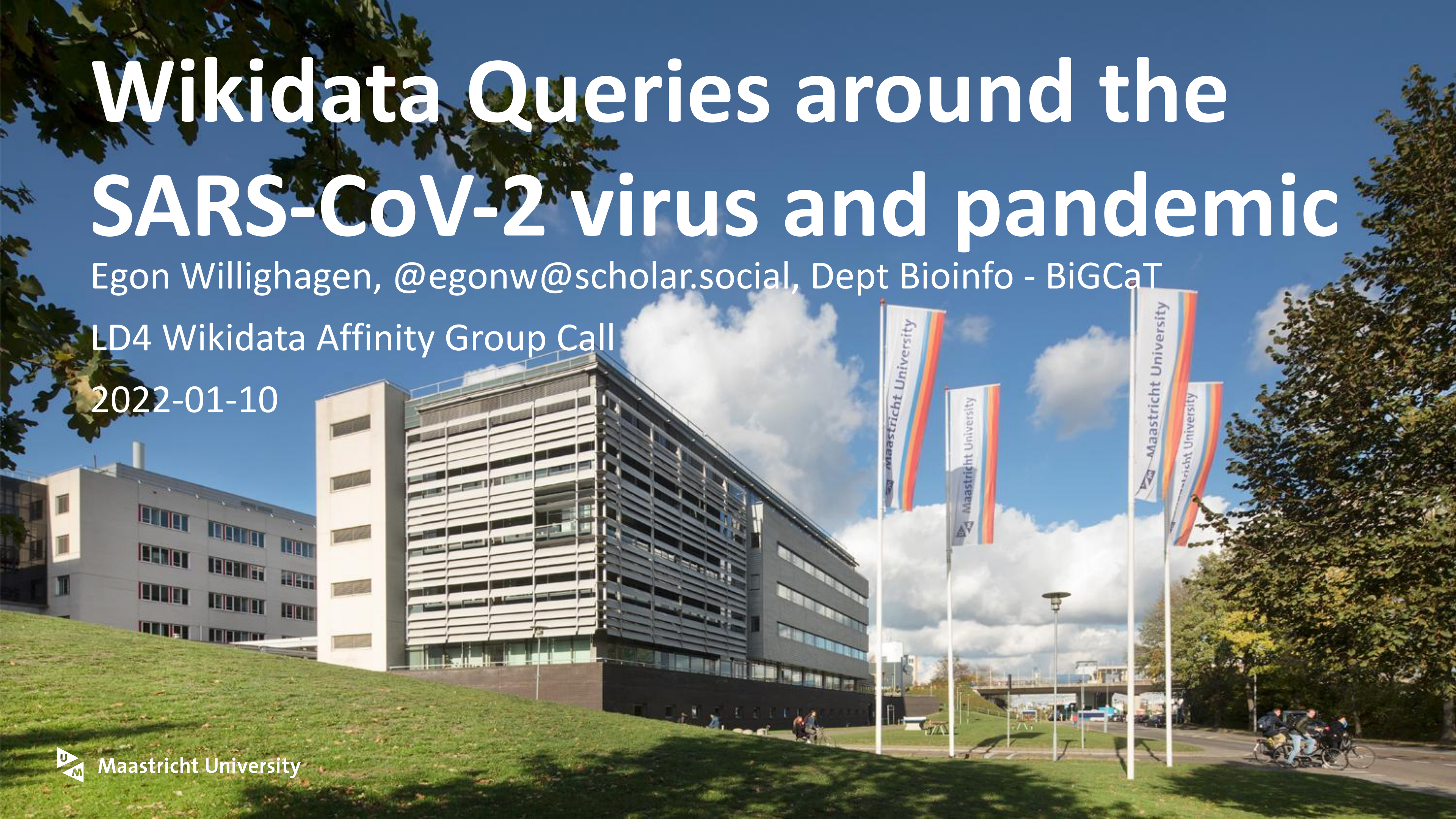


Wikidata Queries around the SARS-CoV-2 virus and pandemic

Egon Willighagen, @egonw@scholar.social, Dept Bioinfo - BiGCaT

LD4 Wikidata Affinity Group Call

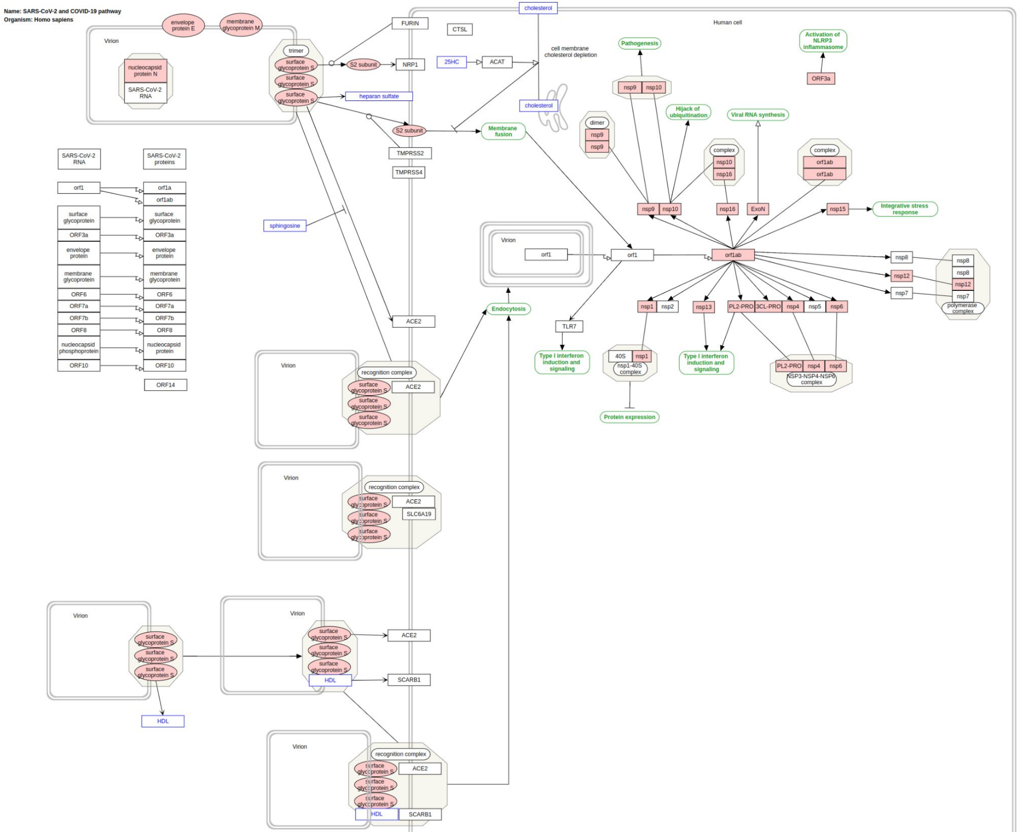
2022-01-10

Maastricht University

The pandemic

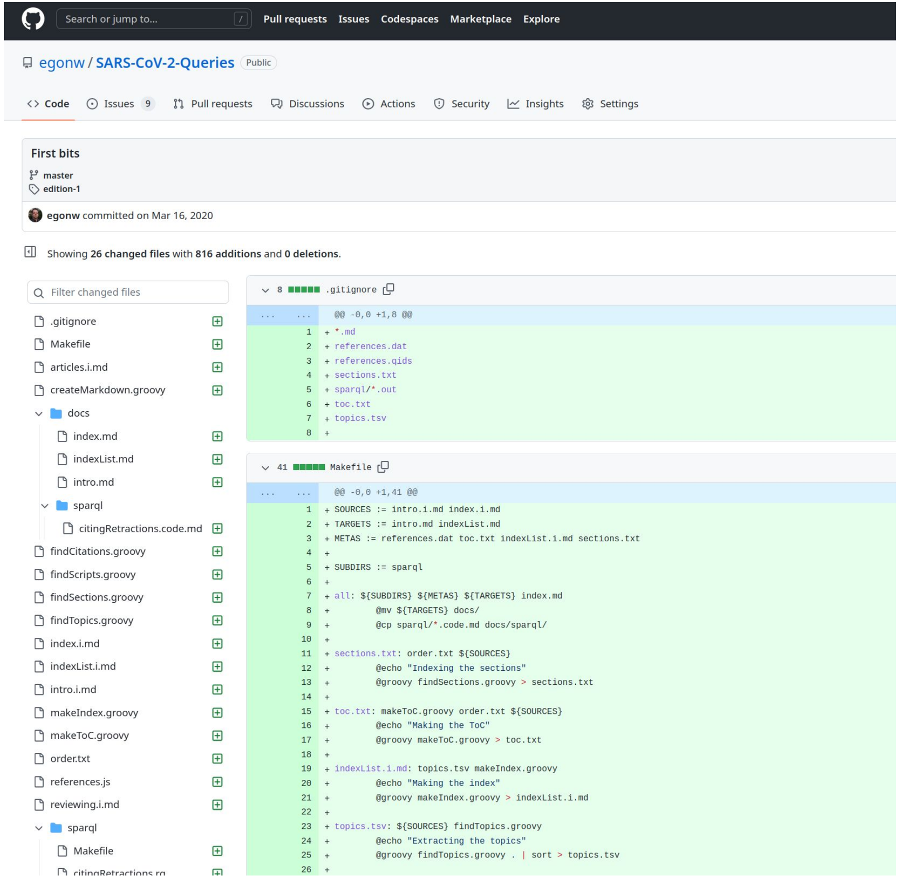
- (cur) (last) ○ 06:27, 30 March 2020 Egonw (Talk | contribs) (31,041 bytes) *(Added another Wikidata identifier)*
- (cur) (last) ○ 18:59, 29 March 2020 Egonw (Talk | contribs) (31,024 bytes) *(Removed another redundant fragment.)*
- (cur) (last) ○ 18:58, 29 March 2020 Egonw (Talk | contribs) (31,601 bytes) *(Removed no-longer existing interaction)*
- (cur) (last) ○ 18:57, 29 March 2020 Egonw (Talk | contribs) (31,937 bytes) *(Added identifiers.)*
- (cur) (last) ○ 17:15, 27 March 2020 Egonw (Talk | contribs) (32,063 bytes) *(A bit more about the virion.)*
- (cur) (last) ○ 14:08, 27 March 2020 Egonw (Talk | contribs) (30,115 bytes) *(Added a bunch of nsps)*
- (cur) (last) ○ 18:49, 26 March 2020 Egonw (Talk | contribs) (19,486 bytes) *(Started drawing the virion structure)*
- (cur) (last) ○ 18:08, 26 March 2020 Egonw (Talk | contribs) (18,062 bytes) *(Split the genes from the proteins, the latter with*
- (cur) (last) ○ 09:53, 17 March 2020 Evelo (Talk | contribs) (15,411 bytes) *(Modified description)*
- (cur) (last) ○ 07:37, 17 March 2020 Egonw (Talk | contribs) (15,572 bytes) *(Modified description)*
- (cur) (last) ○ 20:12, 16 March 2020 Evelo (Talk | contribs) (15,411 bytes) *(Modified description)*
- (cur) (last) ○ 20:10, 16 March 2020 Evelo (Talk | contribs) (15,388 bytes) *(Modified description)*
- (cur) (last) ○ 15:41, 16 March 2020 Egonw (Talk | contribs) (15,124 bytes) *(A bit more of the known mechanism)*
- (cur) (last) ○ 15:23, 16 March 2020 Egonw (Talk | contribs) (13,754 bytes) *(Added TMPRSS2)*
- (cur) (last) ○ 15:12, 16 March 2020 Egonw (Talk | contribs) (13,467 bytes) *(Aligned again.)*
- (cur) (last) ○ 15:10, 16 March 2020 Egonw (Talk | contribs) (13,467 bytes) *(Ordered according to RNA location)*
- (cur) (last) ○ 14:22, 16 March 2020 Egonw (Talk | contribs) (13,467 bytes) *(Removed the link to the other pathway which rai*
- (cur) (last) ○ 14:02, 16 March 2020 Egonw (Talk | contribs) (13,776 bytes) *(Added references about the interaction with ACE*
- (cur) (last) ○ 13:54, 16 March 2020 Egonw (Talk | contribs) (7,415 bytes) *(Added two literature references.)*
- (cur) (last) ○ 13:30, 16 March 2020 L Dupuis (Talk | contribs) (4,767 bytes) *(Ontology Term : 'disease pathway' added !)*
- (cur) (last) ○ 13:28, 16 March 2020 Egonw (Talk | contribs) (4,574 bytes) *(Added the virus, with a link to Wikidata.)*
- (cur) (last) ○ 13:22, 16 March 2020 Egonw (Talk | contribs) (4,340 bytes) *(Other gene identifiers.)*
- (cur) (last) ○ 13:04, 16 March 2020 Egonw (Talk | contribs) (4,133 bytes) *(Added a link to the other pathway.)*
- (cur) (last) ○ 12:59, 16 March 2020 Egonw (Talk | contribs) (3,779 bytes) *(Modified description)*
- (cur) (last) ○ 12:52, 16 March 2020 Egonw (Talk | contribs) (3,699 bytes) *(Added the first two gene identifiers.)*
- (cur) (last) ○ 12:43, 16 March 2020 L Dupuis (Talk | contribs) (3,262 bytes) *(Ontology Term : 'disease by infectious agent' added !)*
- (cur) (last) ○ 12:40, 16 March 2020 Egonw (Talk | contribs) (3,053 bytes) *(Added the virus genes.)*
- (cur) (last) ○ 12:16, 16 March 2020 Egonw (Talk | contribs) (447 bytes) *(New pathway)*

(Latest | Earliest) View (newer 50) (older 50) (20 | 50 | 100 | 250 | 500)



https://new.wikipathways.org/pathways/WP4846.html

The eBook



egonw / SARS-CoV-2-Queries

First bits

master
edition-1

egonw committed on Mar 16, 2020

Showing **26 changed files** with **816 additions** and **0 deletions**.

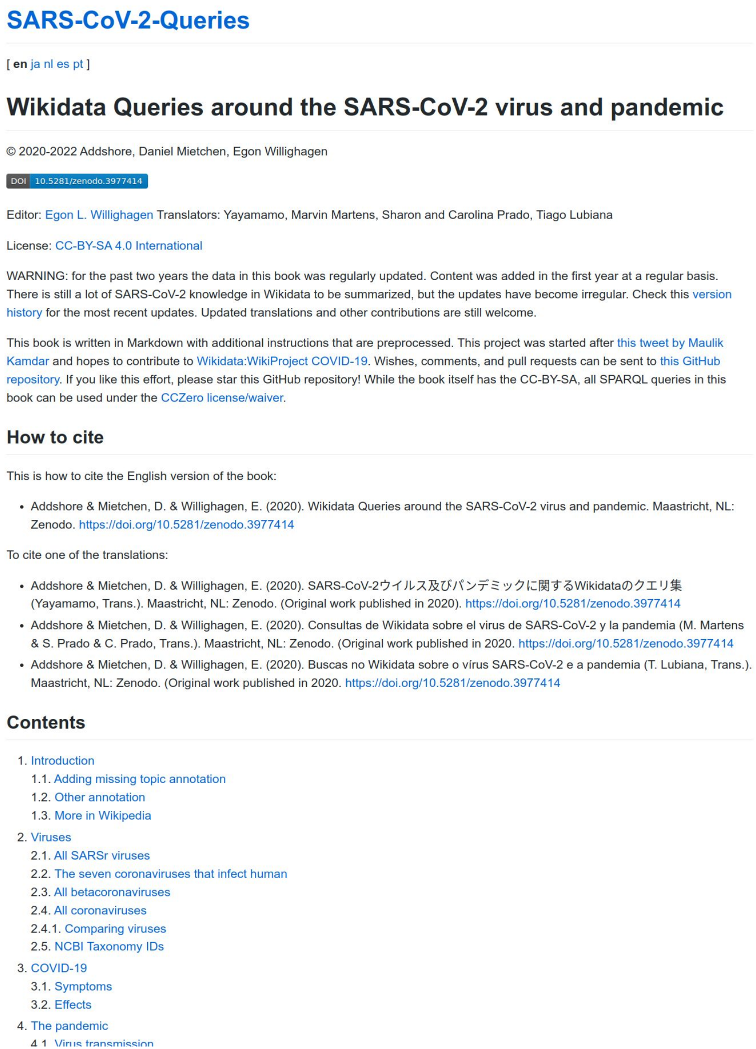
- .gitignore
- Makefile
- articles.i.md
- createMarkdown.groovy
- docs
  - index.md
  - indexList.md
  - intro.md
  - sparql
    - citingRetractions.code.md
- findCitations.groovy
- findScripts.groovy
- findSections.groovy
- findTopics.groovy
- index.i.md
- indexList.i.md
- intro.i.md
- makeIndex.groovy
- makeToC.groovy
- order.txt
- references.js
- reviewing.i.md
- sparql
  - Makefile
  - citingRetractions.rq

8 .gitignore

```
@@ -0,0 +1,8 @@
+ *.md
+ references.dat
+ references.qids
+ sections.txt
+ sparql/*.out
+ toc.txt
+ topics.tsv
+
```

41 Makefile

```
@@ -0,0 +1,41 @@
+ SOURCES := intro.i.md index.i.md
+ TARGETS := intro.md indexList.md
+ METAS := references.dat toc.txt indexList.i.md sections.txt
+
+ SUBDIRS := sparql
+
+ all: ${SUBDIRS} ${METAS} ${TARGETS} index.md
+         @mv ${TARGETS} docs/
+         @cp sparql/*.code.md docs/sparql/
+
+ sections.txt: order.txt ${SOURCES}
+         @echo "Indexing the sections"
+         @groovy findSections.groovy > sections.txt
+
+ toc.txt: makeToC.groovy order.txt ${SOURCES}
+         @echo "Making the ToC"
+         @groovy makeToC.groovy > toc.txt
+
+ indexList.i.md: topics.tsv makeIndex.groovy
+         @echo "Making the index"
+         @groovy makeIndex.groovy > indexList.i.md
+
+ topics.tsv: ${SOURCES} findTopics.groovy
+         @echo "Extracting the topics"
+         @groovy findTopics.groovy . | sort > topics.tsv
+
```



SARS-CoV-2-Queries

[**en** ja nl es pt]

Wikidata Queries around the SARS-CoV-2 virus and pandemic

© 2020-2022 Addshore, Daniel Mietchen, Egon Willighagen

DOI 10.5281/zenodo.3977414

Editor: Egon L. Willighagen Translators: Yayamamo, Marvin Martens, Sharon and Carolina Prado, Tiago Lubiana

WARNING: for the past two years the data in this book was regularly updated. Content was added in the first year at a regular basis. There is still a lot of SARS-CoV-2 knowledge in Wikidata to be summarized, but the updates have become irregular. Check this version history for the most recent updates. Updated translations and other contributions are still welcome.

This book is written in Markdown with additional instructions that are preprocessed. This project was started after this tweet by Maulik Kamdar and hopes to contribute to Wikidata:WikiProject COVID-19. Wishes, comments, and pull requests can be sent to this GitHub repository. If you like this effort, please star this GitHub repository! While the book itself has the CC-BY-SA, all SPARQL queries in this book can be used under the CCZero license/waiver.

How to cite

This is how to cite the English version of the book:

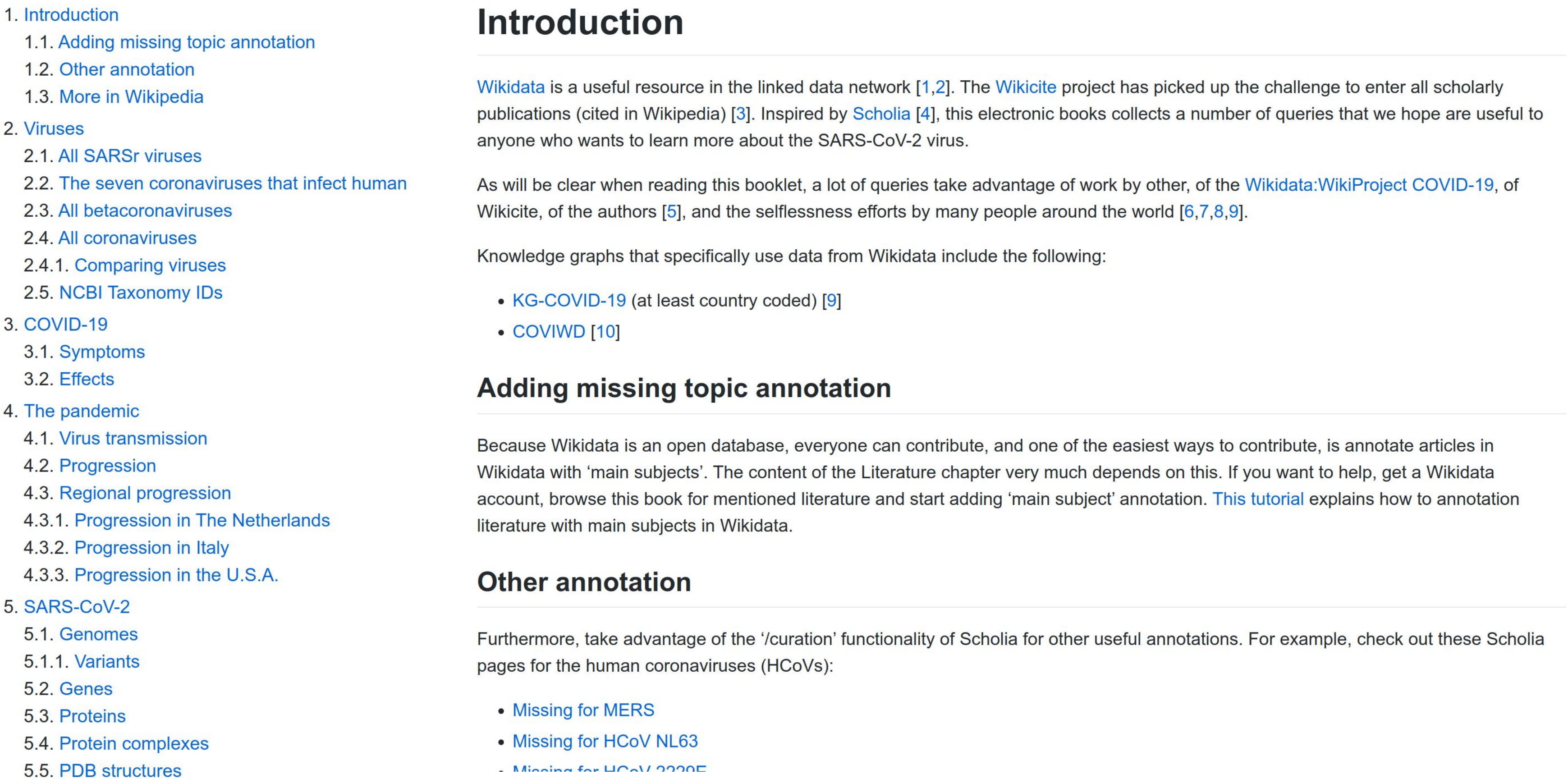
- Addshore & Mietchen, D. & Willighagen, E. (2020). Wikidata Queries around the SARS-CoV-2 virus and pandemic. Maastricht, NL: Zenodo. https://doi.org/10.5281/zenodo.3977414

To cite one of the translations:

- Addshore & Mietchen, D. & Willighagen, E. (2020). SARS-CoV-2ウイルス及びパンデミックに関するWikidataのクエリ集 (Yayamamo, Trans.). Maastricht, NL: Zenodo. (Original work published in 2020). https://doi.org/10.5281/zenodo.3977414
- Addshore & Mietchen, D. & Willighagen, E. (2020). Consultas de Wikidata sobre el virus de SARS-CoV-2 y la pandemia (M. Martens & S. Prado & C. Prado, Trans.). Maastricht, NL: Zenodo. (Original work published in 2020. https://doi.org/10.5281/zenodo.3977414
- Addshore & Mietchen, D. & Willighagen, E. (2020). Buscas no Wikidata sobre o vírus SARS-CoV-2 e a pandemia (T. Lubiana, Trans.). Maastricht, NL: Zenodo. (Original work published in 2020. https://doi.org/10.5281/zenodo.3977414

Contents

The eBook



Introduction

Wikidata is a useful resource in the linked data network [1,2]. The Wikicite project has picked up the challenge to enter all scholarly publications (cited in Wikipedia) [3]. Inspired by Scholia [4], this electronic books collects a number of queries that we hope are useful to anyone who wants to learn more about the SARS-CoV-2 virus.

As will be clear when reading this booklet, a lot of queries take advantage of work by other, of the Wikidata:WikiProject COVID-19, of Wikicite, of the authors [5], and the selflessness efforts by many people around the world [6,7,8,9].

Knowledge graphs that specifically use data from Wikidata include the following:

- KG-COVID-19 (at least country coded) [9]
- COVIWD [10]

Adding missing topic annotation

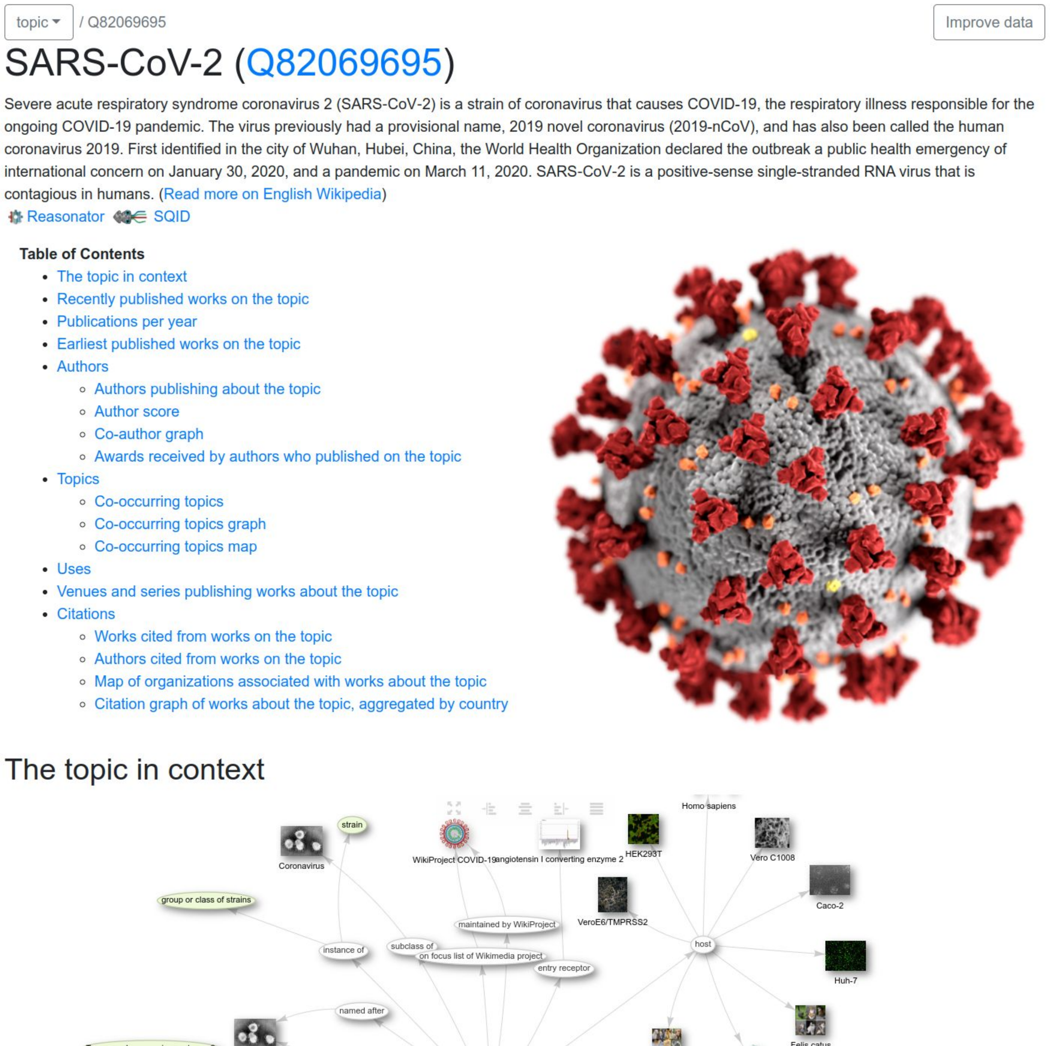
Because Wikidata is an open database, everyone can contribute, and one of the easiest ways to contribute, is annotate articles in Wikidata with ‘main subjects’. The content of the Literature chapter very much depends on this. If you want to help, get a Wikidata account, browse this book for mentioned literature and start adding ‘main subject’ annotation. This tutorial explains how to annotation literature with main subjects in Wikidata.

Other annotation

Furthermore, take advantage of the ‘/curation’ functionality of Scholia for other useful annotations. For example, check out these Scholia pages for the human coronaviruses (HCoVs):

- Missing for MERS
- Missing for HCoV NL63
- Missing for HCoV 229E

Scholia



/ Q82069695

SARS-CoV-2 (Q82069695)

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a strain of coronavirus that causes COVID-19, the respiratory illness responsible for the ongoing COVID-19 pandemic. The virus previously had a provisional name, 2019 novel coronavirus (2019-nCoV), and has also been called the human coronavirus 2019. First identified in the city of Wuhan, Hubei, China, the World Health Organization declared the outbreak a public health emergency of international concern on January 30, 2020, and a pandemic on March 11, 2020. SARS-CoV-2 is a positive-sense single-stranded RNA virus that is contagious in humans. (Read more on English Wikipedia)

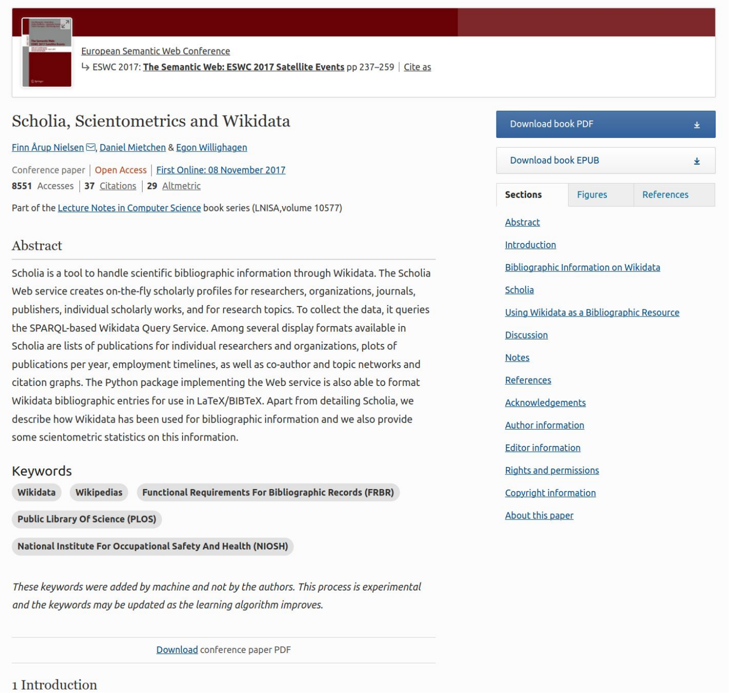
Reasonator SQID

Table of Contents

The topic in context

Maastricht University

Scholia



European Semantic Web Conference
↳ ESWC 2017: **The Semantic Web: ESWC 2017 Satellite Events** pp 237–259 | Cite as

Scholia, Scientometrics and Wikidata

Finn Årup Nielsen, Daniel Mietchen & Egon Willighagen

Conference paper | Open Access | First Online: 08 November 2017
8551 Accesses | **37** Citations | **29** Altmetric

Part of the Lecture Notes in Computer Science book series (LNISA,volume 10577)

Abstract

Scholia is a tool to handle scientific bibliographic information through Wikidata. The Scholia Web service creates on-the-fly scholarly profiles for researchers, organizations, journals, publishers, individual scholarly works, and for research topics. To collect the data, it queries the SPARQL-based Wikidata Query Service. Among several display formats available in Scholia are lists of publications for individual researchers and organizations, plots of publications per year, employment timelines, as well as co-author and topic networks and citation graphs. The Python package implementing the Web service is also able to format Wikidata bibliographic entries for use in LaTeX/BIBTeX. Apart from detailing Scholia, we describe how Wikidata has been used for bibliographic information and we also provide some scientometric statistics on this information.

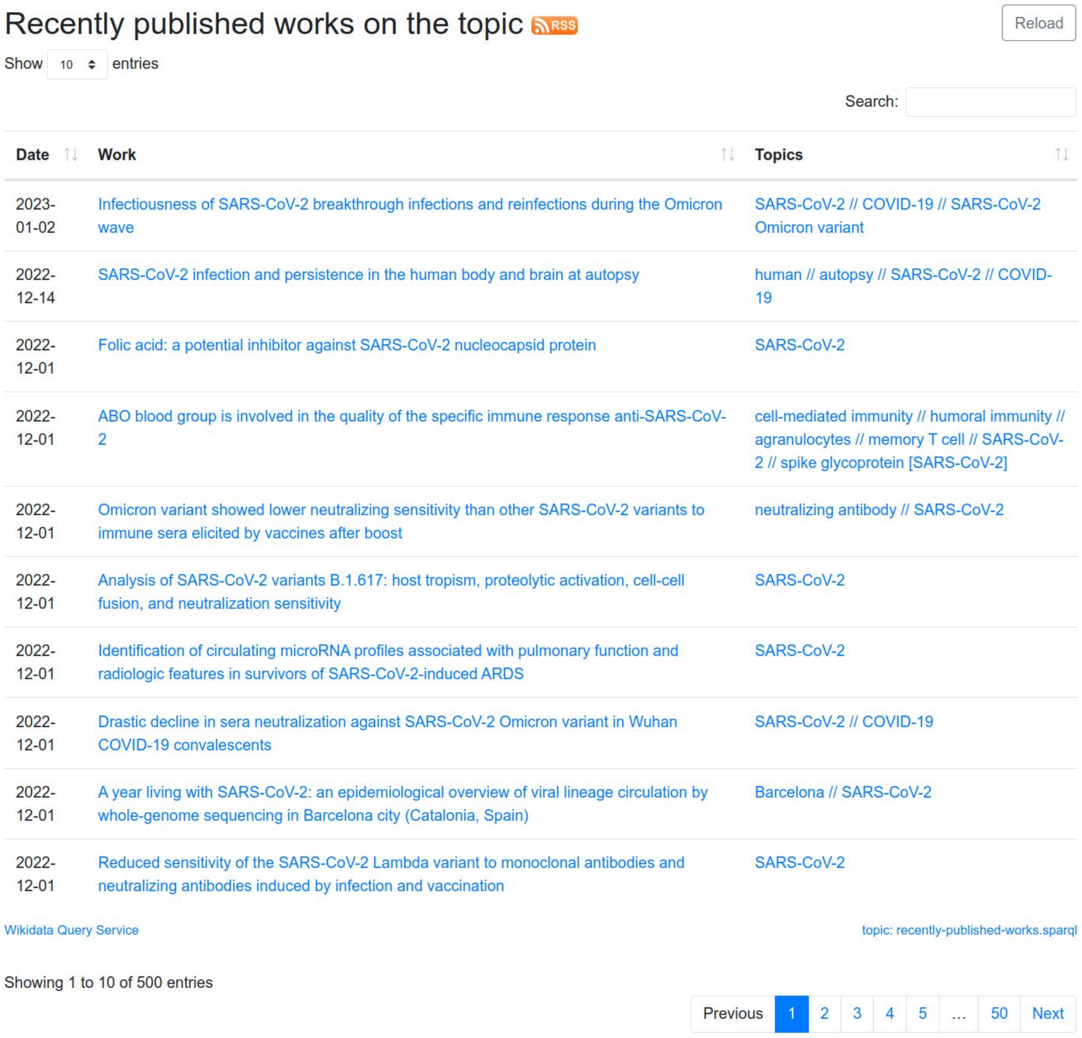
Keywords

Wikidata **Wikipedias** **Functional Requirements For Bibliographic Records (FRBR)** **Public Library Of Science (PLOS)** **National Institute For Occupational Safety And Health (NIOSH)**

These keywords were added by machine and not by the authors. This process is experimental and the keywords may be updated as the learning algorithm improves.

Download conference paper PDF

1 Introduction



Recently published works on the topic

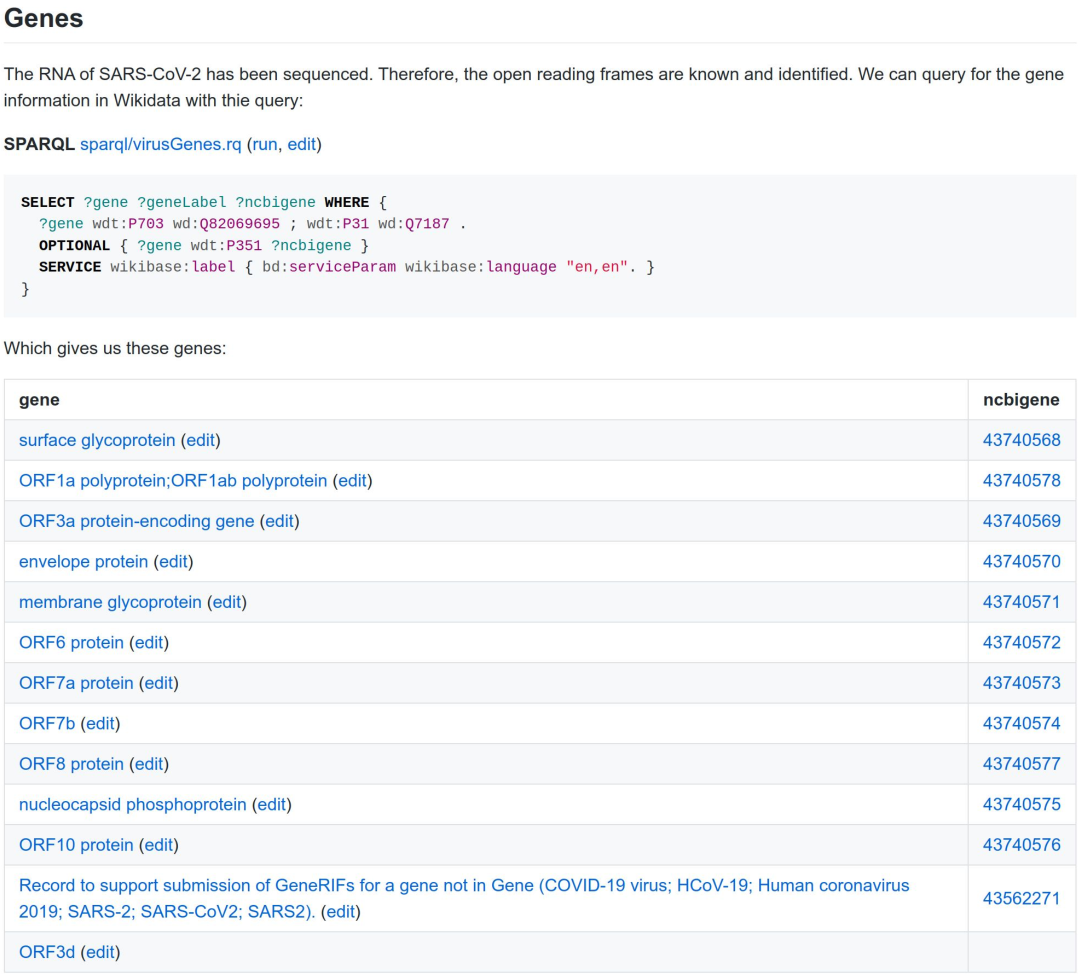
Date	Work	Topics
2023-01-02	Infectiousness of SARS-CoV-2 breakthrough infections and reinfections during the Omicron wave	SARS-CoV-2 // COVID-19 // SARS-CoV-2 Omicron variant
2022-12-14	SARS-CoV-2 infection and persistence in the human body and brain at autopsy	human // autopsy // SARS-CoV-2 // COVID-19
2022-12-01	Folic acid: a potential inhibitor against SARS-CoV-2 nucleocapsid protein	SARS-CoV-2
2022-12-01	ABO blood group is involved in the quality of the specific immune response anti-SARS-CoV-2	cell-mediated immunity // humoral immunity // agranulocytes // memory T cell // SARS-CoV-2 // spike glycoprotein [SARS-CoV-2]
2022-12-01	Omicron variant showed lower neutralizing sensitivity than other SARS-CoV-2 variants to immune sera elicited by vaccines after boost	neutralizing antibody // SARS-CoV-2
2022-12-01	Analysis of SARS-CoV-2 variants B.1.617: host tropism, proteolytic activation, cell-cell fusion, and neutralization sensitivity	SARS-CoV-2
2022-12-01	Identification of circulating microRNA profiles associated with pulmonary function and radiologic features in survivors of SARS-CoV-2-induced ARDS	SARS-CoV-2
2022-12-01	Drastic decline in sera neutralization against SARS-CoV-2 Omicron variant in Wuhan COVID-19 convalescents	SARS-CoV-2 // COVID-19
2022-12-01	A year living with SARS-CoV-2: an epidemiological overview of viral lineage circulation by whole-genome sequencing in Barcelona city (Catalonia, Spain)	Barcelona // SARS-CoV-2
2022-12-01	Reduced sensitivity of the SARS-CoV-2 Lambda variant to monoclonal antibodies and neutralizing antibodies induced by infection and vaccination	SARS-CoV-2

Wikidata Query Service

topic: recently-published-works.sparql

Showing 1 to 10 of 500 entries

The eBook



Genes

The RNA of SARS-CoV-2 has been sequenced. Therefore, the open reading frames are known and identified. We can query for the gene information in Wikidata with thie query:

SPARQL sparql/virusGenes.rq (run, edit)

```
SELECT ?gene ?geneLabel ?ncbigene WHERE {
  ?gene wdt:P703 wd:Q82069695 ; wdt:P31 wd:Q7187 .
  OPTIONAL { ?gene wdt:P351 ?ncbigene }
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en,en". }
}
```

Which gives us these genes:

gene	ncbigene
surface glycoprotein (edit)	43740568
ORF1a polyprotein;ORF1ab polyprotein (edit)	43740578
ORF3a protein-encoding gene (edit)	43740569
envelope protein (edit)	43740570
membrane glycoprotein (edit)	43740571
ORF6 protein (edit)	43740572
ORF7a protein (edit)	43740573
ORF7b (edit)	43740574
ORF8 protein (edit)	43740577
nucleocapsid phosphoprotein (edit)	43740575
ORF10 protein (edit)	43740576
Record to support submission of GeneRIFs for a gene not in Gene (COVID-19 virus; HCoV-19; Human coronavirus 2019; SARS-2; SARS-CoV2; SARS2). (edit)	43562271
ORF3d (edit)	

Virus proteins and protein complexes

BMC Biology

Home About Articles Submission Guidelines

Methodology article | Open Access | Published: 22 January 2021

A protocol for adding knowledge to Wikidata: aligning resources on human coronaviruses

Andra Waagmeester, Egon L. Willighagen, Andrew I. Su, Martina Kutmon, Jose Emilio Labra Gayo, Daniel Fernández-Álvarez, Quentin Groom, Peter J. Schaap, Lisa M. Verhagen & Jasper J. Koehorst

BMC Biology **19**, Article number: 12 (2021) | Cite this article

3470 Accesses | **9** Citations | **52** Altmetric | Metrics

Download PDF

Sections Figures References

- Abstract
- Background
- Results
- Discussion
- Conclusion
- Methods
- Availability of data and materials
- References
- Acknowledgements
- Funding
- Author information
- Ethics declarations
- Additional information
- Supplementary Information
- Rights and permissions
- About this article

Abstract

Background

Pandemics, even more than other medical problems, require swift integration of knowledge. When caused by a new virus, understanding the underlying biology may help finding solutions. In a setting where there are a large number of loosely related projects and initiatives, we need common ground, also known as a “commons.” Wikidata, a public knowledge graph aligned with Wikipedia, is such a commons and uses unique identifiers to link knowledge in other knowledge bases. However, Wikidata may not always have the right schema for the urgent questions. In this paper, we address this problem by showing how a data schema required for the integration can be modeled with entity schemas represented by Shape Expressions.

Results

As a telling example, we describe the process of aligning resources on the genomes and proteomes of the SARS-CoV-2 virus and related viruses as well as how Shape Expressions can be defined for Wikidata to model the knowledge, helping others studying the SARS-CoV-2 pandemic. How this model can be used to make data between various resources interoperable is demonstrated by integrating data from NCBI (National Center for Biotechnology Information) Taxonomy, NCBI Genes, UniProt, and WikiPathways. Based on that model, a set of automated applications or bots were written for regular updates of these sources in Wikidata and added to a platform for automatically running these updates.

Conclusions

Although this workflow is developed and applied in the context of the COVID-19 pandemic, to demonstrate its broader applicability it was also applied to other human coronaviruses (MERS, SARS, human coronavirus NL63, human coronavirus 229E, human coronavirus HKU1, human coronavirus OC4).

Background

The coronavirus disease 2019 (COVID-19) pandemic, caused by the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) virus, is leading to a burst of swiftly released scientific publications on the matter [1]. In response to the pandemic, many research groups have started projects to understand the SARS-CoV-2 virus life cycle and to find solutions. Examples of the numerous projects include outbreak.info [2], Virus Outbreak Data Network (VODAN) [3], CORD-19-on-FHIR [4], KG-COVID-19 knowledge graph [5], and the COVID-19 Disease Map [6]. Many research papers and preprints get published every week and many call for more Open Science [7]. The Dutch universities went a step further and want to make any previously published research openly available, in whatever way related to COVID-19 research.

However, this swift release of research findings comes with an increased number of incorrect interpretations [8] which can be problematic when new research articles are picked up by

Nucleic Acids Research

Issues Section browse ▾ Advance articles Submit ▾ Purchase About ▾ Nucleic Acids Research

Volume 50, Issue D1
7 January 2022

Article Contents

- Abstract
- INTRODUCTION
- CONTENT
- WEBSITE SEARCH AND DATA VISUALIZATION
- DATA ACCESS OPTIONS
- COLLABORATION AND COMMUNITY INVOLVEMENT
- SUMMARY AND FUTURE PLANS
- DATA AVAILABILITY
- ACKNOWLEDGEMENTS
- FUNDING
- REFERENCES

Comments (0)

< Previous Next >

JOURNAL ARTICLE

Complex Portal 2022: new curation frontiers

Birgit H M Meldal, Livia Perfetto, Colin Combe, Tiago Lubiana, João Vitor Ferreira Cavalcante, Hema Bye-A-Jee, Andra Waagmeester, Noemi del-Toro, Anjali Shrivastava, Elisabeth Barrera ... Show more

Nucleic Acids Research, Volume 50, Issue D1, 7 January 2022, Pages D578–D586, https://doi.org/10.1093/nar/gkab991

Published: 29 October 2021 **Article history** ▾

PDF Split View Cite Permissions Share ▾

Abstract

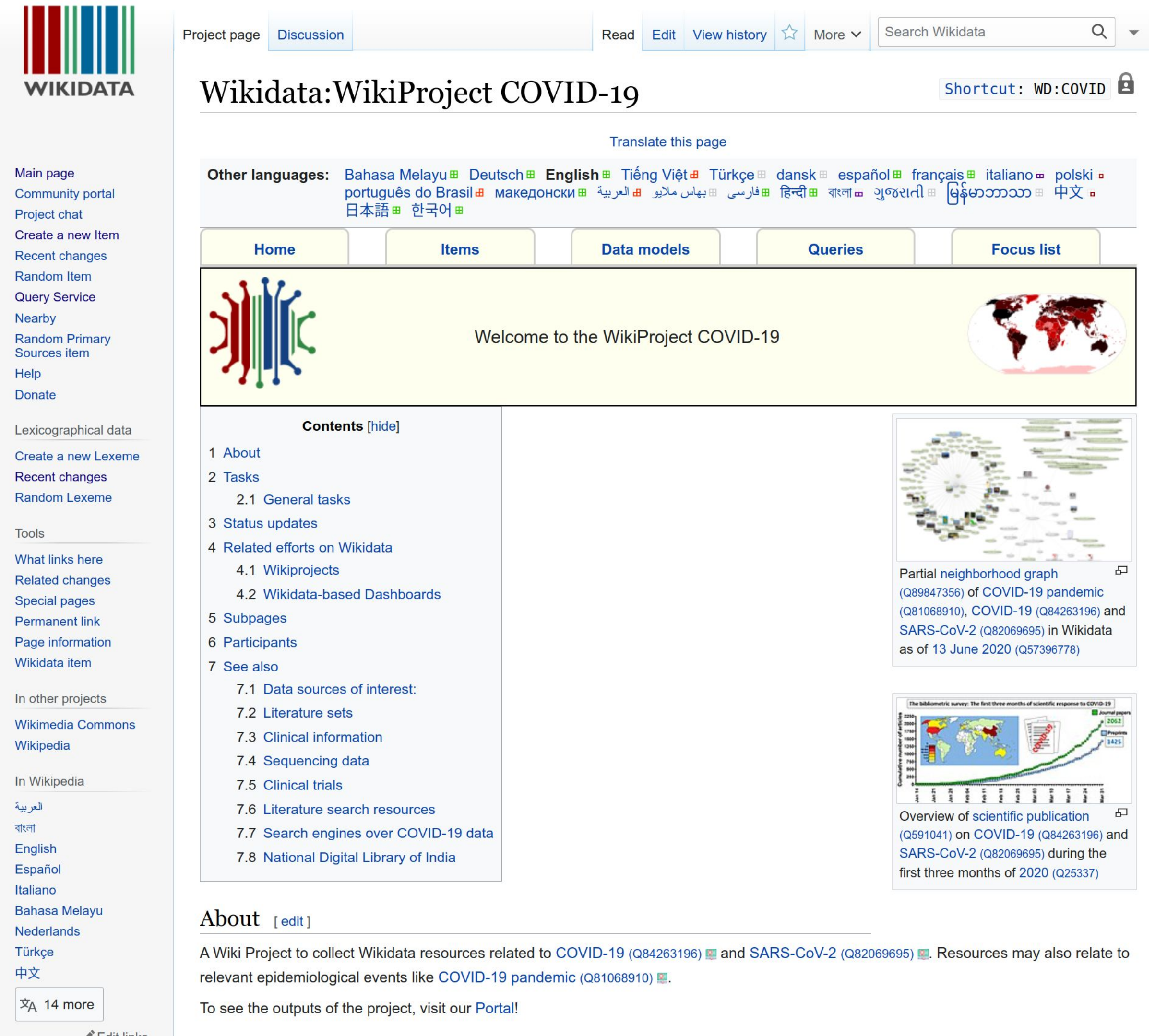
The Complex Portal (www.ebi.ac.uk/complexportal) is a manually curated, encyclopaedic database of macromolecular complexes with known function from a range of model organisms. It summarizes complex composition, topology and function along with links to a large range of domain-specific resources (i.e. wwPDB, EMDB and Reactome). Since the last update in 2019, we have produced a first draft complexome for *Escherichia coli*, maintained and updated that of *Saccharomyces cerevisiae*, added over 40 coronavirus complexes and increased the human complexome to over 1100 complexes that include approximately 200 complexes that act as targets for viral proteins or are part of the immune system. The display of protein features in ComplexViewer has been improved and the participant table is now colour-coordinated with the nodes in ComplexViewer. Community collaboration has expanded, for example by contributing to an analysis of putative transcription cofactors and providing data accessible to semantic web tools through Wikidata which is now populated with manually curated Complex Portal content through a new bot. Our data license is now CC0 to encourage data reuse. Users are encouraged to get in touch, provide us with feedback and send curation requests through the ‘Support’ link.

Issue Section: Database Issue

INTRODUCTION

Protein complexes, stable functional assemblies consisting of two or more associated polypeptide chains, are responsible for driving and regulating many

SPARQL queries



Wikidata:WikiProject COVID-19

Shortcut: WD:COVID

Translate this page

Other languages: Bahasa Melayu Deutsch **English** Tiếng Việt Türkçe dansk español français italiano polski português do Brasil македонски العربية بهاس ملايو فارسی हिन्दी বাংলা ગુજરાતી မြန်မာဘာသာ 中文 日本語 한국어

Welcome to the WikiProject COVID-19

Contents [hide]

Partial neighborhood graph (Q89847356) of COVID-19 pandemic (Q81068910), COVID-19 (Q84263196) and SARS-CoV-2 (Q82069695) in Wikidata as of 13 June 2020 (Q57396778)

Overview of scientific publication (Q591041) on COVID-19 (Q84263196) and SARS-CoV-2 (Q82069695) during the first three months of 2020 (Q25337)

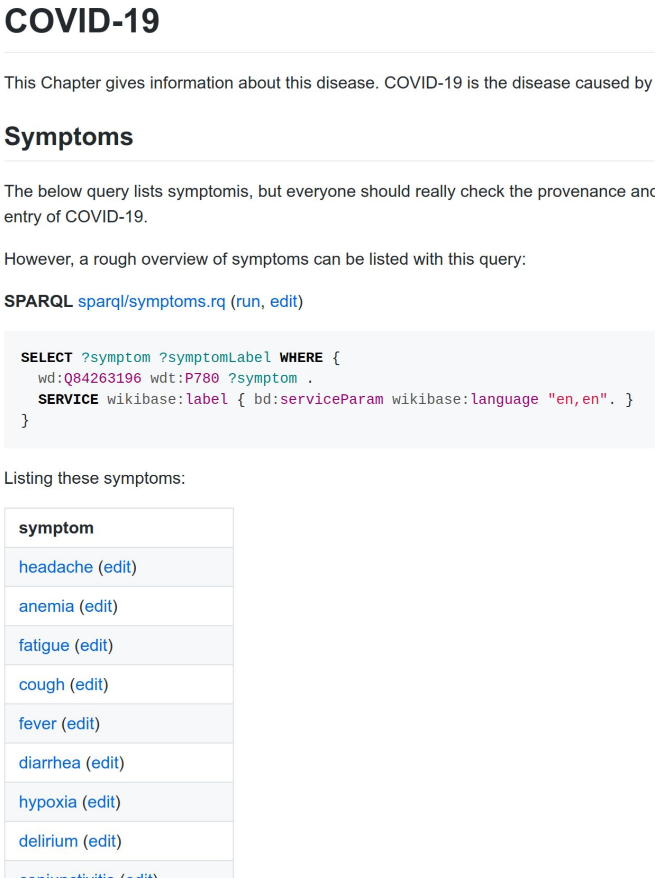
About [edit]

A Wiki Project to collect Wikidata resources related to COVID-19 (Q84263196) and SARS-CoV-2 (Q82069695). Resources may also relate to relevant epidemiological events like COVID-19 pandemic (Q81068910).

To see the outputs of the project, visit our Portal!

Maastricht University

SPARQL queries



COVID-19

This Chapter gives information about this disease. COVID-19 is the disease caused by

Symptoms

The below query lists symptomis, but everyone should really check the provenance anc entry of COVID-19.

However, a rough overview of symptoms can be listed with this query:

SPARQL sparql/symptoms.rq (run, edit)

```
SELECT ?symptom ?symptomLabel WHERE {
  wd:Q84263196 wdt:P780 ?symptom .
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en,en". }
}
```

Listing these symptoms:

symptom
headache (edit)
anemia (edit)
fatigue (edit)
cough (edit)
fever (edit)
diarrhea (edit)
hypoxia (edit)
delirium (edit)

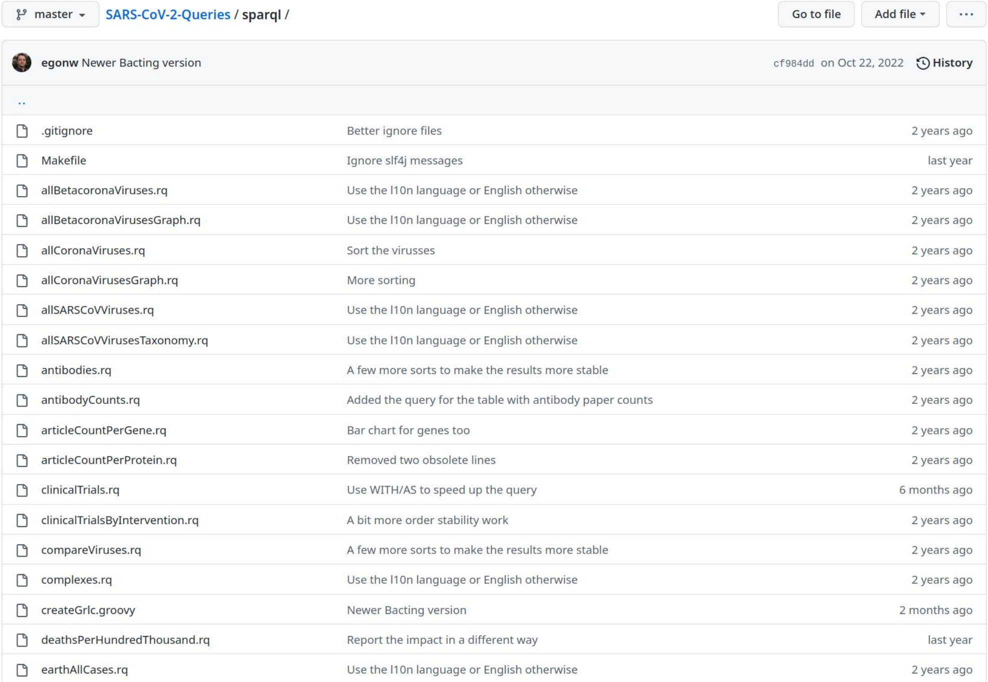
Code examples

curl

```
curl -s https://raw.githubusercontent.com/egonw/SARS-CoV-2-Queries/master/sparql/symptoms.rq | sed 's+<lang/>+en+' > s

curl -H "Accept: text/tab-separated-values" -G https://query.wikidata.org/bigdata/namespace/wdq/sparql --data-urlencod
```

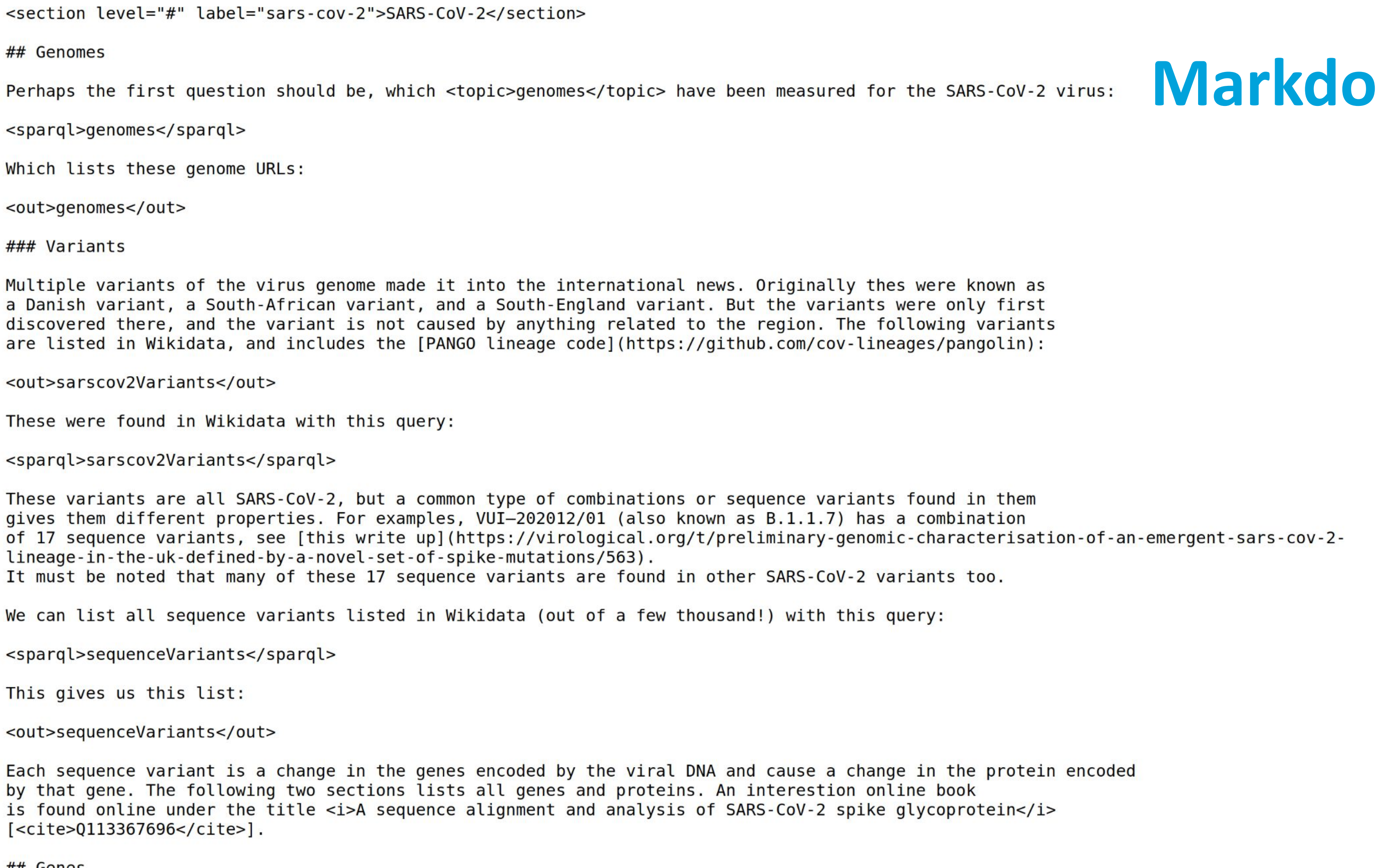
This SPARQL query is available under CCZero.



egonw Newer Bacting version — cf984dd on Oct 22, 2022

..		
.gitignore	Better ignore files	2 years ago
Makefile	Ignore slf4j messages	last year
allBetacoronaViruses.rq	Use the l10n language or English otherwise	2 years ago
allBetacoronaVirusesGraph.rq	Use the l10n language or English otherwise	2 years ago
allCoronaViruses.rq	Sort the virusses	2 years ago
allCoronaVirusesGraph.rq	More sorting	2 years ago
allSARSCoVViruses.rq	Use the l10n language or English otherwise	2 years ago
allSARSCoVVirusesTaxonomy.rq	Use the l10n language or English otherwise	2 years ago
antibodies.rq	A few more sorts to make the results more stable	2 years ago
antibodyCounts.rq	Added the query for the table with antibody paper counts	2 years ago
articleCountPerGene.rq	Bar chart for genes too	2 years ago
articleCountPerProtein.rq	Removed two obsolete lines	2 years ago
clinicalTrials.rq	Use WITH/AS to speed up the query	6 months ago
clinicalTrialsByIntervention.rq	A bit more order stability work	2 years ago
compareViruses.rq	A few more sorts to make the results more stable	2 years ago
complexes.rq	Use the l10n language or English otherwise	2 years ago
createGrlc.groovy	Newer Bacting version	2 months ago
deathsPerHundredThousand.rq	Report the impact in a different way	last year
earthAllCases.rq	Use the l10n language or English otherwise	2 years ago

Markdown



```
<section level="#" label="sars-cov-2">SARS-CoV-2</section>

## Genomes

Perhaps the first question should be, which <topic>genomes</topic> have been measured for the SARS-CoV-2 virus:

<sparql>genomes</sparql>

Which lists these genome URLs:

<out>genomes</out>

### Variants

Multiple variants of the virus genome made it into the international news. Originally thes were known as
a Danish variant, a South-African variant, and a South-England variant. But the variants were only first
discovered there, and the variant is not caused by anything related to the region. The following variants
are listed in Wikidata, and includes the [PANGO lineage code](https://github.com/cov-lineages/pangolin):

<out>sarscov2Variants</out>

These were found in Wikidata with this query:

<sparql>sarscov2Variants</sparql>

These variants are all SARS-CoV-2, but a common type of combinations or sequence variants found in them
gives them different properties. For examples, VUI–202012/01 (also known as B.1.1.7) has a combination
of 17 sequence variants, see [this write up](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-
lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563).
It must be noted that many of these 17 sequence variants are found in other SARS-CoV-2 variants too.

We can list all sequence variants listed in Wikidata (out of a few thousand!) with this query:

<sparql>sequenceVariants</sparql>

This gives us this list:

<out>sequenceVariants</out>

Each sequence variant is a change in the genes encoded by the viral DNA and cause a change in the protein encoded
by that gene. The following two sections lists all genes and proteins. An interestion online book
is found online under the title <i>A sequence alignment and analysis of SARS-CoV-2 spike glycoprotein</i>
[<cite>Q113367696</cite>].

## Genes
```

Concept: Maintainable book

Groovy Cheminformatics...

Update: the fourth edition is out.

Some project are never finished. Neither is this one, but it is never too late to change how things work, so, taking advantage of publishing-on-demand, here I introduce the release-soon, release-often equivalent of cheminformatics books, my Groovy Cheminformatics with the Chemistry Development Kit book:

Groovy Cheminformatics with the Chemistry Development Kit

By Egon Willighagen

Paperback, 72 pages ☆☆☆☆☆ This item has not been rated yet

List Price: ~~$50.13~~
Price: **$30.08**
You Save: $20.05 (40%)
Ships in 3–5 business days

This book introduces the reader to the wide variety of functionality available in the Chemistry Development Kit (CDK) library. It will discuss parts of the data model, basic cheminformatics algorithms, chemical file formats, etc. It discusses bits of chemical graph theory, computer representation, etc. But the goal of this book is not to provide an introduction into cheminformatics. However, this book does require a basic chemical education. It assumes that you know what atoms are, how they are connected by chemical bonds, and it assumes some basic computer knowledge. This book is about learning how to perform cheminformatics tasks using the CDK. But to keep the required knowledge to a minimum, the examples will be pretty verbose. The book details the CDK data model, discusses input and output, atom types, graph properties, missing information, and substructure searching. It demonstrates the CDK library with 75 code examples.

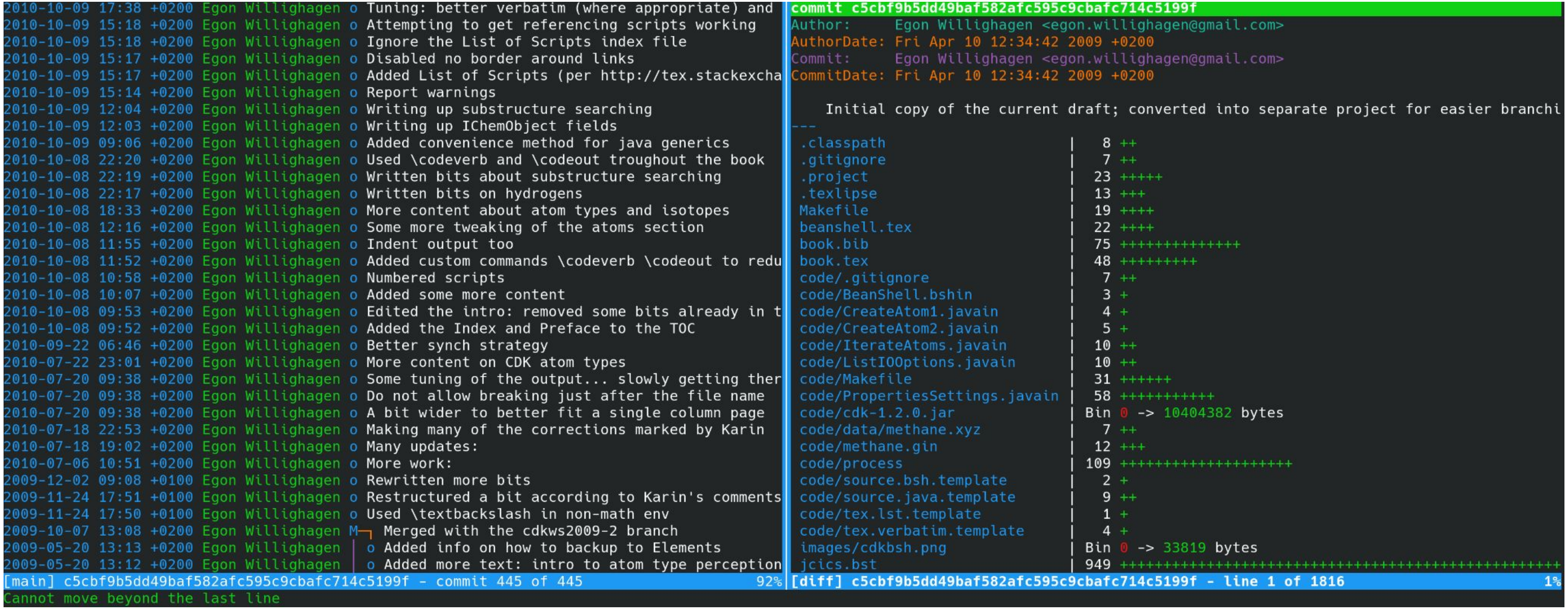
With a serious discount for just being the first edition (1.3.8-0), but still counting at 72 pages with 75 code examples, this edition marks a personal milestone (and probably not much more than that). There remains much to do, but I promised a release by tomorrow, so here it is. Next releases will contain more code examples, more functionality descriptions, and more literature reviewing where such code is used in science. The plan is to make new editions with each new CDK release, as well as new editions when I added a new chapter, section, or just paragraph. But, there will not be a Nightly build service anytime soon.

The current table of content is as follows:

Maastricht University

https://chem-bla-ics.blogspot.com/2011/02/groovy-cheminformatics.html?q=groovy+cheminformatics

Concept:
Maintainable book



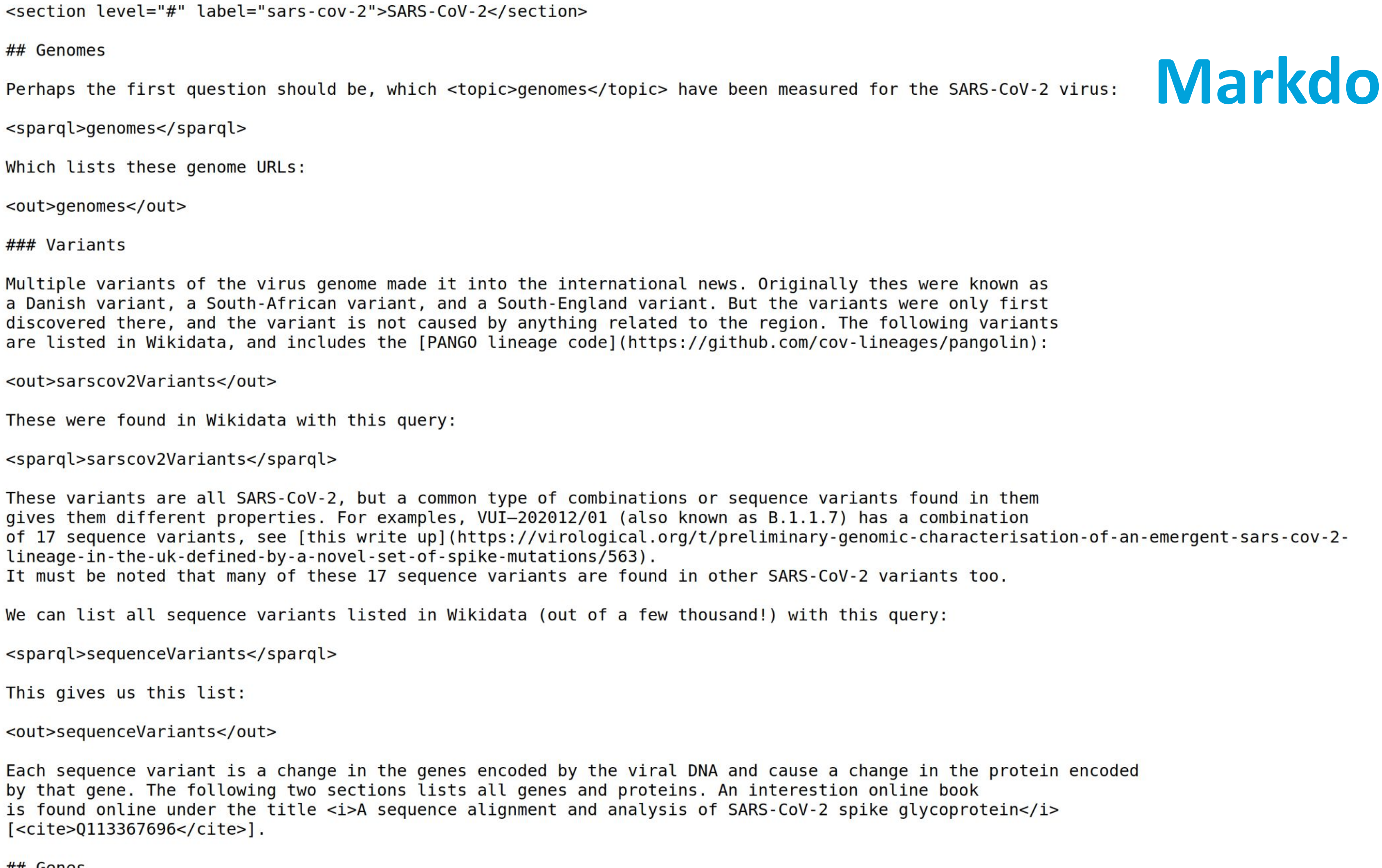
```
2010-10-09 17:38 +0200 Egon Willighagen o Tuning: better verbatim (where appropriate) and
2010-10-09 15:18 +0200 Egon Willighagen o Attempting to get referencing scripts working
2010-10-09 15:18 +0200 Egon Willighagen o Ignore the List of Scripts index file
2010-10-09 15:17 +0200 Egon Willighagen o Disabled no border around links
2010-10-09 15:17 +0200 Egon Willighagen o Added List of Scripts (per http://tex.stackexcha
2010-10-09 15:14 +0200 Egon Willighagen o Report warnings
2010-10-09 12:04 +0200 Egon Willighagen o Writing up substructure searching
2010-10-09 12:03 +0200 Egon Willighagen o Writing up IChemObject fields
2010-10-09 09:06 +0200 Egon Willighagen o Added convenience method for java generics
2010-10-08 22:20 +0200 Egon Willighagen o Used \codeverb and \codeout troughout the book
2010-10-08 22:19 +0200 Egon Willighagen o Written bits about substructure searching
2010-10-08 22:17 +0200 Egon Willighagen o Written bits on hydrogens
2010-10-08 18:33 +0200 Egon Willighagen o More content about atom types and isotopes
2010-10-08 12:16 +0200 Egon Willighagen o Some more tweaking of the atoms section
2010-10-08 11:55 +0200 Egon Willighagen o Indent output too
2010-10-08 11:52 +0200 Egon Willighagen o Added custom commands \codeverb \codeout to redu
2010-10-08 10:58 +0200 Egon Willighagen o Numbered scripts
2010-10-08 10:07 +0200 Egon Willighagen o Added some more content
2010-10-08 09:53 +0200 Egon Willighagen o Edited the intro: removed some bits already in t
2010-10-08 09:52 +0200 Egon Willighagen o Added the Index and Preface to the TOC
2010-09-22 06:46 +0200 Egon Willighagen o Better synch strategy
2010-07-22 23:01 +0200 Egon Willighagen o More content on CDK atom types
2010-07-20 09:38 +0200 Egon Willighagen o Some tuning of the output... slowly getting ther
2010-07-20 09:38 +0200 Egon Willighagen o Do not allow breaking just after the file name
2010-07-20 09:38 +0200 Egon Willighagen o A bit wider to better fit a single column page
2010-07-18 22:53 +0200 Egon Willighagen o Making many of the corrections marked by Karin
2010-07-18 19:02 +0200 Egon Willighagen o Many updates:
2010-07-06 10:51 +0200 Egon Willighagen o More work:
2009-12-02 09:08 +0100 Egon Willighagen o Rewritten more bits
2009-11-24 17:51 +0100 Egon Willighagen o Restructured a bit according to Karin's comments
2009-11-24 17:50 +0100 Egon Willighagen o Used \textbackslash in non-math env
2009-10-07 13:08 +0200 Egon Willighagen M─┐ Merged with the cdkws2009-2 branch
2009-05-20 13:13 +0200 Egon Willighagen │ o Added info on how to backup to Elements
2009-05-20 13:12 +0200 Egon Willighagen │ o Added more text: intro to atom type perception
[main] c5cbf9b5dd49baf582afc595c9cbafc714c5199f - commit 445 of 445                    92%
Cannot move beyond the last line
```

```
commit c5cbf9b5dd49baf582afc595c9cbafc714c5199f
Author:     Egon Willighagen <egon.willighagen@gmail.com>
AuthorDate: Fri Apr 10 12:34:42 2009 +0200
Commit:     Egon Willighagen <egon.willighagen@gmail.com>
CommitDate: Fri Apr 10 12:34:42 2009 +0200

    Initial copy of the current draft; converted into separate project for easier branchi
---
 .classpath                      |   8 ++
 .gitignore                      |   7 ++
 .project                        |  23 +++++
 .texlipse                       |  13 +++
 Makefile                        |  19 ++++
 beanshell.tex                   |  22 ++++
 book.bib                        |  75 ++++++++++++++
 book.tex                        |  48 +++++++++
 code/.gitignore                 |   7 ++
 code/BeanShell.bshin            |   3 +
 code/CreateAtom1.javain         |   4 +
 code/CreateAtom2.javain         |   5 +
 code/IterateAtoms.javain        |  10 ++
 code/ListIOOptions.javain       |  10 ++
 code/Makefile                   |  31 ++++++
 code/PropertiesSettings.javain  |  58 +++++++++++
 code/cdk-1.2.0.jar              | Bin 0 -> 10404382 bytes
 code/data/methane.xyz           |   7 ++
 code/methane.gin                |  12 +++
 code/process                    | 109 ++++++++++++++++++++
 code/source.bsh.template        |   2 +
 code/source.java.template       |   9 ++
 code/tex.lst.template           |   1 +
 code/tex.verbatim.template      |   4 +
 images/cdkbsh.png               | Bin 0 -> 33819 bytes
 jcics.bst                       | 949 +++++++++++++++++++++++++++++++++++++++++++++++++++++
[diff] c5cbf9b5dd49baf582afc595c9cbafc714c5199f - line 1 of 1816                      1%
```

https://chem-bla-ics.blogspot.com/2011/02/groovy-cheminformatics.html?q=groovy+cheminformatics

Markdown



```
<section level="#" label="sars-cov-2">SARS-CoV-2</section>

## Genomes

Perhaps the first question should be, which <topic>genomes</topic> have been measured for the SARS-CoV-2 virus:

<sparql>genomes</sparql>

Which lists these genome URLs:

<out>genomes</out>

### Variants

Multiple variants of the virus genome made it into the international news. Originally thes were known as
a Danish variant, a South-African variant, and a South-England variant. But the variants were only first
discovered there, and the variant is not caused by anything related to the region. The following variants
are listed in Wikidata, and includes the [PANGO lineage code](https://github.com/cov-lineages/pangolin):

<out>sarscov2Variants</out>

These were found in Wikidata with this query:

<sparql>sarscov2Variants</sparql>

These variants are all SARS-CoV-2, but a common type of combinations or sequence variants found in them
gives them different properties. For examples, VUI–202012/01 (also known as B.1.1.7) has a combination
of 17 sequence variants, see [this write up](https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-
lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563).
It must be noted that many of these 17 sequence variants are found in other SARS-CoV-2 variants too.

We can list all sequence variants listed in Wikidata (out of a few thousand!) with this query:

<sparql>sequenceVariants</sparql>

This gives us this list:

<out>sequenceVariants</out>

Each sequence variant is a change in the genes encoded by the viral DNA and cause a change in the protein encoded
by that gene. The following two sections lists all genes and proteins. An interestion online book
is found online under the title <i>A sequence alignment and analysis of SARS-CoV-2 spike glycoprotein</i>
[<cite>Q113367696</cite>].

## Genes
```

Citations

18 lines (16 sloc) | 629 Bytes

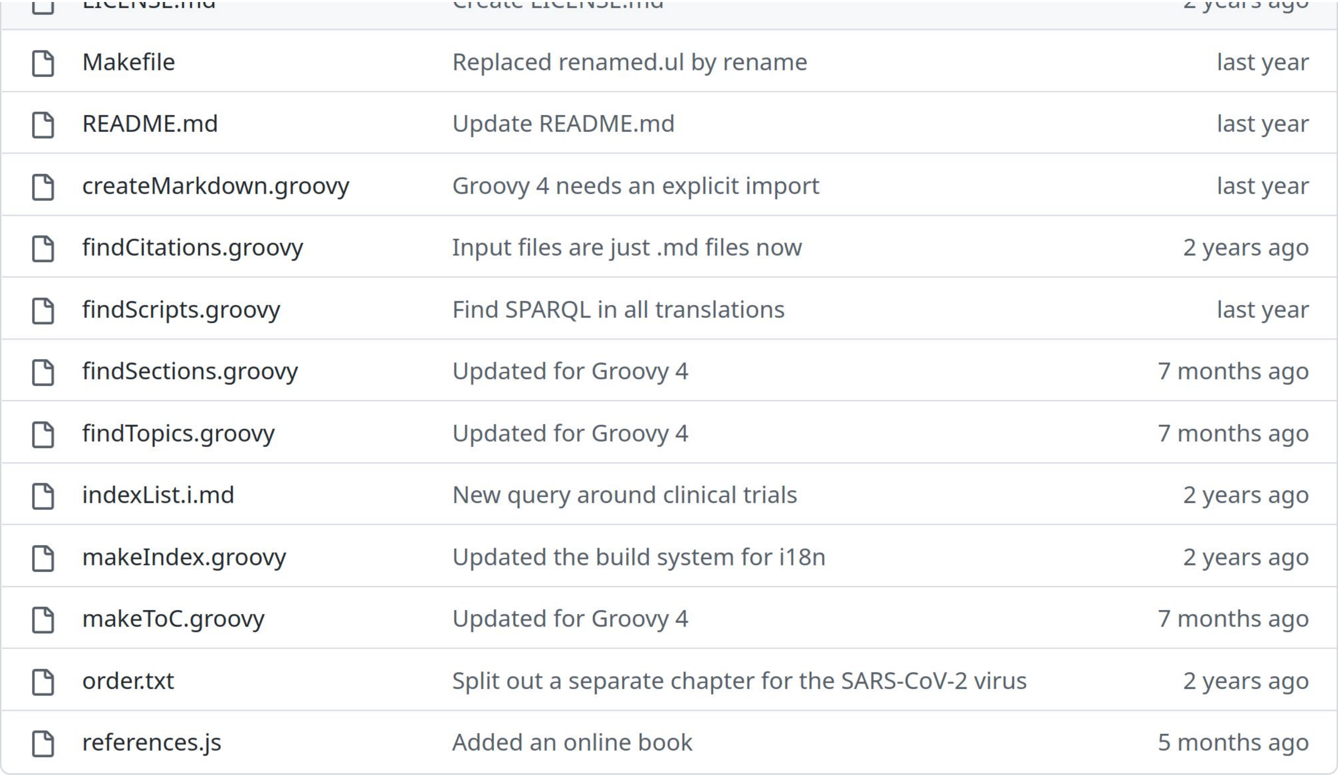
```
const fs = require('fs').promises;

const Cite = require('citation-js');

(async function () {
  const file = await fs.readFile('references.qids', 'utf8')
  const data = Array.from(await Cite.async(file))
    .map(item => item.id + '=' + Cite(item).format('bibliography', {
      template: 'vancouver',
      append: function (entry) {
        return entry.DOI
          ? ' doi:[' + entry.DOI + '](https://doi.org/' + entry.DOI + ')' +
            ' ([Scholia](https://scholia.toolforge.org/doi/' + entry.DOI + "))"
          : '';
      }
    }))
  await fs.writeFile('references.dat', data.join(''))
})().catch(console.error)
```

Willighagen, L. G. Citation.js: a format-independent, modular bibliography tool for the browser and command line. PeerJ Computer Science 5, e214 (2019).

Pre-Markdown → Markdown → GitHub Pages



Makefile	Replaced renamed.ul by rename	last year
README.md	Update README.md	last year
createMarkdown.groovy	Groovy 4 needs an explicit import	last year
findCitations.groovy	Input files are just .md files now	2 years ago
findScripts.groovy	Find SPARQL in all translations	last year
findSections.groovy	Updated for Groovy 4	7 months ago
findTopics.groovy	Updated for Groovy 4	7 months ago
indexList.i.md	New query around clinical trials	2 years ago
makeIndex.groovy	Updated the build system for i18n	2 years ago
makeToC.groovy	Updated for Groovy 4	7 months ago
order.txt	Split out a separate chapter for the SARS-CoV-2 virus	2 years ago
references.js	Added an online book	5 months ago

Internationalization (i18n) / Localization (l10n)

しかし、SARS-CoV-2はヒトから放出されたのちに、様々なものの表面などでしばらくの間生き続けることが判明しています[要引用]。次のクエリを発行して、平らな表面でヒトコロナウイルスが生存することについて調査している論文を取得します。

SPARQL sparql/surfacesCounts.rq (実行, 編集)

```
SELECT ?virus ?virusLabel ?count WITH {
  SELECT ?virus (COUNT(DISTINCT ?work) AS ?count) WHERE {
    VALUES ?virus {
      wd:Q82069695 # SARS-CoV-2
      wd:Q16983360 # HKU1
      wd:Q16991954 # OC43
      wd:Q8351095  # NL63
      wd:Q16983356 # 229E
      wd:Q4902157  # MERS-CoV
      wd:Q278567   # SARS-CoV
    }
    ?work wdt:P921 ?virus ;
          wdt:P921 wd:Q484298 .
  } GROUP BY ?virus
} AS %ARTICLES WHERE {
  INCLUDE %ARTICLES
  SERVICE wikibase:label { bd:serviceParam wikibase:language "ja,en". }
}
ORDER BY DESC(?count)
```

結果は以下の通りになります。

virus	count
ヒトコロナウイルス229E (edit)	4
SARSコロナウイルス2 (edit)	3
ベータコロナウイルス１ (edit)	2
MERSコロナウイルス (edit)	1

物質表面上でのウイルス生存に関する論文を実際に取得するには以下のクエリを発行します。

SPARQL sparql/surfaces.rq (実行, 編集)

```
SELECT ?virus ?virusLabel ?work ?workLabel WITH {
```

Acknowledgments

- Wikidata:WikiProject COVID-19
- WikiPathways community
- Contributors to the book: Addshore, Daniel Mietchen
- Translators: Yayamamo, Marvin Martens, Sharon and Carolina Prado, Tiago Lubiana, Nuno Nunes
- Birgit Meldal, Andra Waagmeester, Egon L. Willighagen, Andrew I. Su, Martina Kutmon, Jose Emilio Labra Gayo, Daniel Fernández-Álvarez, Quentin Groom, Peter J. Schaap, Lisa M. Verhagen & Jasper J. Koehorst
- https://citation.js.org/

Maastricht University

ALFRED P. SLOAN
FOUNDATION

ZonMw

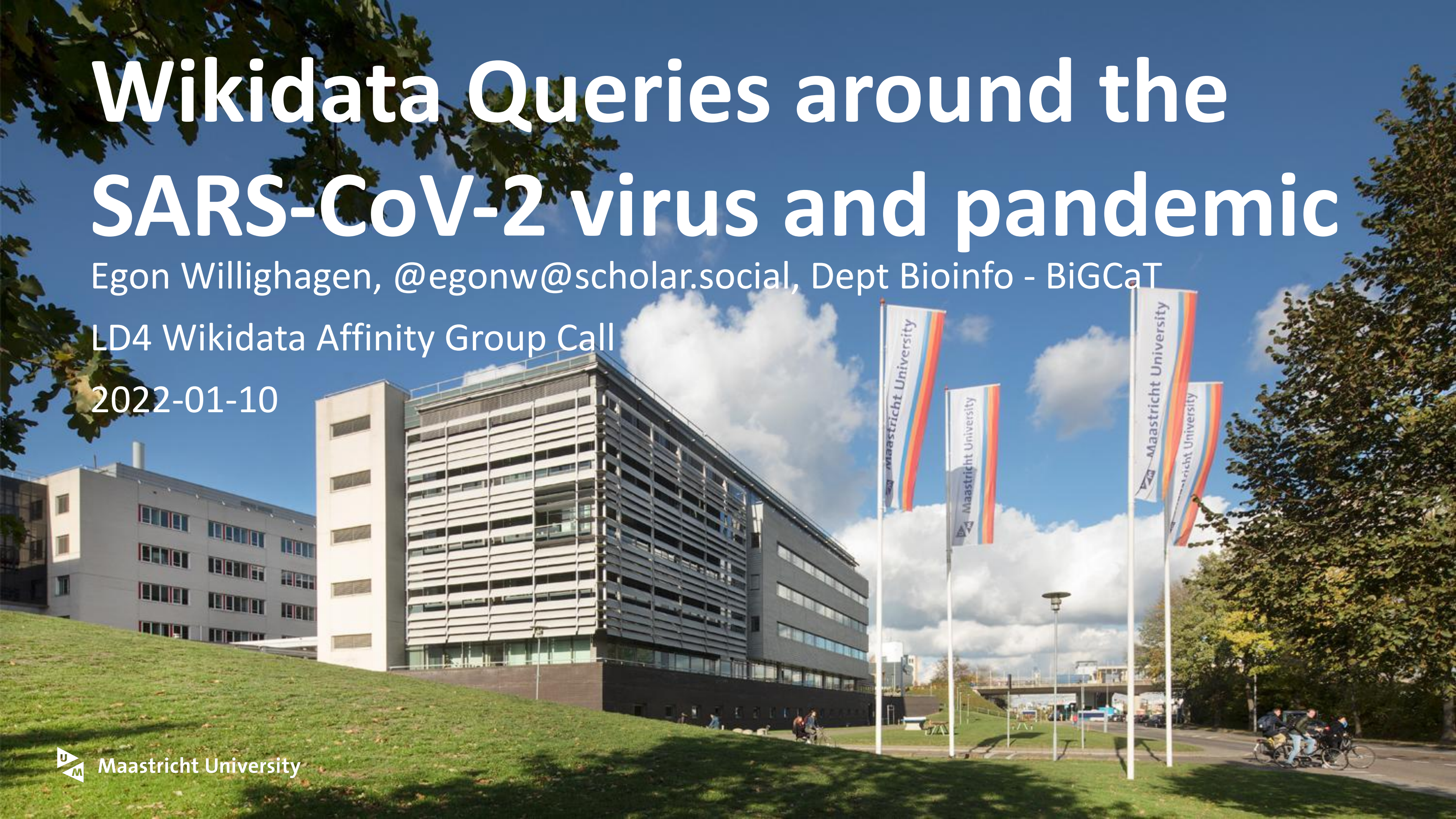
NWO

Wikidata Queries around the SARS-CoV-2 virus and pandemic

Egon Willighagen, @egonw@scholar.social, Dept Bioinfo - BiGCaT

LD4 Wikidata Affinity Group Call

2022-01-10



Maastricht University

Further reading (from the QA)

- *WikiPathways: connecting communities*, 10.1093/nar/gkaa1024, covid.wikipathways.org
- EBI COVID-19 Data Portal, www.covid19dataportal.org
- COVID-19 Disease Maps project, 10.15252/msb.202110387
- Research from my colleagues [preprint]: *Different B cell activation patterns in asymptomatic and symptomatic COVID-19 patients* 10.1101/2022.12.19.521064