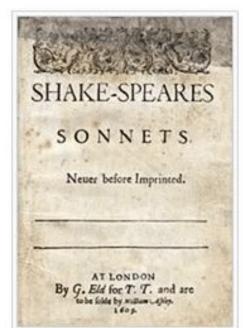
# {} wikicite

## Wikidata as a structured repository of bibliographic data

Dario Taraborelli • Daniel Mietchen • Lydia Pintscher

WikidataCon 2017 • Berlin, 28 October 2017





Title page from 1609 Gedition of Shake-Speares Sonnets.

Published in 1609, the *Sonnets* were the last of Shakespeare's non-dramatic works to be the 154 sonnets was composed, but evidence suggests that Shakespeare wrote sonnets [132] Even before the two unauthorised sonnets appeared in *The Passionate Pilgrim* in 18 Shakespeare's "sugred Sonnets among his private friends". [133] Few analysts believe the intended sequence. [134] He seems to have planned two contrasting series: one about un complexion (the "dark lady"), and one about conflicted love for a fair young man (the "fair represent real individuals, or if the authorial "I" who addresses them represents Shakesp with the sonnets "Shakespeare unlocked his heart". [135]

The 1609 edition was dedicated to a "Mr. W.H.", credited as "the only begetter" of the pool it is not known whether this was written by Shakespeare himself or by the publisher, Thorpe, whose initials appear at the foot of the dedication page; nor is it known who Mr. was, despite numerous theories, or whether Shakespeare even authorised the publication [136] Critics praise the *Sonnets* as a profound meditation on the nature of love, sexual passion, procreation, death, and time. [137]



to curtail the investigation, Trump dismissed Comey as head of the FBI on May 9, 2017.<sup>[27]</sup> In a memo written by James 0 to end the investigation into Mike Flynn..<sup>[28][29][30]</sup> On May 17, 2017, Deputy Attorney General Rod Rosenstein appointed counsel in its investigation.<sup>[31]</sup> The White House has expressed interest in using legal laws to slow and block parts of the

On March 20, 2017, FBI director James Comey testified to the House Intelligence Committee that the FBI has been cond

Russian interference since July 2016, including possible coordination between associates of Trump and Russia. [25][26] Ir

### Russian involvement

stopping the investigation on Jared Kushner and Paul Manafort. [32]

#### Vladimir Putin involvement

We assess Russian President Vladimir Putin ordered an influence campaign in 2016 aimed at the US presidential of Russia's goals were to undermine public faith in the US democratic process, denigrate Secretary Clinton, and harm electability and potential presidency. We further assess Putin and the Russian Government developed a clear preference of President-elect Trump. We have high confidence in these judgments.<sup>[4]</sup>

This assessment<sup>[4]</sup> was published in public, non-classified form in January 2017 by the Office of the Director of National I representing the work of the Federal Bureau of Investigation (FBI), the Central Intelligence Agency (CIA), and the National

Agency (NSA). The FBI and CIA gave the assessment with high confidence and the NSA with moderate confidence.

er cells including liver cells, fibroblasts, and adrenal gland cells.[77] Viral replication triggers tory chemical signals and leads to a septic state. [25] ough contact with mucous membranes or through skin breaks.[40] Once infected, endothelial essels), liver cells, and several types of immune cells such as macrophages, monocytes, ts of infection.[40] Following infection with the virus, the immune cells carry the virus to production of the virus takes place.[40] From there, the virus can enter the bloodstream and nout the body. [40] Macrophages are the first cells infected with the virus, and this infection Other types of white blood cells, such as lymphocytes, also undergo programmed cell death ration of lymphocytes in the blood.[40] This contributes to the weakened immune response nin three days after exposure to the virus. [35] The breakdown of endothelial cells leading to to EBOV glycoproteins. This damage occurs due to the synthesis of Ebola virus glycoprotein f specific integrins responsible for cell adhesion to the intercellular structure and causes liver The widespread bleeding that occurs in affected people causes swelling and shock due to tion in bleeding and clotting commonly seen in EVD has been attributed to increased the coagulation cascade due to excessive tissue factor production by macrophages and Pathogenesis sch

licates very efficiently in many cells, producing large amounts of virus in monocytes,

Uninfected

Endothelial Cell Tox



#### fake news solutions



Q

All

News

Images

Videos

Shopping

More

Settings

Tools

About 5,470,000 results (0.50 seconds)

#### What causes fake news, and what are its solutions? Journalists from ...

www.poynter.org/2017/what-causes-fake-news-and-what...solutions.../458582/ ▼ May 4, 2017 - Folkenflik's comments, made during a Thursday evening panel at The Poynter Institute about the rise of **fake news**, underscored a troubling ...

#### Solutions that can stop fake news spreading - BBC News

www.bbc.com/news/blogs-trending-38769996 •

Jan 30, 2017 - **Fake news** - from false celebrity gossip to the fabricated story of Pope Francis endorsing Donald Trump - became a huge issue during the US ...

#### How to solve Facebook's fake news problem: experts pitch their ideas ...

https://www.theguardian.com → Technology → Facebook ▼

Nov 29, 2016 - If publishing **fake news** was punishable with bans on Facebook then it ... of trying to think up **solutions** to the problem, starting with a rambling ...

#### **Hurricane Kirk**

On August 22, the National Hurricane Center noted a strong tropical wave moving off the coast of Africa, and began monitoring it for further development.[117] The wave traveled west-northwestward into the Central Atlantic while slowly becoming more organized. By August 28th, it finally gained sufficient organization to be classified as a tropical depression by the National Hurricane Center.[118] Later that evening, a satellite pass determined that the depression had strengthened into a tropical storm, and was given the name Kirk at 11 p.m AST (0300 UTC).[119] Kirk later strengthened into a hurricane on August 30.<sup>[120]</sup> It later further intensified into a category 2 hurricane on August 31.[121] Due to cold SSTs and wind shear, Kirk started to weaken. By September 2nd, Kirk was not expected to "live long and prosper" by the NHC as it began to undergo extratropical transition.[122] It became a post-tropical cyclone on September 2 at 5:00 p.m. EDT. Kirk's remnants would hit the UK days later, bringing high winds. [citation needed]



Citations for the sum of all human knowledge

## References in Wikipedia

```
The molecular origins of insulin go at least as far back as
the simplest unicellular [[eukaryotes]].<ref
name='LeRoith'>{{cite journal | vauthors = LeRoith D, Shiloach
J, Heffron R, Rubinovitz C, Tanenbaum R, Roth J │ title =
Insulin-related material in microbes: similarities and
differences from mammalian insulins | journal = Can. J.
Biochem. Cell Biol. | volume = 63 | issue = 8 | pages = 839-49
| year = 1985 | pmid = 3933801 | doi = 10.1139/o85-106
}}</ref> Apart from animals, insulin-like proteins are also
known to exist in Fungi and Protista kingdoms.
```

### 2005

#### Related projects [edit]

and historical discussion

Over the last six years, there have been a number of proposals for new Wikimedia projects and for technological changes to Mediawiki that would provide spaces for summaries, metadata, and discussion of references cited in Wikipedia articles and in other Wikimedia projects. These proposals include:

- Wikicite (metadata proposal) a 2005 proposal to build an "expansion" to Mediawiki that would store and track bibliographic information for citations used in Wikipedia articles
- Wikicat, Wikicite (2006 proposal), and WikiProject Wikicite a series of related new project proposals based on the earlier Wikicite metadata proposal
- Wikitextrose, a model for capturing citations as they are made on other Projects, and for recording the Type of citations [evidentiary, supporting, refuting, &c], intended for use with the above projects.

Outside of Wikimedia, AcaWiki P is an active non-Wikimedia CC BY-licensed wiki that hosts summaries of scholarly articles along with metadata on scholarly works. AcaWiki current hosts about 800 summaries from over 100 contributors.

Other related Wikimedia projects whose goals would be served by a new project proving information citations include:

- Research:Oral Citations a research project offering a rough proposal for building spaces to discuss oral citations in Wikipedia articles
- 'Fact and Reference Check' (en:wp)





Vision
Technology
Community
Scale
Licensing
Independence

the sum of all knowledge.

Imagine a world in which

everyone can freely share in

Imagine a world in which

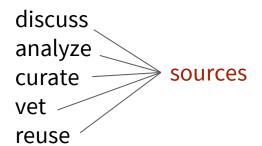
everyonething can freely share in

the sum of all structured knowledge.

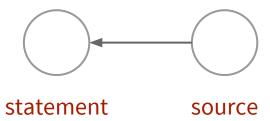
## {} wikicite

Build a structured repository of sources to serve as a foundation of free knowledge

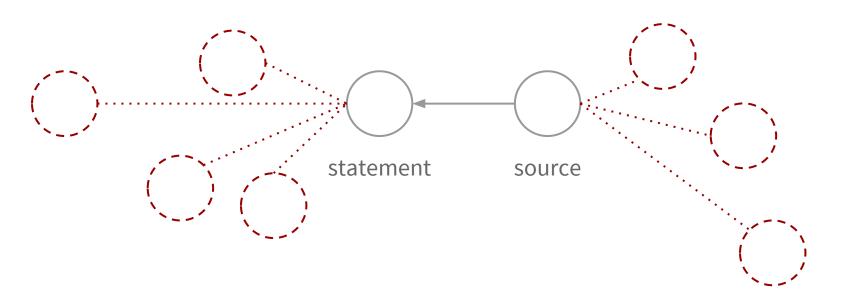
## {} wikicite



## { } wikicite



## { } wikicite



all statements citing a New York Times article

most popular journals cited by statements of any item that is a subclass of economics

all statements citing the works of Joseph Stiglitz

all statements citing journal articles by physicists at Oxford University in the 1970s

all statements citing a journal article that was retracted

which sources are cited in Wikipedia across languages?

which sources are cited in Wikipedia across articles of different quality and importance?

which sources are cited in Wikipedia across topics?

what kinds of citation-related activity do Wikimedia contributors engage in?

what kinds of citation-related activity should Wikimedia contributors prioritize?

are there gaps in citation quality that should be addressed?

UNDERSTANDING THE CONTEXT OF CITATIONS IN WIKIPEDIA <a href="https://meta.wikimedia.org/wiki/Research:Understanding\_the\_context\_of\_citations\_in\_Wikipedia">https://meta.wikimedia.org/wiki/Research:Understanding\_the\_context\_of\_citations\_in\_Wikipedia</a>





(in five years the verb "to wikidata" means look up a fact with literature provenance) "let me wikidata that for you"

RETWEETS LIKES

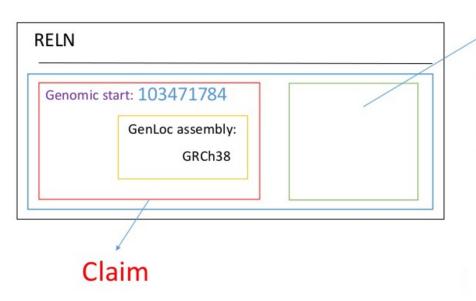
5 8

💸 🦝 AZ 🤎 😂 👬 🔄 👰

9:25 AM - 8 Apr 2016

Zika virus • Q202864 **TAXON** has natural reservoir • P1605 Aedes hensilli • Q14573674 TAXON stated in • P248 Aedes hensilli as a potential vector of Chikungunya and Zika viruses • Q22330738 SCIENTIFIC ARTICLE funded by • P859 **Centers for Disease Control and Prevention • Q583725 GOVERNMENT AGENCY** published in • P1433 PLOS Neglected Tropical Diseases • Q3359737 SCIENTIFIC JOURNAL publisher • P123 **Public Library of Science • Q2333358 PUBLISHER** 

### Computable trust



#### Add References

- 1. Add references
- Check that references concur with the claim or not
- 3. Estimate 'truthiness' of claim
- 4. Provide humans with sources to follow up.

## { } wikicite

Berlin, 25-26 May 2016



## { } wikicite

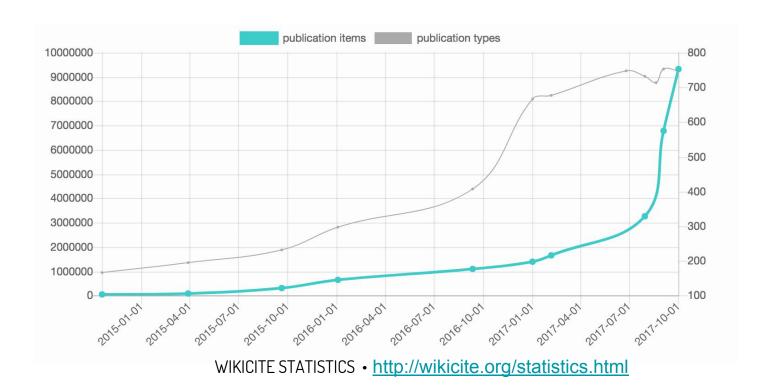
Vienna, 23-25 May 2017



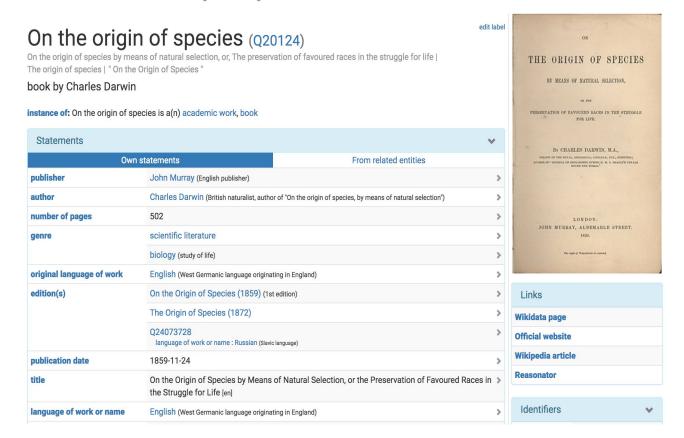


## A growing database of sources stored as Wikidata items

## 9.8 million items, 26% of all Wikidata



## An extensive list of properties to model creative works



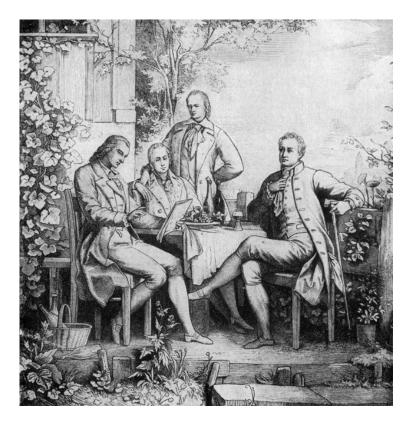
ON THE ORIGIN OF SPECIES • <a href="https://www.wikidata.org/wiki/Q20124">https://www.wikidata.org/wiki/Q20124</a>

The elder Geoffroy and Goethe propounded, at about the same period, their law of compensation or balancement of growth; or, as Goethe expressed it, "in order to spend on one side, nature is forced to economise on the other side." I think this holds true to a certain ex-

From Chapter XI:

species belonging to European genera occur. On the highest mountains of Brazil, some few European genera were found by Gardner, which do not exist in the wide intervening hot countries. So on the Silla of Caraccas the illustrious Humboldt long ago found species belong-

#### Flashback to the WikidataCon opening keynote



- The image <u>depicts</u> <u>Q22670</u> (Friedrich Schiller)

Q77888 (Wilhelm von Humboldt)

**Q6694** (Alexander von Humboldt)

Q5879 (Johann Wolfgang von Goethe)

- They are all authors (P50) of multiple works.
- Darwin <u>cites</u> two of them.
- We don't have a text equivalent of "depicts".

(Adolph Müller; ca. 1797; Scene is in Q2235783 in Jena)

## Usage of some key bibliographic properties

(% growth since Wikimania 2017)

author (P50)	<u>861,086</u>	(+78%)
--------------	----------------	--------

author name string (P2093) 40,276,503 (+135%)

ISBN-13 (P212) often used for books DOI (P356) often used for journal articles

OI (P356) often used for journal articles <u>6,895,447</u> (+135%)

<u>29,933</u> (unchanged)

main subject (P921) 411,545 (+6%)

### All PubMed / PubMed Central citations in Wikipedia

half a million sources cited on the English Wikipedia using a PubMed or PubMed Central identifier are now available as Wikidata entries

- 5. ^ Penicillin at the Free Dictionary ₽
- 6. ^ Garrod, L. P. (1960). "Relative Antibacterial Activity of Three Penicillins". *British Medical Journal*. 1 (5172): 527–29. doi:10.1136/bmj.1.5172.527 doi:
- 7. \* Garrod, L. P. (1960). "The Relative Antibacterial Activity of Four Penicillins" & British Medical Journal. 2 (5214): 1695–6. doi:10.1136/bmj.2.5214.1695 & PMC 2098302 & PMID 13703756 &.
- 8. ^ "Penicillin G and Penicillin V" &. *livertox.nih.gov*. Retrieved 2016-09-25.
- 10. \* "Penicillin G sodium salt Susceptibilty and Resistance Data" (PDF). TOKU-E. Retrieved 4 March 2014.
- 11. ^ a b c Rossi S, ed. (2006). *Australian Medicines Handbook*. Adelaide: Australian Medicines Handbook. ISBN 0-9757919-2-3.

## Broad coverage of scholarly journals



Web of Science currently includes 32,840 scholarly journals –@habib #force2017

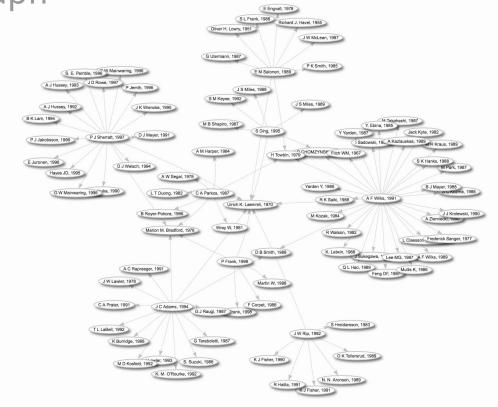




There are now 42,650 items about scholarly journals in @wikidata, up +80% since last year.

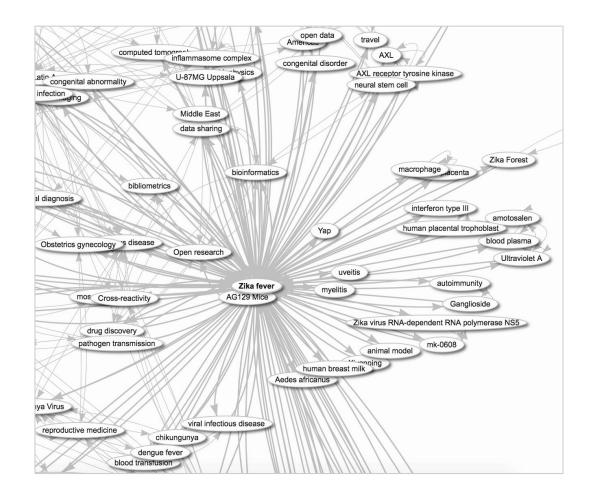
https://twitter.com/Wikicite/status/923134697680629760 https://twitter.com/Wikicite/status/921520882392088576 The Wikidata citation graph

36 million citation links using the *cites* (P2860) property



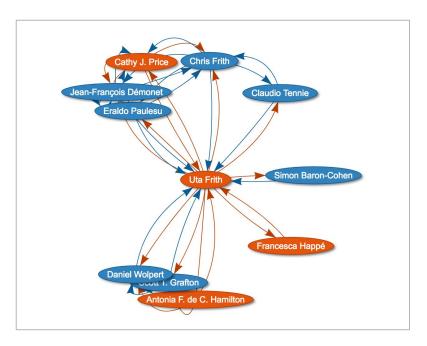
## The Zika corpus

the complete, annotated scholarly publication corpus on Zika virus (Q202864)



### Scholia

### an open source, Wikidata-driven scholarly profile application



CO-AUTHOR GRAPH FOR **UTA FRITH** <a href="https://tools.wmflabs.org/scholia/author/Q8219">https://tools.wmflabs.org/scholia/author/Q8219</a>



LOCATION OF **TURING AWARD** RECIPIENTS <a href="https://tools.wmflabs.org/scholia/award/Q185667">https://tools.wmflabs.org/scholia/award/Q185667</a>

## New ways to indicate the impact of a paper

### Supports the following statement(s)

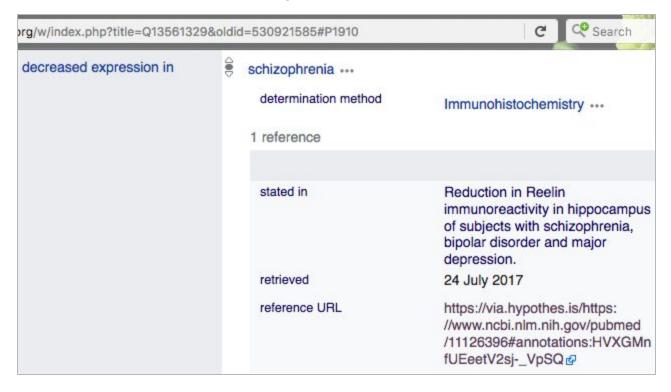
Statements in Wikidata supported by references to this work. Only a maximum of around 2000 statements are shown.

Search:

Item	Property	Value	
Cathelicidin antimicrobial peptide	biological process	antimicrobial humoral immune response mediated by antimicrobial peptide	
Cathelicidin antimicrobial peptide	biological process	innate immune response	
Amyloid beta precursor protein	biological process	antifungal humoral response	
Amyloid beta precursor protein	biological process	antibacterial humoral response	
Amyloid beta precursor protein	biological process	defense response to Gram-negative bacterium	
Amyloid beta precursor protein	biological process	response to yeast	
Amyloid beta precursor protein	biological process	defense response to Gram-positive bacterium	
Amyloid beta precursor protein	biological process	innate immune response	Example: <u>Q21090025</u>

Edit on query.Wikidata.org

### Annotations as a potential sources for statements



Example: Q13561329 (Reelin)

Paper: Reduction in Reelin immunoreactivity in hippocampus of subjects with schizophrenia, bipolar disorder and major depression

### The Initiative for Open Citations

half a billion scholarly citations from 20+ major publishers are now available under CC0







































CORE



















































zotero

### Outreach

**15+ talks and keynotes** at national and international conferences, including:

ELAG 2016 • Wikimania 2016 • VIVO 2016 • COASP 2016 • NIH Data Science lecture series • Crossref LIVE 2016 • DLF 2016 • OpenCon 2016 • Convegno delle Stelline 2017 • OER 2017 • ElixirCONF 2017 • the 2017 European GLAM-Wiki Coordinators meeting • AAAS 2017 • Wikimania 2017 • FORCE 2017 • WikidataCon 2017 • Crossref LIVE 2017 • SWIB 2017







### Technical collaborations













### Funding

### Seven funders and organizations supporting the initiative















# A community







DARIO TARABORELLI /CCO

# Challenges and next steps

1. How to represent **every publication type**?

## Things being cited

book
play
poem
song
psalm
gospel
single
strike
article

chapter edition data set hardback painting software biography manifesto legal case

publication free license open content academic work literary work disease burden literary genre MediaWiki site religious text review article social movement student protest tertiary source research article

literary movement literature review unique identifier Wikimedia project scientific article scientific journal technical standard animated short film political statement animated feature film Internet encyclopedia academic journal article **Creative Commons license** United States executive order United States Military Standard public domain equivalent license **United States Supreme Court decision** Federal Information Processing Standard

import and curate bibliographic information?

2. What **tools and processes** do we need to

#### Source MetaData

Source ID: 10.1016/S0092-8674(0) Check source

You can create a new item, or update an existing one, for this source in QuickStatements:

```
CREATE
LAST
        P356
                "10.1016/S0092-8674(00)80705-9"
LAST
        P31 Q13442814
LAST
                en: "A Novel Family of Mammalian Taste Receptors"
        P1476
LAST
        Len "A Novel Family of Mammalian Taste Receptors"
LAST
        P433
                "6"
LAST
        P304
                "693-702"
LAST
        P478
                "100"
LAST
        P577
                +2000-03-00T00:00:00Z/10
LAST
        P2093
                "Elliot Adler" P1545 "1"
LAST
        P2093
                "Mark A Hoon"
                                 P1545
LAST
        P2093
                "Ken L Mueller"
                                 P1545
                                          "3"
                                                  "4"
LAST
        P2093
                "Jayaram Chandrashekar"
                                         P1545
LAST
        P2093
                "Nicholas J.P Ryba"
                                   P1545
LAST
        P2093
                "Charles S Zuker" P1545
```

Open in QuickStatements

https://tools.wmflabs.org/sourcemd

data we already have vs data we need?

3. How do we get a better understanding of

### Coverage differences between Wikipedia and Wikidata

ISBN citations in Wikipedia

**3 million** in English Wikipedia **90,000** in Dutch Wikipedia

m:R:Scholarly article citations in Wikipedia

ISBN items in Wikidata

**30,000** items with an ISBN-13 (P212)

http://tinyurl.com/y99ukpbn



Wikimedia projects can work together

4. Understand how Wikidata and other

```
Journal [edit source]

{{Cite Q|Q15625490}}

Jeffrey T. Williams; Kent E. Carpenter; James L. Van Tassell; Paul Hoetjes; Wes Toller; Peter Etnoyer; Michael Smith (21 May 2010),

"Biodiversity Assessment of the Fishes of Saba Bank Atoll, Netherlands Antilles" PLOS ONE, 5 (5), PMC 2873961 A, PMID 20505760 A,

doi:10.1371/JOURNAL.PONE.0010676 A, Wikidata Q15625490

{{Cite Q|Q15625490|page=42}}

Jeffrey T. Williams; Kent E. Carpenter; James L. Van Tassell; Paul Hoetjes; Wes Toller; Peter Etnoyer; Michael Smith (21 May 2010),
```

```
{{Cite Q|Q15625490|access-date=18 May 2017}}

Jeffrey T. Williams; Kent E. Carpenter; James L. Van Tassell; Paul Hoetjes; Wes Toller; Peter Etnoyer; Michael Smith (21 May 2010),

"Biodiversity Assessment of the Fishes of Saba Bank Atoll, Netherlands Antilles" & PLOS ONE, 5 (5), PMC 2873961 & PMID 20505760 & PMID 2050
```

"Biodiversity Assessment of the Fishes of Saba Bank Atoll, Netherlands Antilles" 과, PLOS ONE, 5 (5): 42, PMC 2873961 금, PMID 20505760 과,

doi:10.1371/JOURNAL.PONE.0010676 ☑, retrieved 18 May 2017, Wikidata Q15625490

```
{{Cite Q|Q15625490|quote=lorem ipsum}}
```

doi:10.1371/JOURNAL.PONE.0010676 & Wikidata Q15625490

{{Cite Q|Q15625490|page=42|access-date=18 May 2017|quote=lorem ipsum}}

Jeffrey T. Williams; Kent E. Carpenter; James L. Van Tassell; Paul Hoetjes; Wes Toller; Peter Etnoyer; Michael Smith (21 May 2010), "Biodiversity Assessment of the Fishes of Saba Bank Atoll, Netherlands Antilles"型, PLOS ONE, 5 (5), PMC 2873961 ③, PMID 20505760型, doi:10.1371/JOURNAL.PONE.0010676型, "lorem ipsum", Wikidata Q15625490

```
Jeffrey T. Williams; Kent E. Carpenter; James L. Van Tassell; Paul Hoetjes; Wes Toller; Peter Etnoyer; Michael Smith (21 May 2010), "Biodiversity Assessment of the Fishes of Saba Bank Atoll, Netherlands Antilles"虚, PLOS ONE, 5 (5): 42, PMC 2873961 合, PMID 20505760 虚, doi:10.1371/JOURNAL.PONE.0010676 虚, retrieved 18 May 2017, "lorem ipsum", Wikidata Q15625490
```



# Who benefits from WikiCite (we need your help!)

Wikidata contributor

Wikimedia contributor

Librarian

Author

Scholarly publisher

Metadata provider

Digital collection curator

Student



# Thank you

D. Taraborelli, D. Mietchen, L. Pintscher (2017) Wikidata as a structured repository of bibliographic data. *WikidataCon '17.* [CC BY 4.0] doi.org/10.6084/m9.figshare.5548591

### Acknowledgments

Elizabeth Seiver, Jonathan Dugan, Cameron Neylon, Finn Årup Nielsen, Andra Waagmeester, Katherine Thornton, Merrilee Proffitt, Jakob Voss, James Hare, Magnus Manske, Egon Willighagen, Antonin Delpeuch, Charles Matthews, Philipp Zumstein, Alessio Melandri, the <a href="WikiCite 2016">WikiCite 2016</a> and <a href="WikiCite 2017">WikiCite 2017</a> participants, the WikiCite funders, Wikidata Souce Metadata contributors

dario@wikimedia.org • @readermeter daniel.mietchen@okfn.org • @EvoMRI lydia.pintscher@wikimedia.de • @nightrose