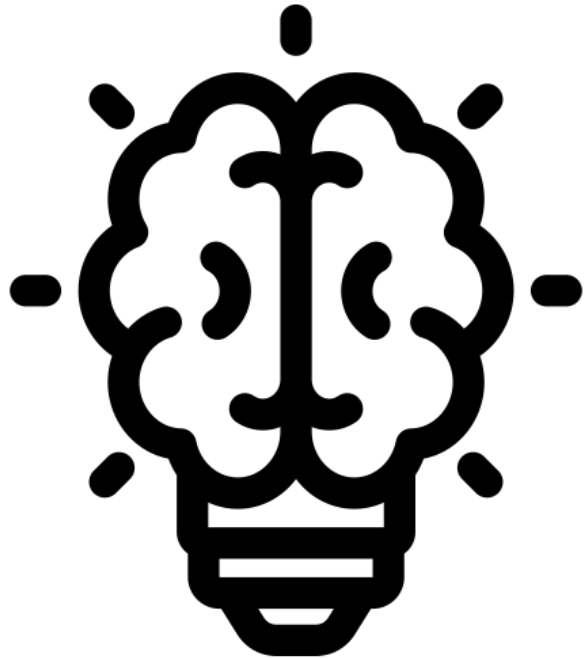
[CRISPR, the disruptor](#)

## Specific research questions:

- Who applied CRISPR to butterflies?
- How to apply CRISPR with minimal costs?
- How do different genome editing techniques compare?



# Time to Rethink Scholarly Communication!



*“The lightbulb was **not** invented by improving the candle.”*

**Oren Harari**

Digitalization is **more** than just Digitization!  
Current and future scientific challenges can not be tackled with an outdated communication system.

---

**Digitalize Knowledge,  
Not Documents!**

# Knowledge Graphs are widely used in industry...



Why not use them for (open) science as well?

# Knowledge Representation in Graphs

From papers...



bioRxiv posts many COVID19-related papers. A reminder: they have not been formally peer-reviewed and should not guide health-related behavior or be reported in the press as conclusive.

New Results [Follow this preprint](#)

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

Linlin Zhang, Robert D. 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](https://doi.org/10.1007/978-981-10-4956-9_8)

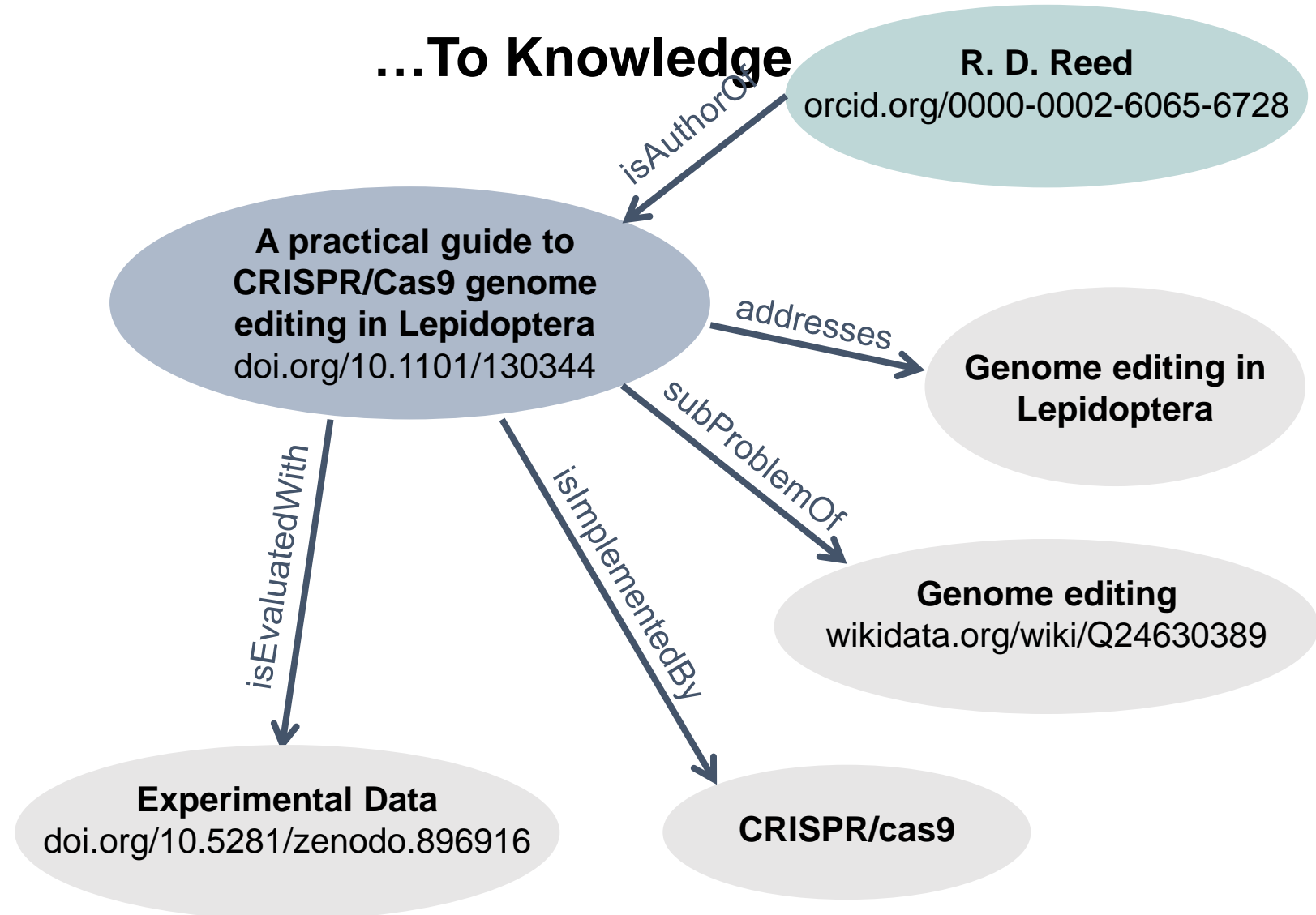


**Abstract** Full Text Info/History Metrics [Preview PDF](#)

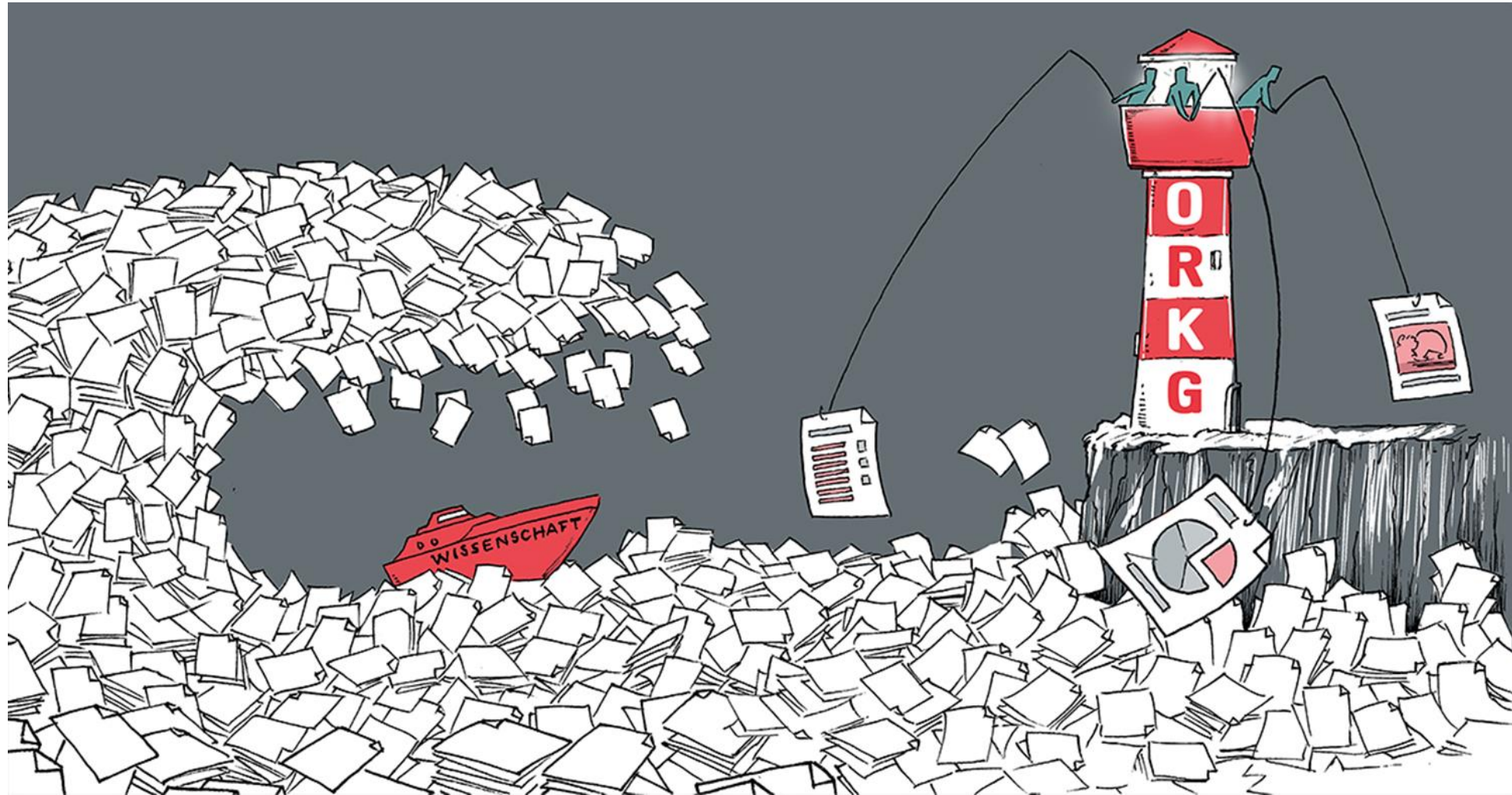
## Abstract

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 knockins that we have achieved in butterflies. Finally, we provide a complete, detailed protocol for producing targeted long deletions in butterflies.

...To Knowledge

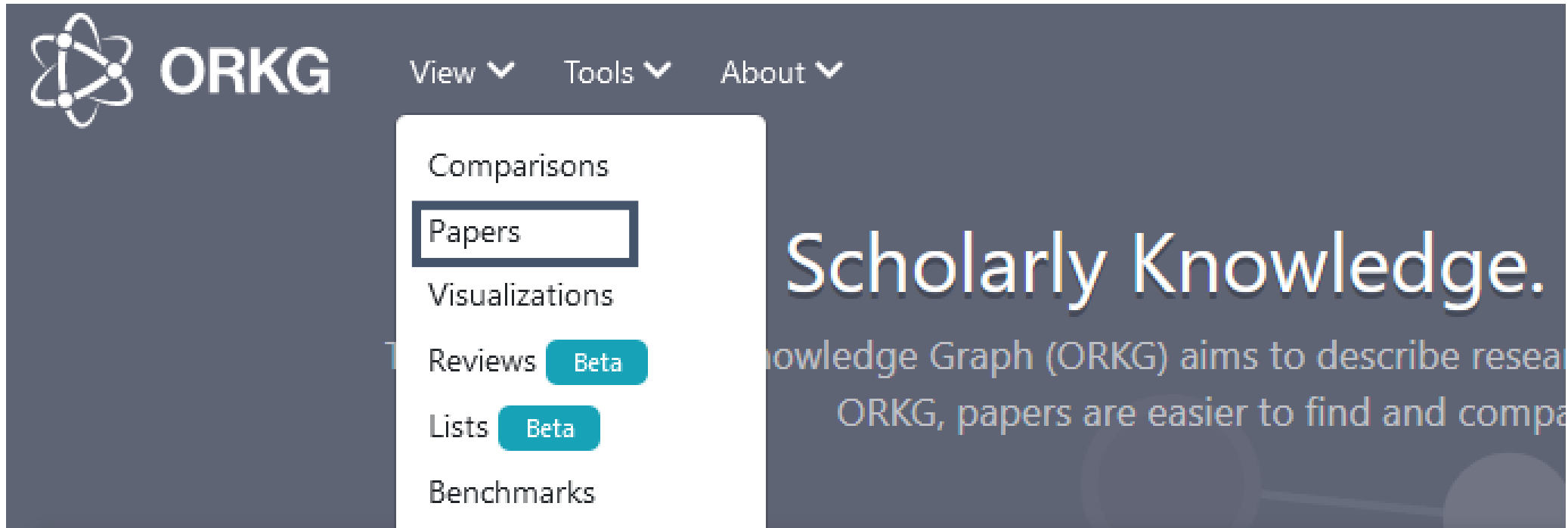


# ORKG: Lighthouse in the Publication Flood





# Features of the ORKG



ORKG papers describe scholarly articles in a structured and semantic manner.

# Structured description of scholarly outputs



## Real-Time PCR Threshold Cycle Cutoffs Help To Identify Agents Causing Acute Childhood Diarrhea in Zanzibar

Kristina Elfving,<sup>a,b</sup> Maria Andersson,<sup>a</sup> Mwinyi I. Msellem,<sup>c</sup> Christina Welinder-Olsson,<sup>a</sup> Max Petzold,<sup>d</sup> Anders Björkman,<sup>a</sup> Birger Trollfors,<sup>a</sup> Andreas Mårtensson,<sup>a,f</sup> Magnus Lindh<sup>a</sup>

<sup>a</sup>Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden; <sup>b</sup>Department of Pediatrics, University of Gothenburg, Gothenburg, Sweden; <sup>c</sup>Zanzibar Malaria Elimination Programme, Ministry of Health, Zanzibar, Tanzania; <sup>d</sup>Academistatistik Centre for Applied Biostatistics, Occupational and Environmental Medicine, University of Gothenburg, Gothenburg, Sweden; <sup>e</sup>Malaria Research, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden; <sup>f</sup>Global Health, Department of Public Health Sciences, Karolinska Institute, Stockholm, Sweden

Molecular assays might improve the identification of causes of acute diarrheal disease but might lead to more frequent detection of asymptomatic infections. In the present study, real-time PCR targeting 14 pathogens was applied to rectal swabs from 330 children aged 2 to 59 months in Zanzibar, including 165 patients with acute diarrhea and 165 asymptomatic control subjects. At least one pathogen was detected for 94% of the patients and 84% of the controls, with higher rates among patients for norovirus genogroup II (20% versus 2.4%;  $P < 0.0001$ ), rotavirus (10% versus 1.8%;  $P = 0.003$ ), and *Cryptosporidium* (30% versus 11%;  $P < 0.0001$ ). Detection rates did not differ significantly for enterotoxigenic *Escherichia coli* (ETEC)-*estA* (33% versus 24%), ETEC-*eltB* (44% versus 46%), *Shigella* (35% versus 33%), and *Campylobacter* (35% versus 33%), but for these agents threshold cycle ( $C_T$ ) values were lower (pathogen loads were higher) in sick children than in controls. In a multivariate analysis,  $C_T$  values for norovirus genogroup II, rotavirus, *Cryptosporidium*, ETEC-*estA*, and *Shigella* were independently associated with diarrhea. We conclude that this real-time PCR allows convenient detection of essentially all diarrheagenic agents and provides  $C_T$  values that may be critical for the interpretation of results for pathogens with similar detection rates in patients and controls. The results indicate that the assessment of pathogen loads may improve the identification of agents causing gastroenteritis in children.

Acute diarrheal disease is the second most common cause of death worldwide in children younger than 5 years (1). Most of these deaths occur in low-income countries, where the etiologies of diarrheal infections have been incompletely understood because there are few comprehensive studies (2, 3). Such studies often used traditional diagnostic methods, such as culture, microscopy, or antigen detection, or focused on only one or a few diarrheal pathogens.

New multitargeting molecular PCR methods allow detection of diarrheal pathogens with high specificity and sensitivity (4–7), and their application may lead to improved understanding of diarrheal disease epidemiology. These methods provide better identification of viruses that cannot be cultured (e.g., *Caliciviridae*) or that previously have been diagnosed with methods with relatively low sensitivity (e.g., antigen testing for rotavirus) (8–10). They also have been shown to improve the detection of bacteria because of their higher sensitivity than culture (11–15). However, the mere presence of a pathogen in a fecal sample does not necessarily imply that it is the cause of disease, since high detection rates have been reported also for asymptomatic individuals with both conventional (2, 16) and molecular (12) methods. This is of particular importance in low-income countries, where children may be exposed to multiple enteric pathogens due to poor sanitary conditions. Thus, understanding the causes of diarrheal disease and how test results should be interpreted requires knowledge of the presence of pathogens in feces from both ill and healthy individuals.

In the present study, we used a broad real-time PCR assay to analyze pathogens in children, with or without diarrhea, in Zanzibar. In addition to comparing detection rates, we aimed at evaluating the potential utility of pathogen loads, in terms of real-time PCR threshold cycle ( $C_T$ ) values, to separate symptomatic from

asymptomatic infections, as suggested by studies on norovirus, rotavirus, and *Shigella* infections (17–19).

### MATERIALS AND METHODS

**Study participants.** (i) **Patients.** Children 2 to 59 months of age who presented to the Kivunge Primary Health Care Centre (PHCC) in rural Zanzibar (North A district) with fever (measured axillary temperature of  $\geq 37.5^\circ\text{C}$  or a history of fever during the preceding 24 h, according to the accompanying guardian) and diarrhea (history of loose stools during the preceding 24 h) were eligible for study inclusion. Children with signs of severe disease according to Integrated Management of Childhood Illness (IMCI) guidelines ([http://www.who.int/child\\_adolescent\\_health/documents/IMCI\\_chartbook/en/index.html](http://www.who.int/child_adolescent_health/documents/IMCI_chartbook/en/index.html)) were excluded. Recruitment was performed in April to July 2011, corresponding to the end of the rainy season and the beginning of the dry season.

(ii) **Asymptomatic control subjects.** Control subjects matched for living area and sampling time period, i.e., asymptomatic children 2 to 59 months of age, were recruited once a week during the entire study period, together with local representatives from 8 villages in the study area. No more than 2 children per household were recruited. An asymptomatic child was defined as having no history of diarrhea, cough, running nose, or fever in the preceding 10 days.

The study was approved in Zanzibar by the Zanzibar Medical Research

Received 27 September 2013; Returned for modification 3 November 2013; Accepted 1 January 2014

Published ahead of print 8 January 2014

Editor: P. H. Gilligan

Address correspondence to Magnus Lindh, [magnus.lindh@microbio.gu.se](mailto:magnus.lindh@microbio.gu.se), or Kristina Elfving, [kristina.elfving@region.se](mailto:kristina.elfving@region.se).

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JCM.02697-13

asked from <https://journals.asm.org/journal> on 14 March 2023 by 92.219.130.138.



## Real-Time PCR Threshold Cycle Cutoffs Help To Identify Agents Causing Acute Childhood Diarrhea in Zanzibar

March 2014 45 citations Virology Kristina Elfving Maria Andersson Mwinyi I. Msellem Christina Welinder-Olsson Max Petzold Anders Björkman Birger Trollfors Andreas Mårtensson Magnus Lindh

Published in: *Journal of Clinical Microbiology*

DOI: <https://doi.org/10.1128/jcm.02697-13>

Contribution 1	
approach	Calculate the cut-off Ct value that is associated with symptomatic infection
beginning	Value not reported in paper
end	Value not reported in paper
hypothesis supported	✗
location	Zanzibar
Material	fecal samples
method	Real-time reverse-transcription PCR
Norovirus	GII
patient age	2 - 59 months
research problem	What is the correlation between norovirus viral load and disease severity

Add to comparison

Provenance Timeline

Added on  
27 Oct 2022

Added by  
 Nele Villabruna

Contributors  
Nele Villabruna

<https://orkg.org/paper/R228732>

# Behind the structured description of scholarly contributions



## Real-Time PCR Threshold Cycle Cutoffs Help To Identify Agents Causing Acute Childhood Diarrhea in Zanzibar

Kristina Elfving,<sup>1,2</sup> Maria Andersson,<sup>3</sup> Mwynyi I. Msellem,<sup>4</sup> Christina Welinder-Olsson,<sup>5</sup> Max Petzold,<sup>6</sup> Anders Björkman,<sup>7</sup> Birger Trollfors,<sup>8</sup> Andreas Mårtensson,<sup>9,10</sup> Magnus Lindh<sup>11</sup>

Department of Infectious Diseases, University of Gothenburg, Gothenburg, Sweden<sup>1</sup>; Department of Pediatrics, University of Gothenburg, Gothenburg, Sweden<sup>2</sup>; Zanzibar Malaria Elimination Programme, Ministry of Health, Zanzibar, Tanzania<sup>3</sup>; Akademisktjänst, Centre for Applied Biostatistics, Occupational and Environmental Medicine, University of Gothenburg, Gothenburg, Sweden<sup>4</sup>; Malaria Research, Department of Medicine Solna, Karolinska Institute, Stockholm, Sweden<sup>5</sup>; Global Health, Department of Public Health Sciences, Karolinska Institute, Stockholm, Sweden<sup>6</sup>

Molecular assays might improve the identification of causes of acute diarrheal disease but might lead to more frequent detection of asymptomatic infections. In the present study, real-time PCR targeting 14 pathogens was applied to rectal swabs from 330 children aged 2 to 59 months in Zanzibar, including 165 patients with acute diarrhea and 165 asymptomatic control subjects. At least one pathogen was detected for 94% of the patients and 84% of the controls, with higher rates among patients for norovirus genogroup II (20% versus 2.4%;  $P < 0.0001$ ), rotavirus (10% versus 1.8%;  $P = 0.003$ ), and *Cryptosporidium* (30% versus 11%;  $P < 0.0001$ ). Detection rates did not differ significantly for enterotoxigenic *Escherichia coli* (ETEC)-*estA* (33% versus 24%), ETEC-*eitB* (44% versus 46%), *Shigella* (35% versus 33%), and *Campylobacter* (35% versus 33%), but for these agents threshold cycle ( $C_t$ ) values were lower (pathogen loads were higher) in sick children than in controls. In a multivariate analysis,  $C_t$  values for norovirus genogroup II, rotavirus, *Cryptosporidium*, ETEC-*estA*, and *Shigella* were independently associated with diarrhea. We conclude that this real-time PCR allows convenient detection of essentially all diarrheagenic agents and provides  $C_t$  values that may be critical for the interpretation of results for pathogens with similar detection rates in patients and controls. The results indicate that the assessment of pathogen loads may improve the identification of agents causing gastroenteritis in children.

Acute diarrheal disease is the second most common cause of death worldwide in children younger than 5 years (1). Most of these deaths occur in low-income countries, where the etiologies of diarrheal infections have been incompletely understood because there are few comprehensive studies (2, 3). Such studies often used traditional diagnostic methods, such as culture, microscopy, or antigen detection, or focused on only one or a few diarrheal pathogens.

New multitargeting molecular PCR methods allow detection of diarrheal pathogens with high specificity and sensitivity (4–7), and their application may lead to improved understanding of diarrheal disease epidemiology. These methods provide better identification of viruses that cannot be cultured (e.g., *Caliciviridae*) or that previously have been diagnosed with methods with relatively low sensitivity (e.g., antigen testing for rotavirus) (8–10). They also have been shown to improve the detection of bacteria because of their higher sensitivity than culture (11–15). However, the mere presence of a pathogen in a fecal sample does not necessarily imply that it is the cause of disease, since high detection rates have been reported also for asymptomatic individuals with both conventional (2, 16) and molecular (12) methods. This is of particular importance in low-income countries, where children may be exposed to multiple enteric pathogens due to poor sanitary conditions. Thus, understanding the causes of diarrheal disease and how test results should be interpreted requires knowledge of the presence of pathogens in feces from both ill and healthy individuals.

In the present study, we used a broad real-time PCR assay to analyze pathogens in children, with or without diarrhea, in Zanzibar. In addition to comparing detection rates, we aimed at evaluating the potential utility of pathogen loads, in terms of real-time PCR threshold cycle ( $C_t$ ) values, to separate symptomatic from

asymptomatic infections, as suggested by studies on norovirus, rotavirus, and *Shigella* infections (17–19).

### MATERIALS AND METHODS

**Study participants.** (i) **Patients.** Children 2 to 59 months of age who presented to the Kivunge Primary Health Care Centre (PHCC) in rural Zanzibar (North A district) with fever (measured axillary temperature of  $\geq 37.5^\circ\text{C}$  or a history of fever during the preceding 24 h, according to the accompanying guardian) and diarrhea (history of loose stools during the preceding 24 h) were eligible for study inclusion. Children with signs of severe disease according to Integrated Management of Childhood Illness (IMCI) guidelines ([http://www.who.int/child\\_adolescent\\_health/documents/IMCI\\_charbooklet/en/index.html](http://www.who.int/child_adolescent_health/documents/IMCI_charbooklet/en/index.html)) were excluded. Recruitment was performed in April to July 2011, corresponding to the end of the rainy season and the beginning of the dry season.

(ii) **Asymptomatic control subjects.** Control subjects matched for living area and sampling time period, i.e., asymptomatic children 2 to 59 months of age, were recruited once a week during the entire study period, together with local representatives from 8 villages in the study area. No more than 2 children per household were recruited. An asymptomatic child was defined as having no history of diarrhea, cough, running nose, or fever in the preceding 10 days.

The study was approved in Zanzibar by the Zanzibar Medical Research

Received 27 September 2013; returned for modification 3 November 2013

Accepted 1 January 2014

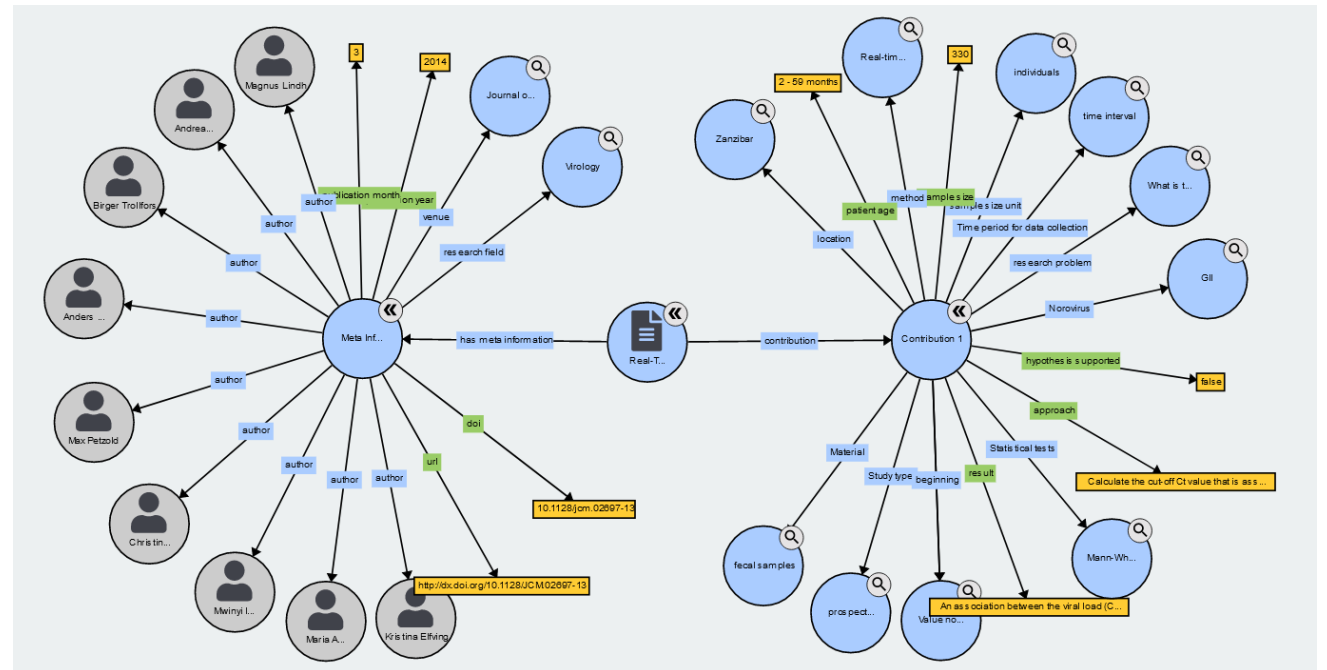
Published ahead of print 8 January 2014

Editor: P. H. Gilligan

Address correspondence to Magnus Lindh, [magnus.lindh@microbiol.gu.se](mailto:magnus.lindh@microbiol.gu.se), or Kristina Elfving, [kristina.elfving@vgregion.se](mailto:kristina.elfving@vgregion.se).

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

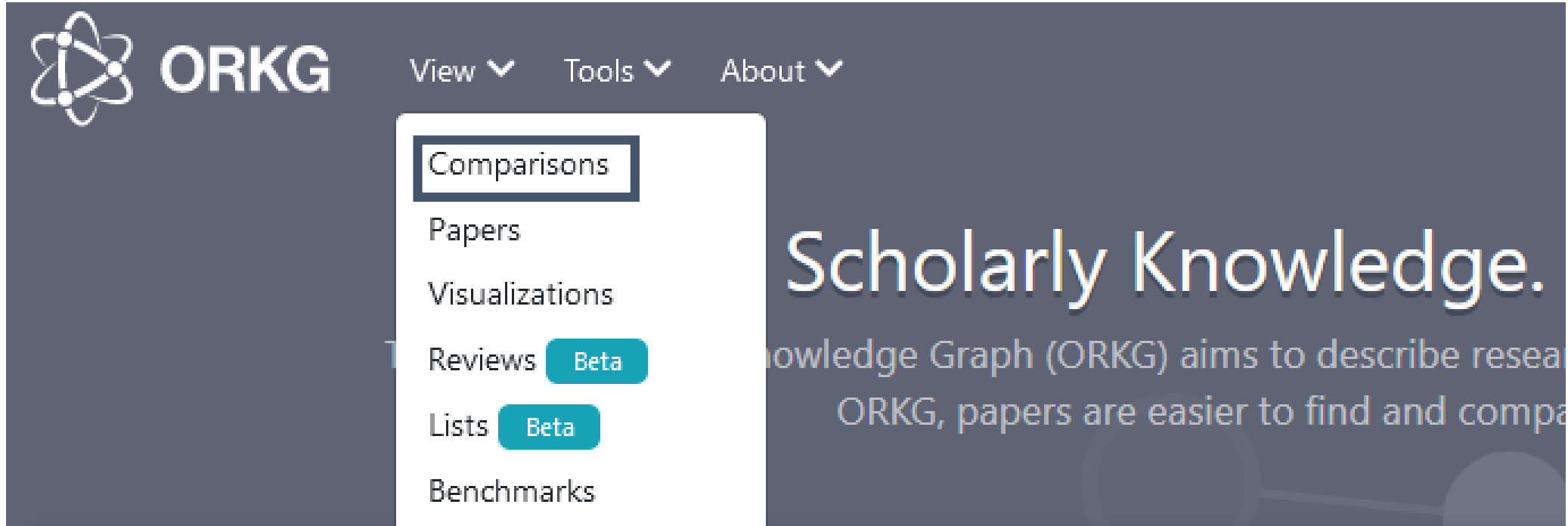
doi:10.1128/JCM.02697-13



Scholarly contributions become machine-actionable and FAIR.

asked from <https://journals.asm.org/journal/jcm> on 14 March 2023 by 92.219.130.138.


# Features of the ORKG





- ORKG comparisons provide condensed tabular overviews of the state-of-the-art for a particular research question.
- Comparison contains: Contributions, Properties, Resources, Literals.




Properties	
<a href="#">location</a>	▼
<a href="#">Time period</a>	▼
<a href="#">has beginning</a>	▼
<a href="#">has end</a>	▼
<a href="#">Basic reproduction number</a>	✕
<a href="#">Has value</a>	▼
<a href="#">Confidence interval (95%)</a>	✕
<a href="#">Lower confidence limit</a>	▼
<a href="#">Upper confidence limit</a>	▼
<a href="#">Method*</a>	▼

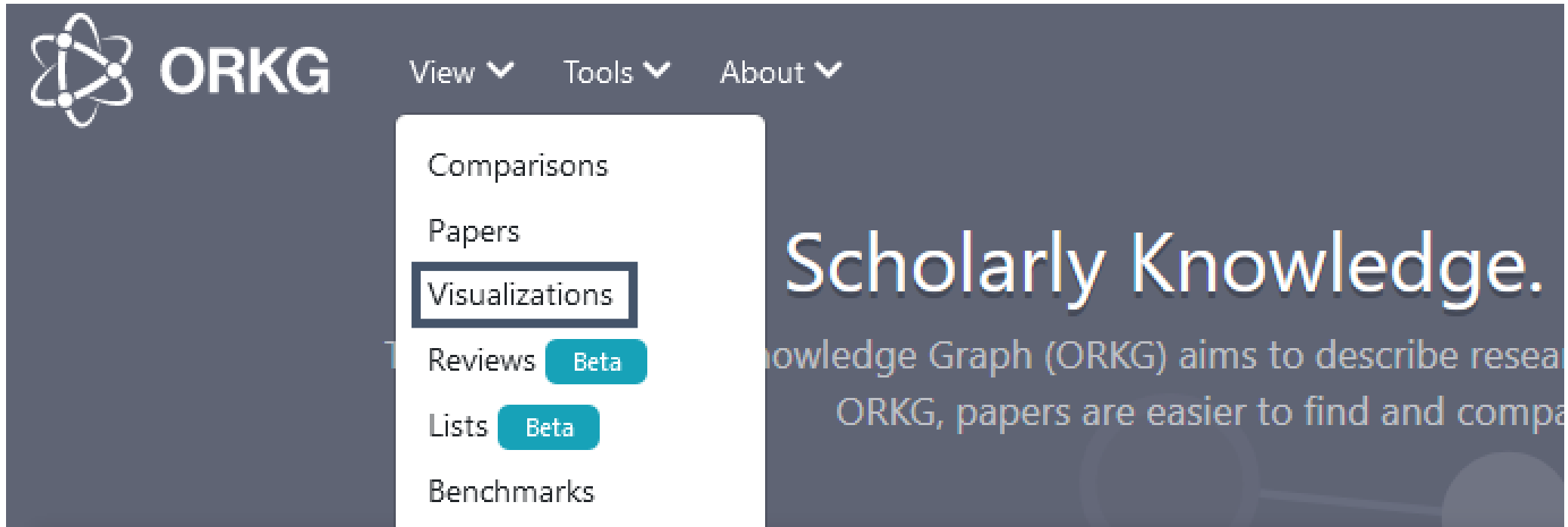
The early phase of the COVID-19 outbreak in Lombardy, Italy <i>2020 - Contribution 1</i>	
	
<a href="#">location</a>	<a href="#">Lombardy, Italy</a>
<a href="#">Time period</a>	<a href="#">Time interval</a>
<a href="#">has beginning</a>	2020-01-14
<a href="#">has end</a>	2020-03-08
<a href="#">Basic reproduction number</a>	<a href="#">Basic reproduction number estimate value specification</a>
<a href="#">Has value</a>	3.1
<a href="#">Confidence interval (95%)</a>	<a href="#">Confidence interval (95%)</a>
<a href="#">Lower confidence limit</a>	2.9
<a href="#">Upper confidence limit</a>	3.2
<a href="#">Method*</a>	

Transmission potential of COVID-19 in Iran <i>2020 - Contribution 1</i>	
	
<a href="#">location</a>	<a href="#">Iran</a>
<a href="#">Time period</a>	<a href="#">Time interval</a>
<a href="#">has beginning</a>	2020-02-19
<a href="#">has end</a>	2020-02-29
<a href="#">Basic reproduction number</a>	<a href="#">Basic reproduction number estimate value specification</a>
<a href="#">Has value</a>	3.6
<a href="#">Confidence interval (95%)</a>	<a href="#">Confidence interval (95%)</a>
<a href="#">Lower confidence limit</a>	3.4
<a href="#">Upper confidence limit</a>	4.2
<a href="#">Method*</a>	generalized growth model

Transmission potential of COVID-19 in Iran <i>2020 - Contribution 2</i>	
	
<a href="#">location</a>	<a href="#">Iran</a>
<a href="#">Time period</a>	<a href="#">Time interval</a>
<a href="#">has beginning</a>	2020-02-19
<a href="#">has end</a>	2020-02-29
<a href="#">Basic reproduction number</a>	<a href="#">Basic reproduction number estimate value specification</a>
<a href="#">Has value</a>	3.58
<a href="#">Confidence interval (95%)</a>	<a href="#">Confidence interval (95%)</a>
<a href="#">Lower confidence limit</a>	1.29
<a href="#">Upper confidence limit</a>	8.46
<a href="#">Method*</a>	based on the calculation of the epidemic's doubling times: estimated epidemic doubling time of 1.20 (95% CI, 1.05, 1.44) days

Estimating the generation interval for COVID-19 based on symptom onset data <i>2020 - Contribution 1</i>	
	
<a href="#">location</a>	<a href="#">Singapore</a>
<a href="#">Time period</a>	<a href="#">Time interval</a>
<a href="#">has beginning</a>	2020-01-21
<a href="#">has end</a>	2020-02-26
<a href="#">Basic reproduction number</a>	<a href="#">Basic reproduction number estimate value specification</a>
<a href="#">Has value</a>	1.27
<a href="#">Confidence interval (95%)</a>	<a href="#">Confidence interval (95%)</a>
<a href="#">Lower confidence limit</a>	1.19
<a href="#">Upper confidence limit</a>	1.36
<a href="#">Method*</a>	generation interval

# Features of the ORKG



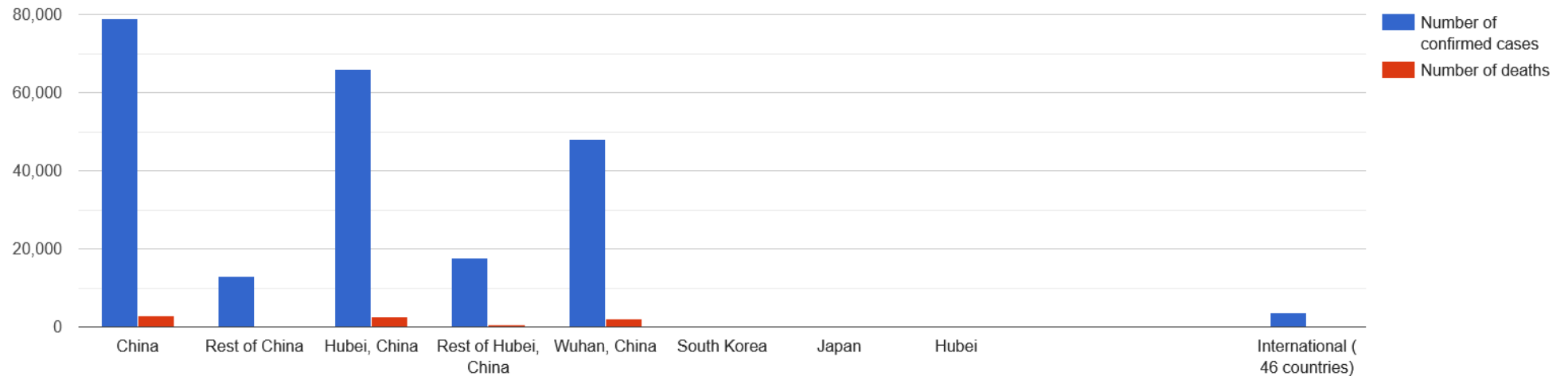
ORKG visualizations are generated from ORKG comparisons.

# Visualization of emerging data and trends in the ORKG

Visualization of the predicted Covid-19 case fatality rate estimates [🔗](#)

Comparison of predicted "Number of confirmed cases" and "Number of deaths" for 2019-nCoV

📅 17 August 2021



<https://orkg.org/comparison/R41466/#VisR139852>

# Publishing State-of-the-Art comparisons

**Acknowledgement of creators** → This overview shows the classification results of approaches that use the machine learning algorithms Naïve Bayes, Support Vector Machines, and Decision Trees C4.5 in combination with the machine learning features Bag of Words or Term Frequency - Inverse Document Frequency to classify user feedback as feature request.

**Citable DOI** → DOI: [10.48366/r112387](https://doi.org/10.48366/r112387)

**Created visualizations** → Visualizations

**Interactive comparison with filtering** → Properties: Has dataset

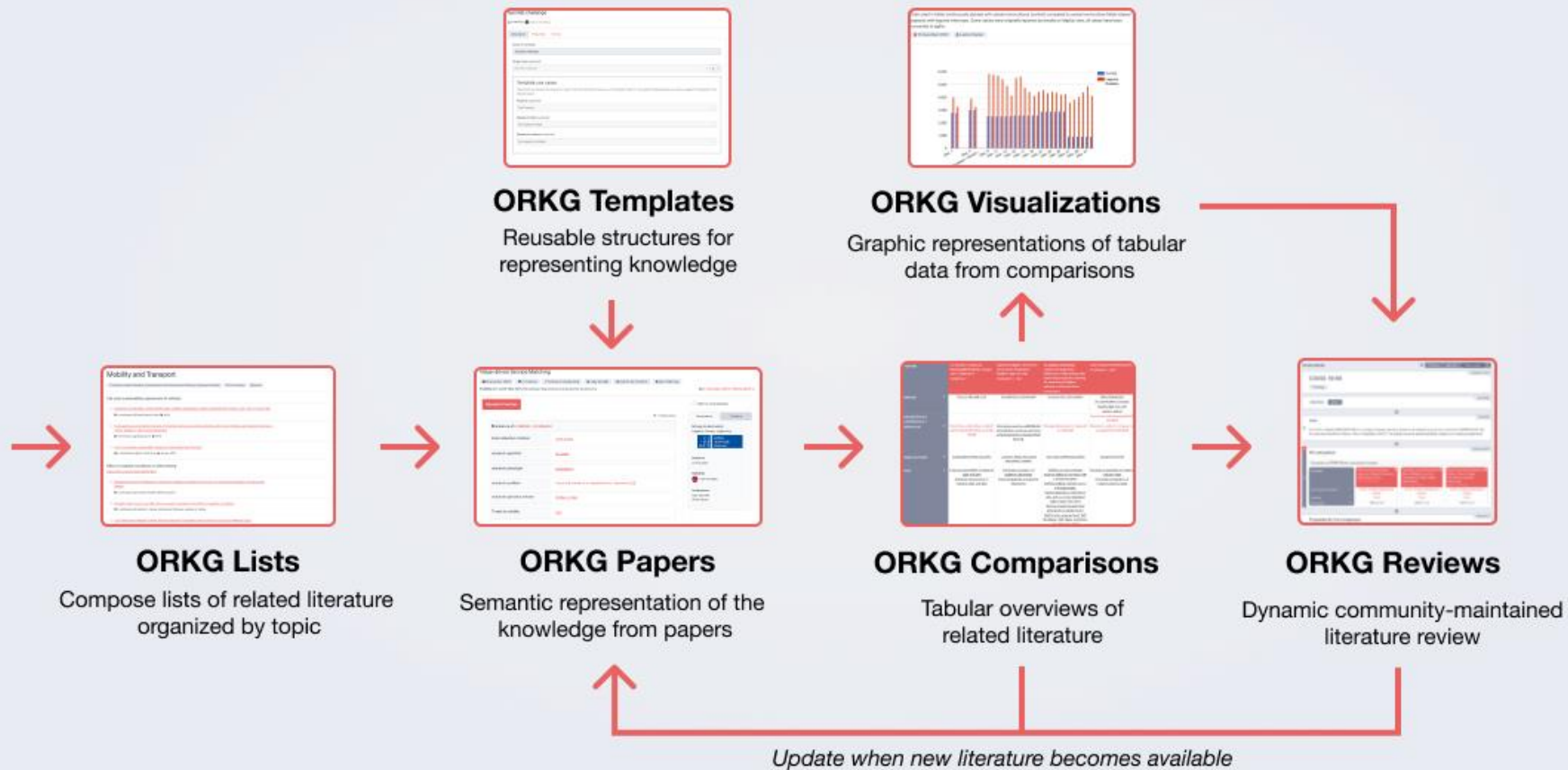
Approach	Year	Classification	URL
Software Feature Request Detection in Issue Tracking Systems	2016	User Feedback Classification	<a href="https://zenodo.org/record/56907#.YKT_NudCRPY">https://zenodo.org/record/56907#.YKT_NudCRPY</a>
Mining User Requirements from Application Store Reviews Using Frame Semantics	2017	User Feedback Classification	<a href="https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip">https://mast.informatik.uni-hamburg.de/wp-content/uploads/2014/03/REJ_data.zip</a>
Mining Twitter Feeds for Software User Requirements	2017	User Feedback Classification	<a href="https://seel.cse.lsu.edu/data/re17.zip">seel.cse.lsu.edu/data/re17.zip</a>
Automatic Classification of Non-Functional Requirements from Augmented Application Reviews	2017	User Feedback Classification	Not available

<https://orkg.org/comparison/R112387/>



# Overview of the ORKG workflow

## Workflow for structured literature reviews using the ORKG



# Current Status of the System

- > 28,000 described papers
- > 6400 research problems
- > 1400 comparisons
- > 350 visualizations
- > 1600 users



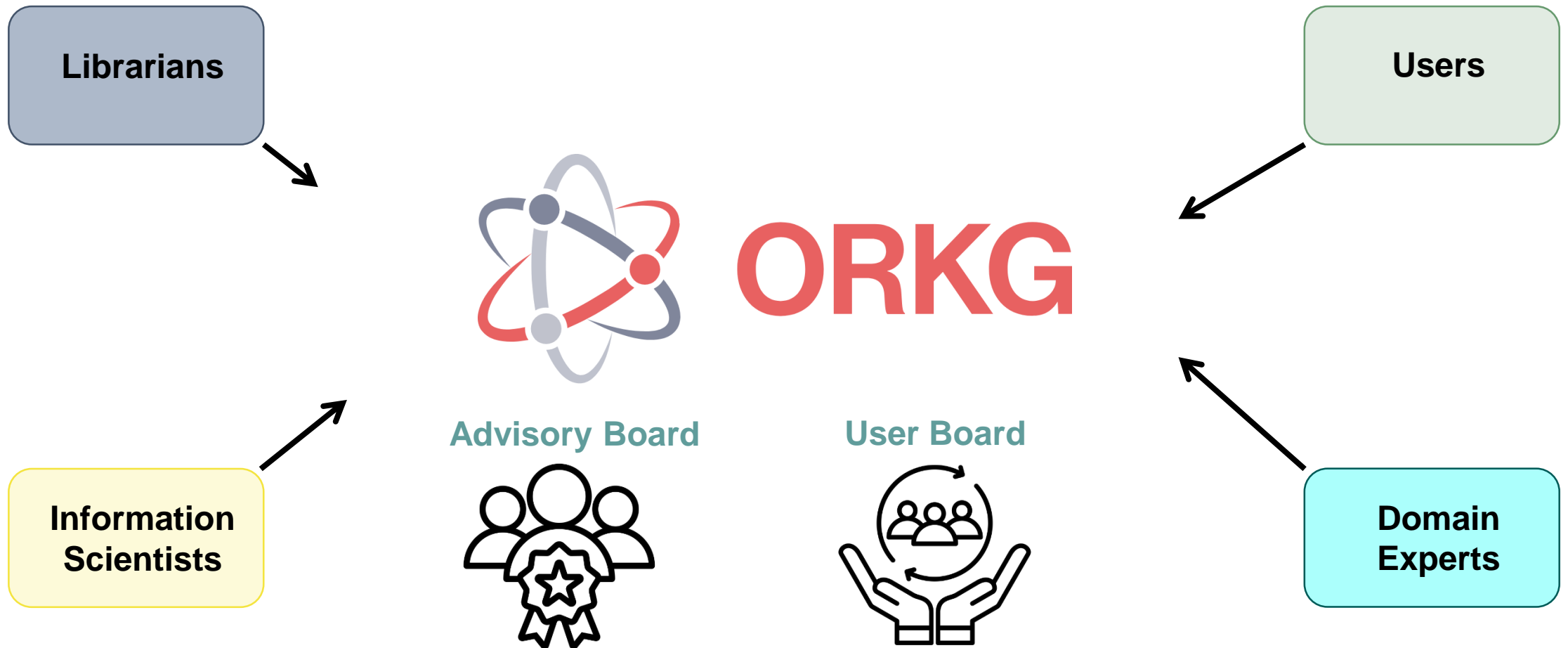
# ORKG Curation – A Crowd-Based Approach

Everyone can create, edit, add, complement, reuse, etc.



# ORKG Curation – A Crowd-Based Approach

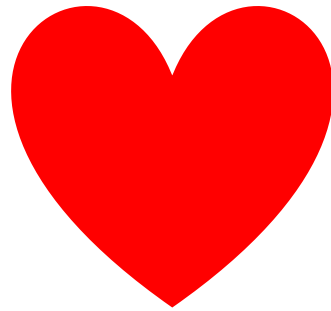
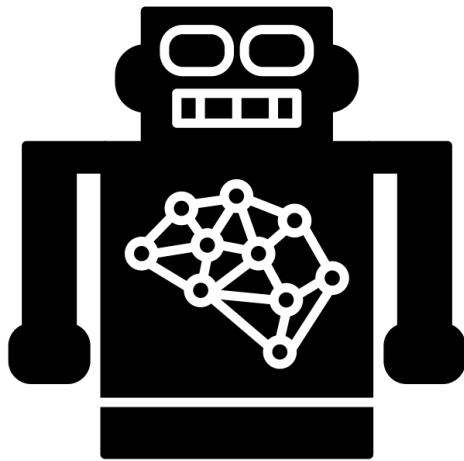
Everyone can create, edit, add, complement, reuse, etc.





# Human-in-the-loop curation model of the ORKG

Better: Machine-actionability AND human-actionability!



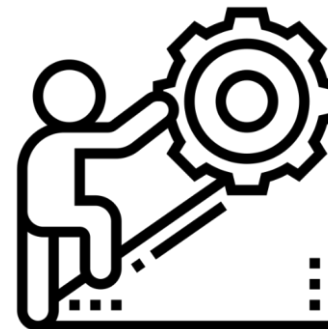
# ORKG Curation Grants



**Domain  
Experts**



**ORKG**

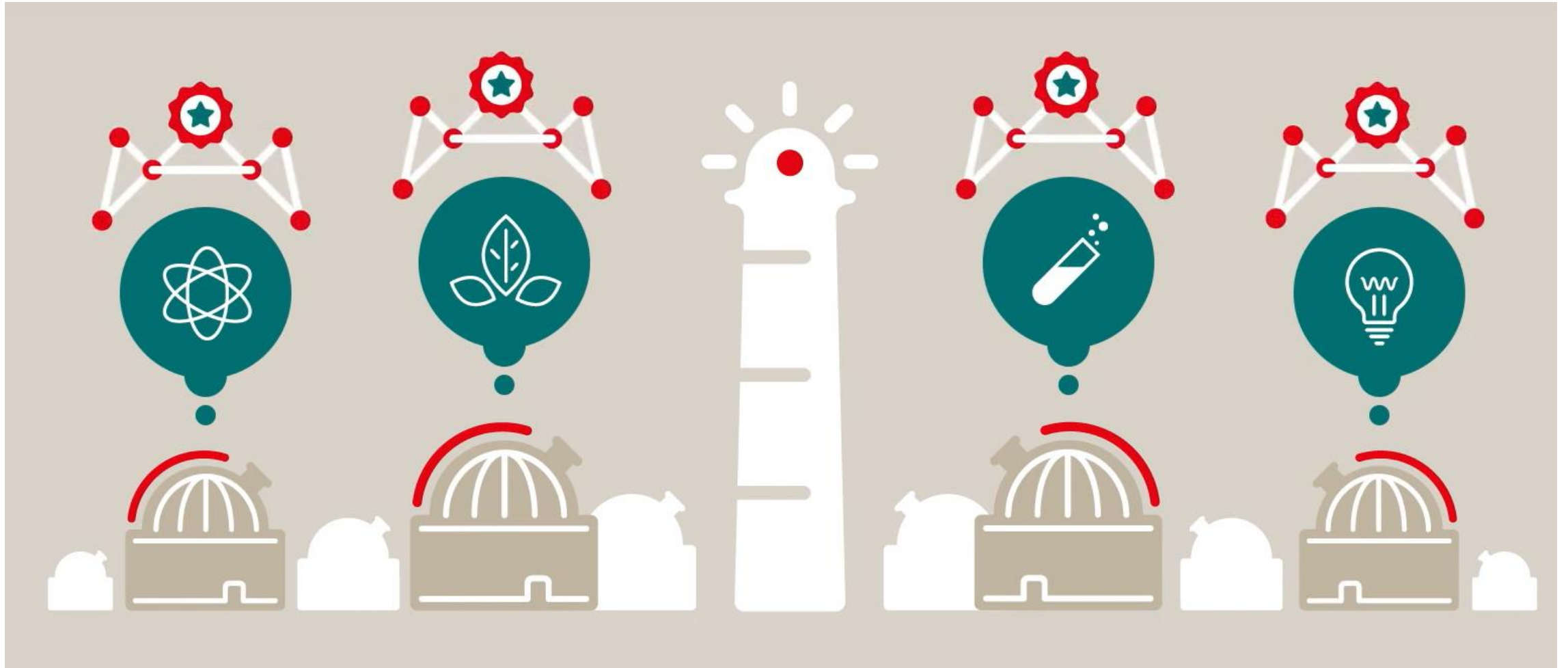


**Semantic  
Experts**

# ORKG comparisons to support Peer-review



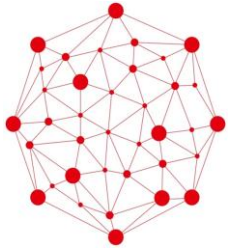
# Observatories: discipline specific knowledge graph



# Scholarly Knowledge: Structured, FAIR, Comparable



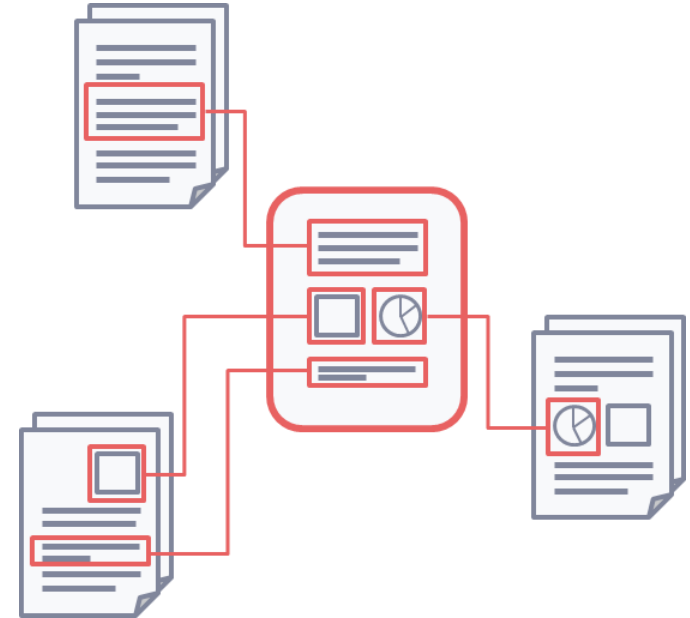
**Rethink scholarly communication:**  
Scholarly work can be realized as expressions other than an article



**Make knowledge  
human and machine actionable**



**Crowd-based approach – Combination  
of domain experts and data curators**





For more information:

[www.orkg.org](http://www.orkg.org)

Follow us:

Mastodon - @orkg@openbiblio.social

Twitter - @orkg\_org

## Contact

[info@orkg.org](mailto:info@orkg.org)

[vinodh.ilangovan@tib.eu](mailto:vinodh.ilangovan@tib.eu)



Creative Commons Attribution 4.0

<https://creativecommons.org/licenses/by/4.0/deed.en>