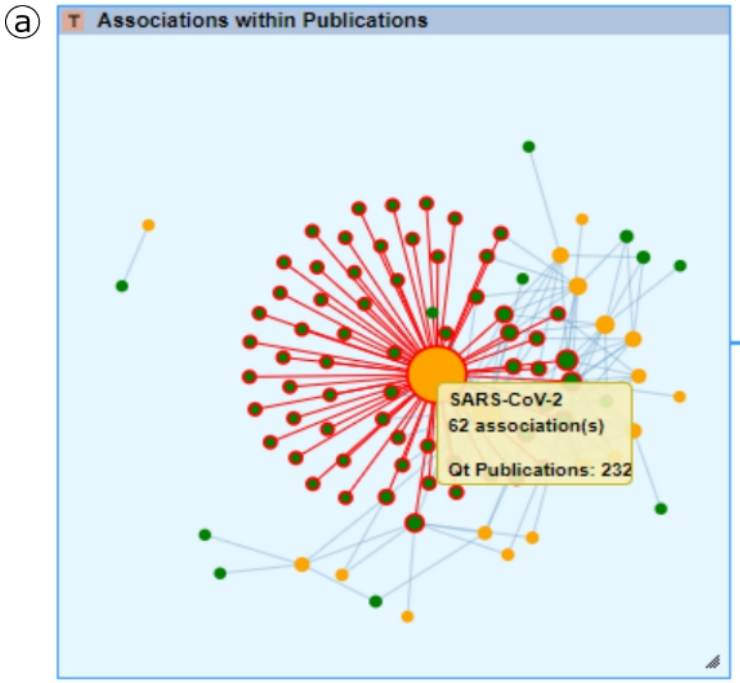
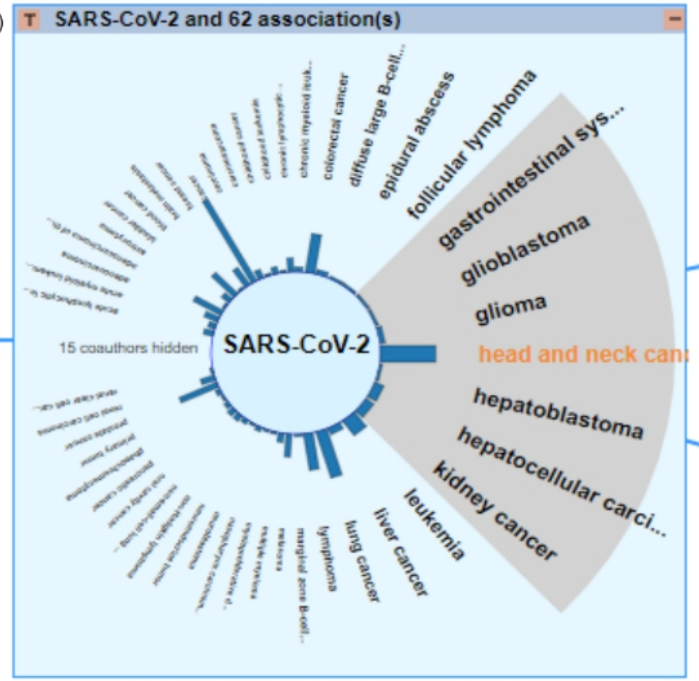
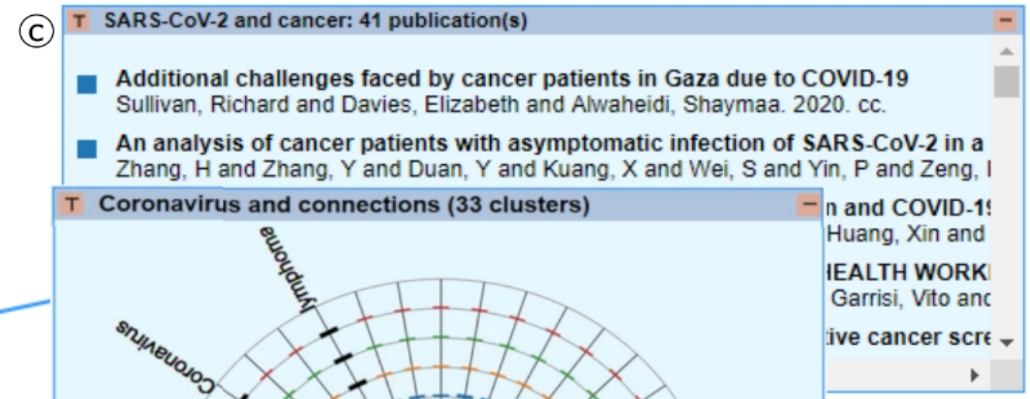
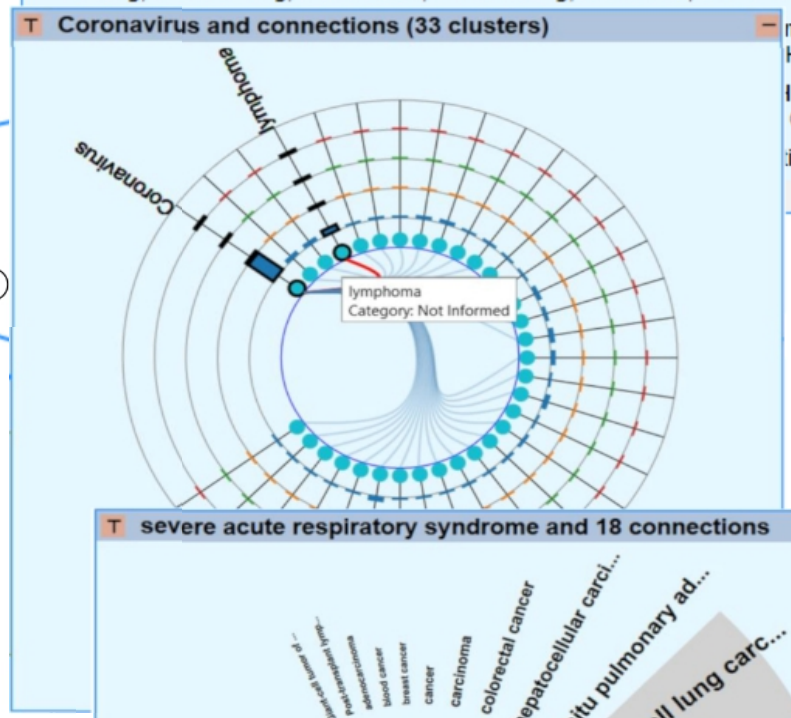
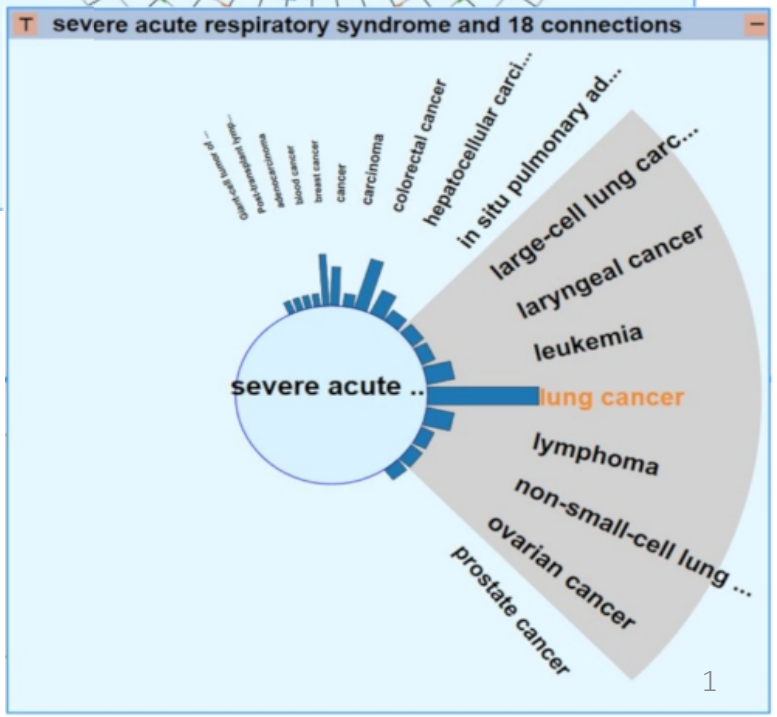










CovidOnTheWeb

[F. Michel, F. Gandon, V. Ah-Kane, A. Bobasheva, E. Cabrio, O. Corby, R. Gazzotti, A. Giboin, S. Marro, T. Mayer, M. Simon, S. Villata, M. Wincker]

~15 persons from the Wimmics Team

vs. use cases…

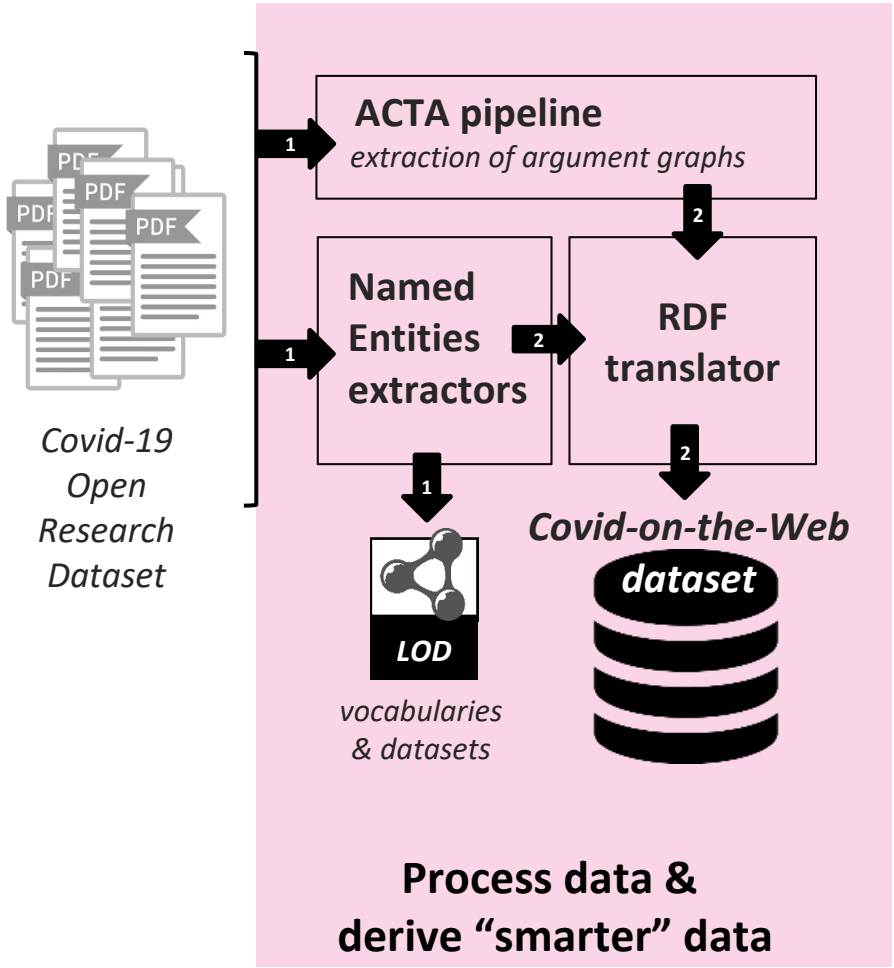
Scenario 1: Help clinicians analyze clinical trials and take evidence-based decisions

Scenario 2: Help hospital physicians to collect ranges of human organism's substances (e.g., cholesterol) from scientific articles.

Scenario 3: Help missions heads from Cancer Institute elaborate research programs to study the links between cancer and coronavirus

[Giboin, et al.]

COVID ON THE WEB [ISWC 2020, IC 2021]



[Michel, Gazzotti, Gandon, Cabrio, Villata, Mayer et al.]

URIs for… things mentioned in papers [Gazzotti, et al.]

AI methods: NLP and IR for *named entity* annotation in text

- DBpedia Spotlight → DBpedia URIs
- Entity-fishing → Wikidata URIs
- BioPortal Annotator → BioPortal URIs

URIs for… things mentioned in papers [Gazzotti, et al.]

AI methods: NLP and IR for *named entity* annotation in text

- DBpedia Spotlight → DBpedia URIs
- Entity-fishing → Wikidata URIs
- BioPortal Annotator → BioPortal URIs

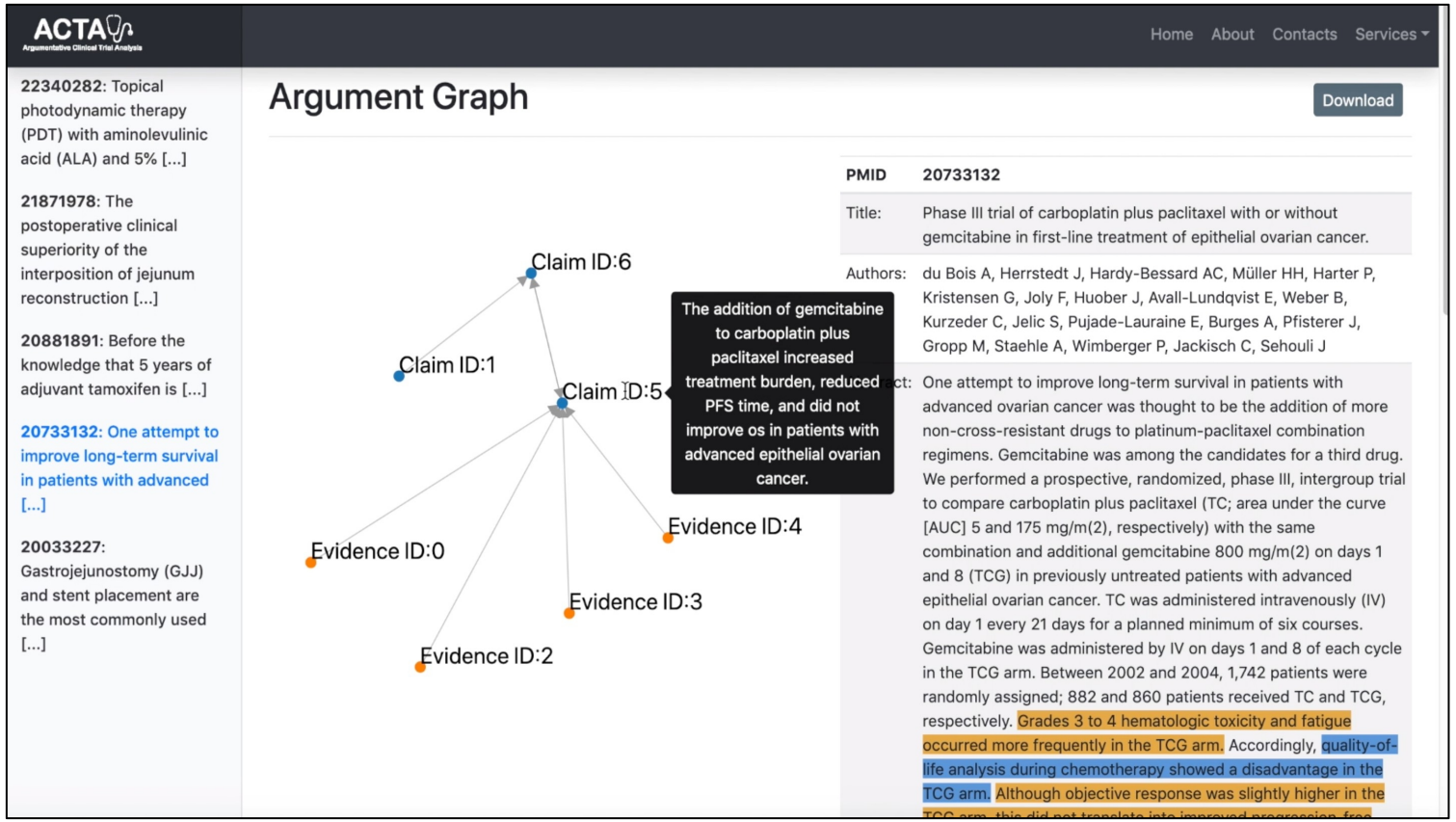
"Effects on QT interval of hydroxychloroquine associated with ritonavir/darunavir or azithromycin in patients with SARS-CoV-2 infection" [Danzi et al.]

http://dbpedia.org/resource/**Severe_acute_respiratory_syndrome_coronavirus_2**
http://dbpedia.org/resource/**Hydroxychloroquine**
http://dbpedia.org/resource/**Azithromycin**
http://dbpedia.org/resource/**Ritonavir**
http://dbpedia.org/resource/**Darunavir**
http://dbpedia.org/resource/**Heart_arrhythmia**
⋮

[Mayer, Cabrio, Villata, Marro, et al.]

Extracting arguments and PICO elements [ECAI 2020]

AI methods: BERT+SciBERT, LSTM, Conditional Random Field



PICO Information

PICO Type	Content
intervention	platinum - paclitaxel combination
intervention	Gemcitabine
intervention	carboplatin plus paclitaxel (TC
intervention	gemcitabine
intervention	TC
intervention	Gemcitabine
intervention	TC
intervention	TCG
outcome	Grades 3 to 4 hematologic toxicity and fatigue
outcome	quality - of - life analysis
outcome	objective response
outcome	progression - free survival (PFS) or overall survival (OS) .
outcome	Median PFS
outcome	Median OS
intervention	gemcitabine
intervention	carboplatin plus paclitaxel
outcome	treatment burden
outcome	PFS time
outcome	OS
intervention	TCG

Evidence-Based Practice (EBP)
- **P**atient Problem, (or Population)
- **I**ntervention,
- **C**omparison or Control, and
- **O**utcome

Integrate in RDF

morph-xr2rml, MongoDB [Michel, Gazzotti et al.]

- Web Annotation Vocabulary

named entities

```
_:b40150806
    a                   oa:Annotation, prov:Entity;
    schema:about        <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>;
    dct:subject         "Engineering", "Biology";

    covidpr:confidence  "1"^^xsd:decimal;
    oa:hasBody          <http://wikidata.org/entity/Q176996>;
    oa:hasTarget [
        oa:hasSource    <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f#abstract>;
        oa:hasSelector  [
            a           oa:TextPositionSelector, oa:TextQuoteSelector;
            oa:exact    "PCR";
            oa:start    "235";
            oa:end      "238"
        ]
    ];
```

Integrate in RDF

morph-xr2rml, MongoDB [Michel, Gazzotti et al.]

named entities

```
_:b40150806
    a                   oa:Annotation, prov:Entity;
    schema:about        <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>;
    dct:subject         "Engineering", "Biology";
```

metadata

```
<http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>
    a                   fabio:ResearchPaper, bibo:AcademicArticle, schema:ScholarlyArticle;
    rdfs:isDefinedBy    <http://ns.inria.fr/covid19/dataset-1-1>;
    dct:title           "A real-time PCR for SARS-coronavirus incorporating target gene pre-amplification";
    schema:publication  "Biochemical and Biophysical Research Communications";
    dce:creator         "Wong, Freda Pui-Fan", "Tam, Siu-Lun", "Fung, Yin-Wan", "Li, Hui", "Cheung, Albert", "Chan, Paul", "Lin,
    dct:source          "Elsevier";
    dct:license         "els-covid";

    dct:issued          "2003-12-26"^^xsd:date;
    bibo:doi            "10.1016/j.bbrc.2003.11.064";
    bibo:pmid           "14652014";
    fabio:hasPubMedId   "14652014";
    foaf:sha1           "f74923b3ce82c984a7ae3e0c2754c9e33c60554f";
    schema:url          <https://doi.org/10.1016/j.bbrc.2003.11.064>;

    dct:abstract        <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f#abstract>;
    covidpr:hasTitle    <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f#title>;
    covidpr:hasBody     <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f#body_text>.
```

- Dublin Core vocabulary
- Bibliographic Ontology

Integrate in RDF

morph-xr2rml, MongoDB [Michel, Gazzotti et al.]

named entities

```
_:b40150806
    a                   oa:Annotation, prov:Entity;
    schema:about        <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>;
    dct:subject         "Engineering", "Biology";
```

metadata

```
<http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>
    a                   fabio:ResearchPaper, bibo:AcademicArticle, schema:ScholarlyArticle;
                                                                                    tract>;
 target gene pre-amplification";
s";
an", "Li, Hui", "Cheung, Albert", "Chan, Paul", "Lin,
```

arguments

```
<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5aed66dcc925c4b>
    a                   amo:Argument;
    schema:about        covid:f74923b3ce82c984a7ae3e0c2754c9e33c60554f;
    dct:creator         <https://team.inria.fr/wimmics/>;
    prov:wasGeneratedBy covid:ProvenanceActa.

    # Argumentative components
    amo:hasEvidence     <http://ns.inria.fr/covid19/arg/4f8
    amo:hasEvidence     <http://ns.inria.fr/covid19/arg/4f8
    amo:hasClaim        <http://ns.inria.fr/covid19/arg/4f8
   .

<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5a
    a                   amo:Evidence, sioca:Justification,
    prov:wasQuotedFrom  covid:4f8d24c531d2c334969e09e4b5aed
    aif:formDescription "17 patients discharged in recovere
    # evidence 0 supports claim 6
    sioca:supports      <http://ns.inria.fr/covid19/arg/4f8
    amo:proves          <http://ns.inria.fr/covid19/arg/4f8
    .

<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5a
    a                   amo:Evidence, sioca:Justification,
    prov:wasQuotedFrom  covid:4f8d24c531d2c334969e09e4b5aed
    aif:formDescription "some other evidence"^^xsd:string;
    # evidence 123 attacks claim 6
    sioca:challenges <http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5aed66dcc925c4b/6>.
    .

<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5aed66dcc925c4b/6>
    a                   amo:Claim, sioca:Idea, aif:I-node, aif:KnowledgePosition_Statement;
    prov:wasQuotedFrom  covid:4f8d24c531d2c334969e09e4b5aed66dcc925c4b;
    aif:claimText       "a simple ct scoring method was capable to predict mortality."^^xsd:string;
    .
```

PICO elements

```
[]  a                    oa:Annotation;
    schema:about         <http://ns.inria.fr/covid19/4f8d24c531d2c334969e09e4b5aed66dcc925c4b>;
    covidpr:confidence   1^^xsd:decimal;

    # link to the ULMS concept id (CUI) and semantic type id (TUI)
    oa:hasBody           [ umls:cui "C0026565"; umls:tui "T81" ];
    oa:hasTarget [
        # the source is the claim/evidence
        oa:hasSource     <http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5aed66dcc925c4b/6>;
        oa:hasSelector   [
            a            oa:TextQuoteSelector;
            oa:exact     "mortality";
        ]
    ].
```

- Argument Model Ontology
- SIOC Argumentation Module
- Argument Interchange Format
- Web Annotation Vocabulary (PICO elements)

Integrate in RDF

morph-xr2rml, MongoDB [Michel, Gazzotti et al.]

named entities

```
_:b40150806
    a                   oa:Annotation, prov:Entity;
    schema:about        <http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>;
    dct:subject         "Engineering", "Biology";
```

metadata

```
<http://ns.inria.fr/covid19/f74923b3ce82c984a7ae3e0c2754c9e33c60554f>
    a                   fabio:ResearchPaper, bibo:AcademicArticle, schema:ScholarlyArticle;
                                                                    tract>;
      target gene pre-amplification";
     s";
     an", "Li, Hui", "Cheung, Albert", "Chan, Paul", "Lin,
```

arguments

```
<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5aed66dcc925c4b>
    a                   amo:Argument;
    schema:about        covid:f74923b3ce82c984a7ae3e0c2754c9e33c60554f;
    dct:creator         <https://team.inria.fr/wimmics/>;
    prov:wasGeneratedBy covid:ProvenanceActa.

    # Argumentative components
    amo:hasEvidence     <http://ns.inria.fr/covid19/arg/4f8
    amo:hasEvidence     <http://ns.inria.fr/covid19/arg/4f8
    amo:hasClaim        <http://ns.inria.fr/covid19/arg/4f8
   .

<http://ns.inria.fr/covid19/arg/4f8d24c531d2c334969e09e4b5a
    a                   amo:Evidence, sioca:Justification,
    prov:wasQuotedFrom  covid:4f8d24c531d2c334969e09e4b5aed
    aif:formDescription "17 patients discharged in recovere
    # evidence 0 supports claim 6
    sioca:supports      <http://ns.inria.fr
    amo:proves          <http://ns.inria.fr
    .

<http://ns.inria.fr/covid19/arg/4f8d24c531d
    a                   amo:Evidence, sioca
    prov:wasQuotedFrom  covid:4f8d24c531d2c
    aif:formDescription "some other evidenc
    # evidence 123 attacks claim 6
    sioca:challenges <http://ns.inria.fr/co
    .

<http://ns.inria.fr/covid19/arg/4f8d24c531d
    a                   amo:Claim, sioca:Id
    prov:wasQuotedFrom  covid:4f8d24c531d2c
    aif:claimText       "a simple ct scorin
    .
```

PICO elements

```
[]  a                   oa:Annotation;
    schema:about        <http://ns.inria.fr/covid19/4f8d24c531d2c334969e09e4b5aed66dcc925c4b>;
    covidpr:confidence  1^^xsd:decimal;

    # link to the ULMS concept id (CUI) and semantic type id (TUI)
    oa:hasBody          [ umls:cui "C0026565"; umls:tui "T81" ];
    oa:hasTarget [
```

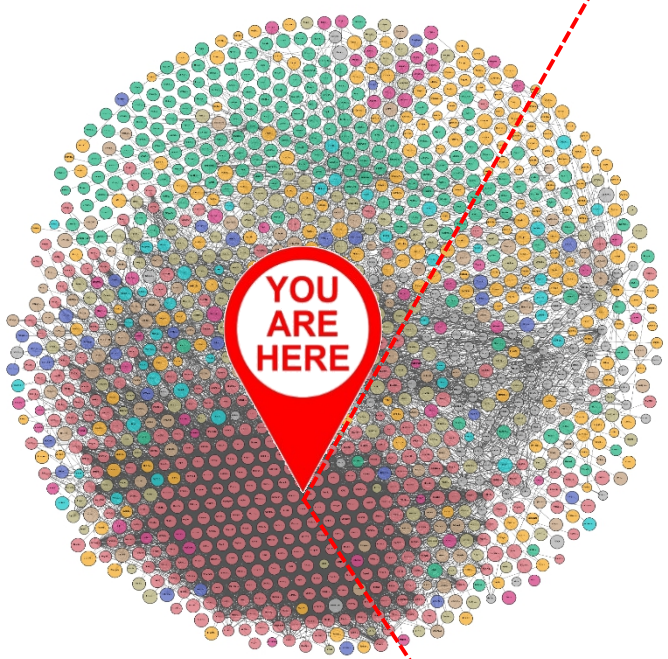
provenance

```
_:b40150806
    rdfs:isDefinedBy    <http://ns.inria.fr/covid19/dataset-1-1>;
    dct:creator         <https://team.inria.fr/wimmics/>;
    prov:wasGeneratedBy [
        a               prov:Activity;
        prov:used       <http://ns.inria.fr/covid19/cord19v7>;
        prov:wasAssociatedWith <https://github.com/kermitt2/entity-fishing>.
    ].

<http://ns.inria.fr/covid19/cord19v7>
    a                    schema:Dataset dcat:Dataset;
    owl:versionInfo      "7";
    dct:title            "COVID-19 Open Research Dataset (CORD-19)";
    dct:issued           "2020-04-10"^^xsd:date;
    schema:url           <https://www.kaggle.com/dataset/08dd9ead3afd4f61ef246bfd6aee098765a19d9f6dbf514f0142965748be859b/version/7>.
```

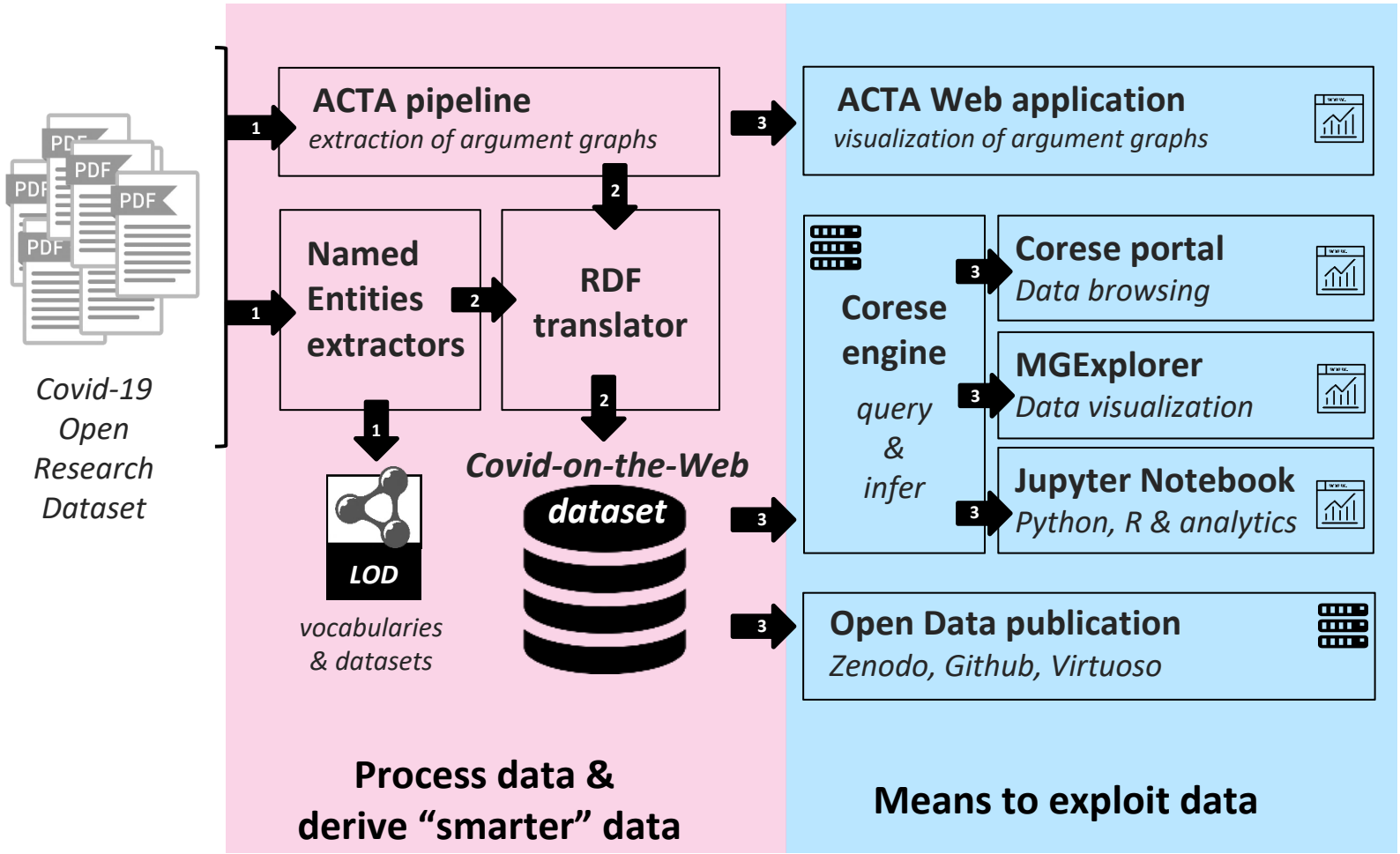
- PROV Ontology

+ other data sources



Dataset description	No. RDF triples
dataset description + definition of a few properties	170
articles metadata (title, authors, DOIs, journal etc.)	3 722 381
Named entities identified by ***Entity-fishing*** in articles titles/abstracts	35 049 832
Named entities identified by ***Entity-fishing*** in articles bodies	1 156 611 321
Named entities identified by ***Bioportal Annotator*** in articles titles/abstracts	104 430 547
Named entities identified by ***DBpedia Spotlight*** in articles titles/abstracts	65 359 664
Argumentative components and PICO elements by ***ACTA*** from articles titles/abstracts	7 469 234
Total	**1 361 451 364**

COVID ON THE WEB [ISWC 2020, IC 2021]



[Corby, Michel, Gazzotti, Gandon et al.]

Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select ?title
where {
  graph <http://ns.inria.fr/covid19/graph/articles> {
    ?paper1 a fabio:ResearchPaper; dct:title ?title.
  }

  graph <http://ns.inria.fr/covid19/graph/entityfishing> {
    ?a1 a oa:Annotation;
        schema:about ?paper1;
        oa:hasBody <http://www.wikidata.org/entity/Q12078> .
  }
} limit 100
```

Sponging: Use only local data (including data retrieved before), but

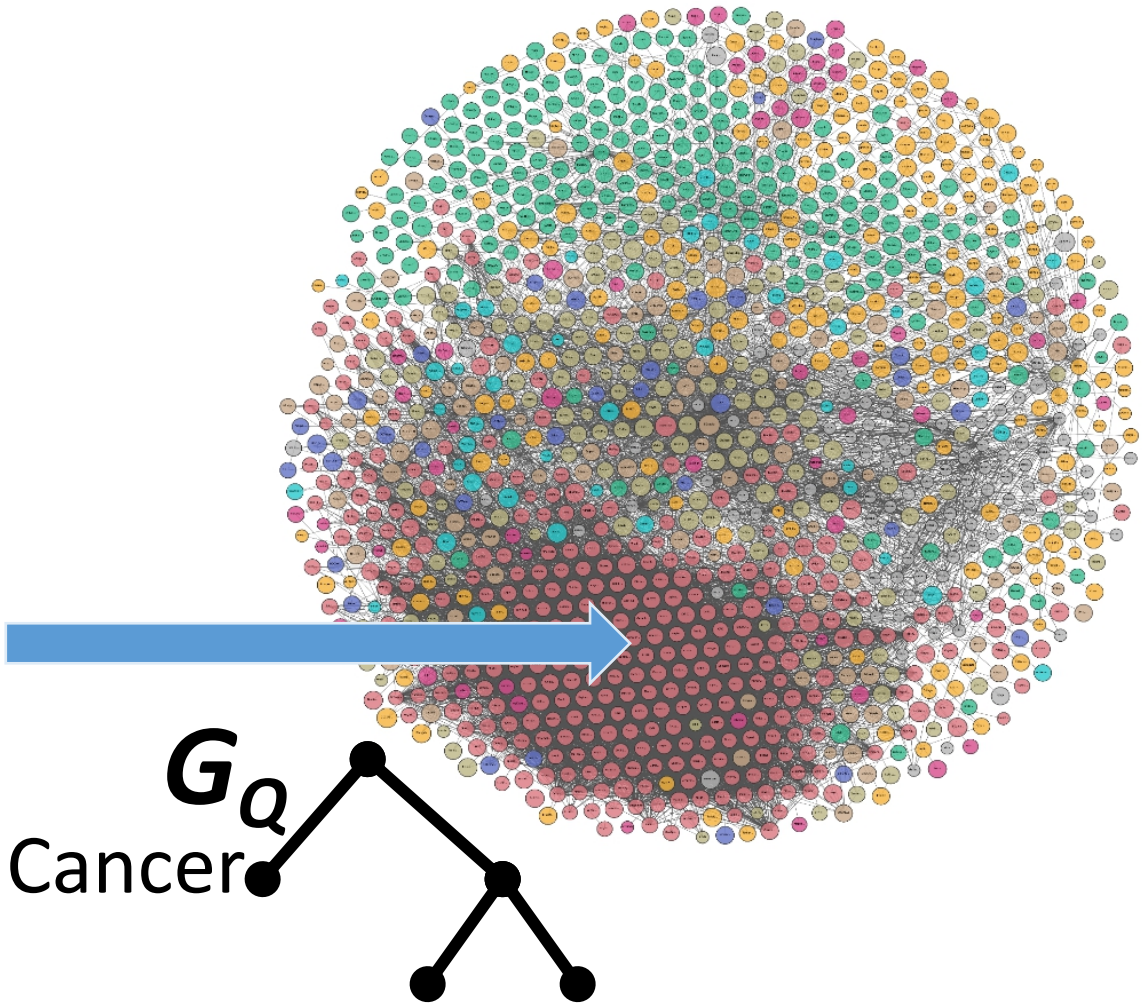
Results Format: HTML

Execution timeout: 0 milliseconds *(values less than 1000*

Options:
- ☑ Strict checking of void variables
- ☐ Log debug info at the end of output (has no effect on
- ☐ Generate SPARQL compilation report (instead of exe

(The result can only be sent back to browser, not saved on the server, see details)

Run Query | Reset



https://covidontheweb.inria.fr/sparql

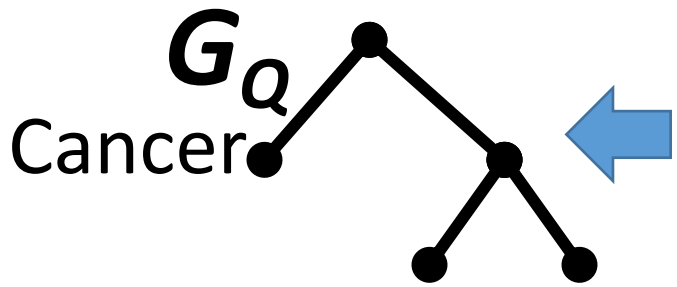
Virtuoso SPARQL Query Editor

Default Data Set Name (Graph IRI)

Query Text

```
select ?title
where {
  graph <http://ns.inria.fr/covid19/graph/articles> {
    ?paper1 a fabio:ResearchPaper; dct:title ?title.
  }

  graph <http://ns.inria.fr/covid19/graph/entityfishing> {
    ?a1 a oa:Annotation;
        schema:about ?paper1;
        oa:hasBody <http://www.wikidata.org/entity/Q12078> .
  }
} limit 100
```



Sponging: Use only local data (including data retrieved before), but

Results Format: HTML

Execution timeout: 0 milliseconds *(values less than 1000*

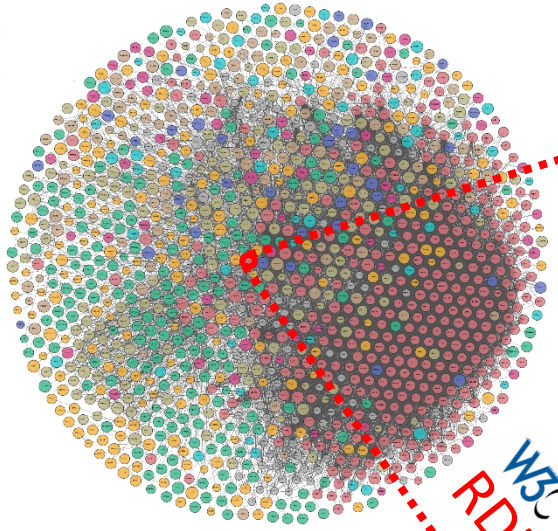
Options:
- [x] Strict checking of void variables
- [ ] Log debug info at the end of output (has no effect on
- [ ] Generate SPARQL compilation report (instead of exe

(The result can only be sent back to browser, not saved on the server, see details)

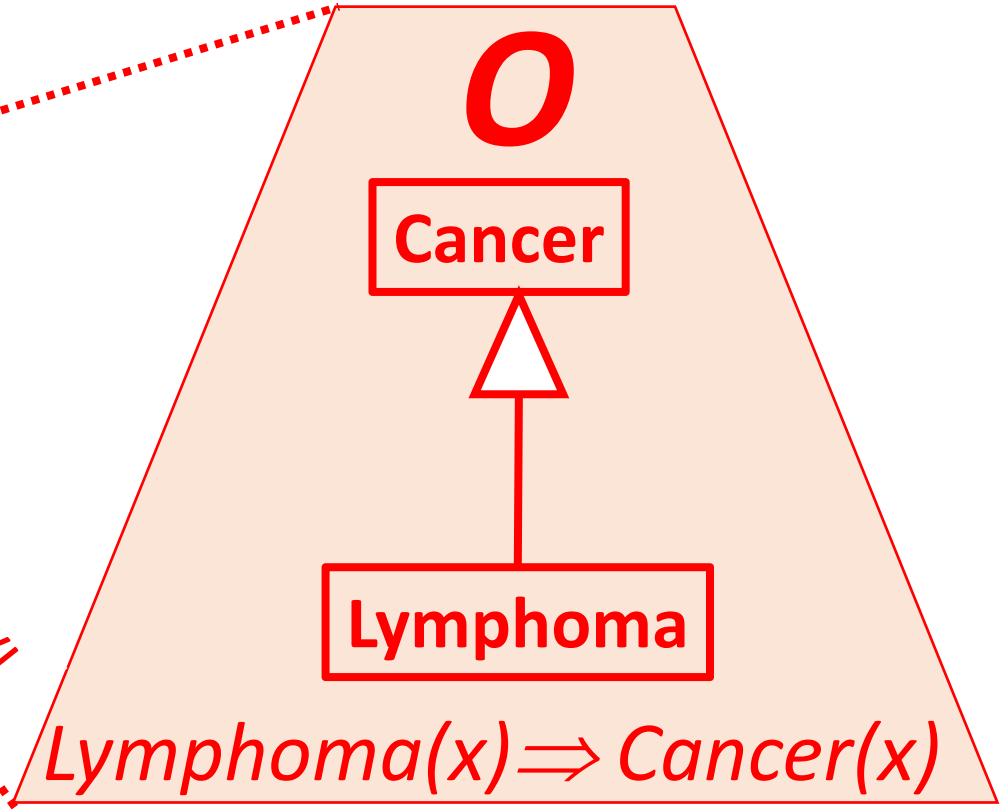
Run Query | Reset

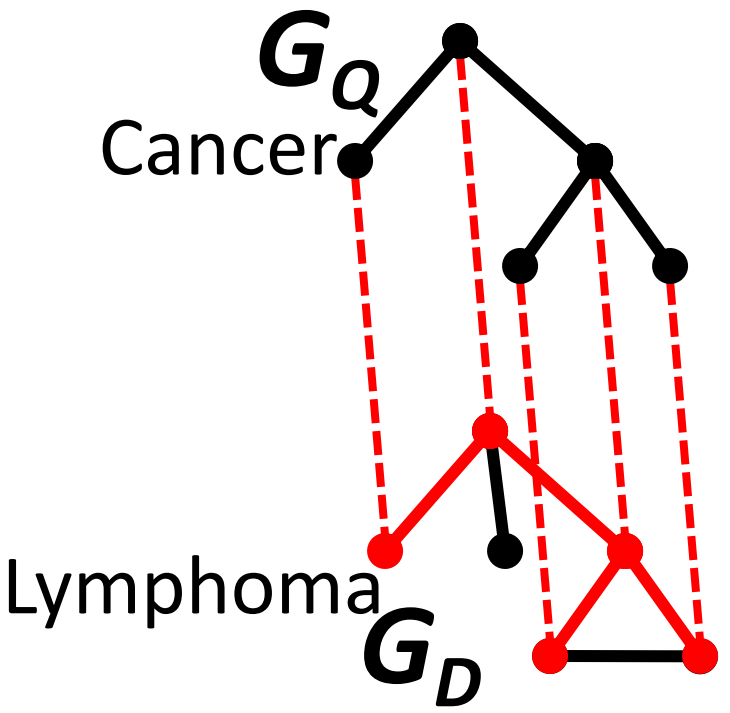
Articles about cancer in the COVID dataset:

- *"Antipsychotic treatment effects on cardiovascular, cancer, infection, and intentional self-harm as cause of death in patients with Alzheimer's dementia"*
- *"Targeting cancer stem cell pathways for cancer therapy"*
- *"Deubiquitinases and cancer: A snapshot"*
- *"The Functional Properties of Preserved Eggs: From Anti-cancer and Anti-inflammatory Aspects"*
- *"The functional role of the novel biomarker karyopherin α 2 (KPNA2) in cancer"*
- *"Biochemical characterisation of lectin from Indian hyacinth plant bulbs with potential inhibitory action against human cancer cells"*
- *"Purification, identification and profiling of serum amyloid A proteins from sera of advanced-stage cancer patients"*
- *"Darwin, medicine and cancer"*
- *"Outcome of Oncology Patients Infected With Coronavirus"*
- *"Review and Meta-Analyses of TAAR1 Expression in the Immune System and Cancers"*
- *"Experimental Data-Mining Analyses Reveal New Roles of Low-Intensity Ultrasound in Differentiating Cell Death Regulatome in Cancer and Non-cancer Cells via Potential Modulation of Chromatin Long-Range Interactions"*
- *"Golgi anti-apoptotic protein: a tale of camels, calcium, channels and cancer"*
- *"Severe novel influenza A (H1N1) infection in cancer patients"*
- *"Molecular Profiling of Multiple Human Cancers Defines an Inflammatory Cancer-Associated Molecular Pattern and Uncovers KPNA2 as a Uniform Poor Prognostic Cancer Marker"*
- *"SARS-CoV-2 transmission in cancer patients of a tertiary hospital in Wuhan"*
- *"Creosote bush lignans for human disease treatment and prevention: Perspectives on combination therapy"*
- *"8 Electrospinning and microfluidics An integrated approach for tissue engineering and cancer"*
- *"Genomic and proteomic approaches for studying human cancer: Prospects for true patient-tailored therapy"*
- *"Community acquired respiratory virus infections in cancer patients—Guideline on diagnosis and management by the Infectious Diseases Working Party of the German Society for haematology and Medical Oncology"*
- *"RIG-I Enhanced Interferon Independent Apoptosis upon Junin Virus Infection"*
- *"Respiratory Viral Infections in Transplant and Oncology Patients"* ....



RDFS, OWL





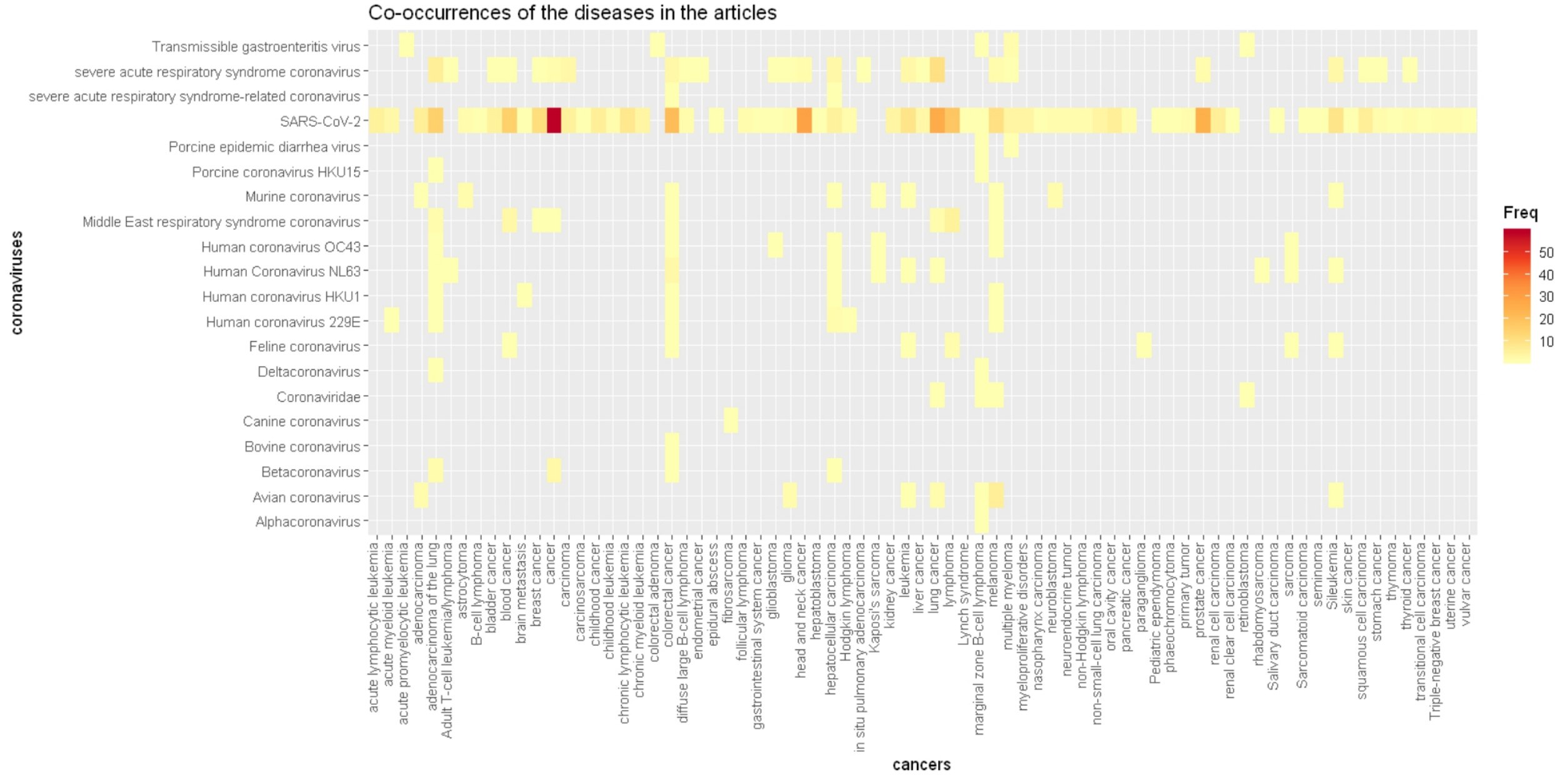
CORESE

[Corby et al.]

$$F \wedge O \rightarrow R \Leftrightarrow G_D \leq G_Q$$

mapping modulo an ontology

AI methods: knowledge graphs, ontology-based formalisms, querying, validating and reasoning



Query for co-occurrence of cancers & coronaviruses with R Jupyter Notebooks…

cancers

Plot heatmap with ggplot

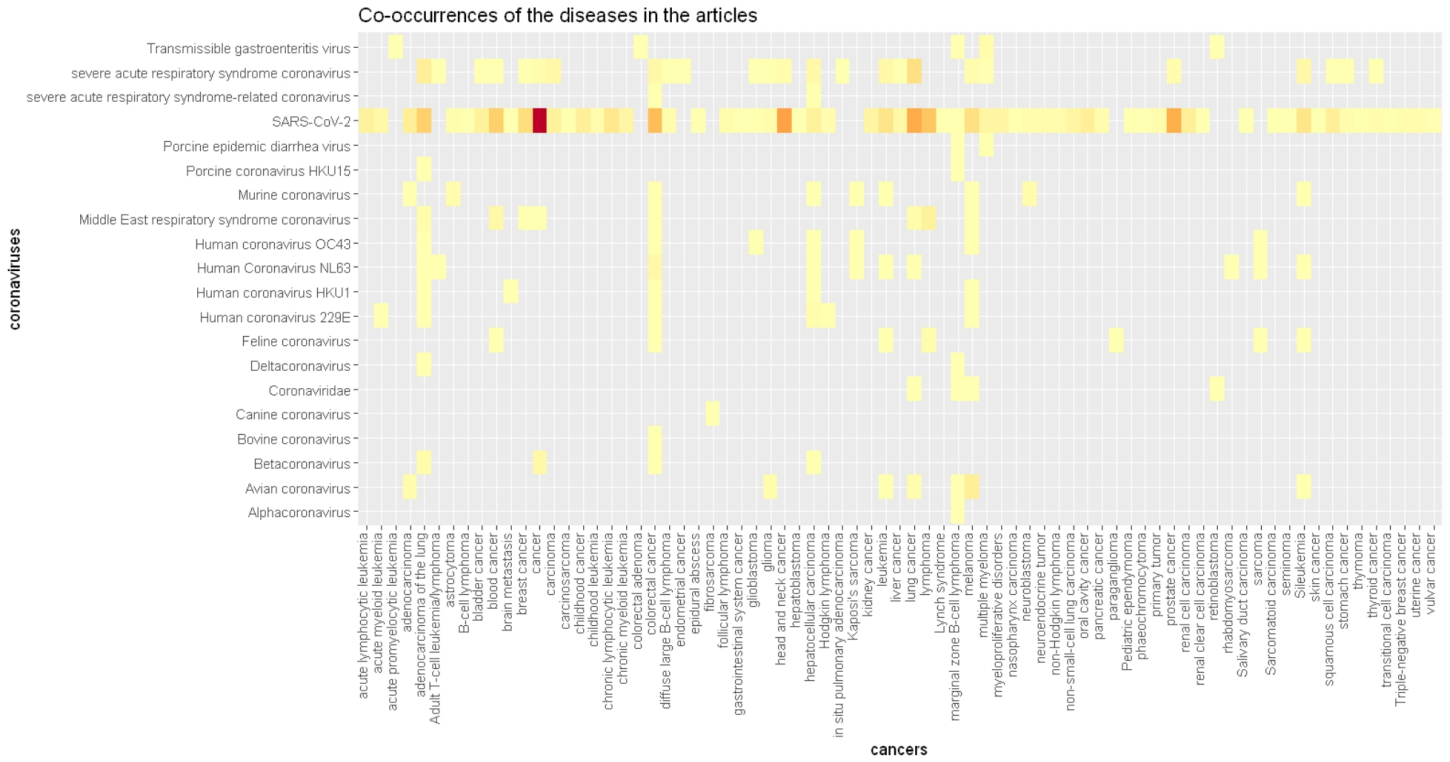
```
options(warn=-1)

library(ggplot2)
```

```
Registered S3 methods overwritten by 'ggplot2':
  method         from
  [.quosures     rlang
  c.quosures     rlang
  print.quosures rlang
```

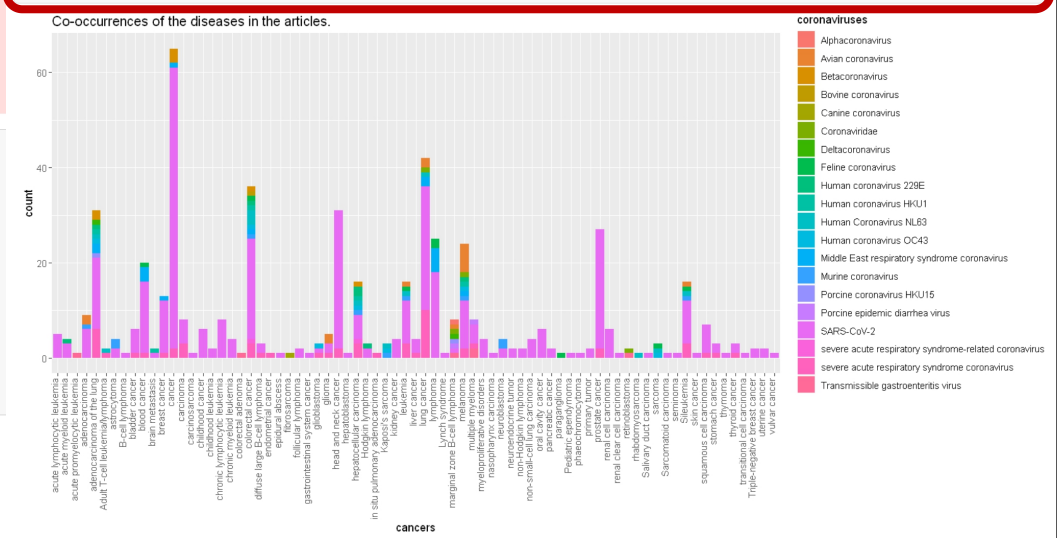
```
counts <- table(res[, c('dis2Label','dis1Label')])
counts <- as.data.frame(counts)
counts <- counts[counts$Freq > 0, ]

ggplot(counts, aes(x=dis1Label, y=dis2Label, fill=Freq)) +
       geom_tile() +
       theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust=0)) +
       scale_fill_distiller(palette = "YlOrRd", direction = 1) +
       ggtitle("Co-occurrences of the diseases in the articles") +
       xlab("cancers") +
       ylab("coronaviruses")
```



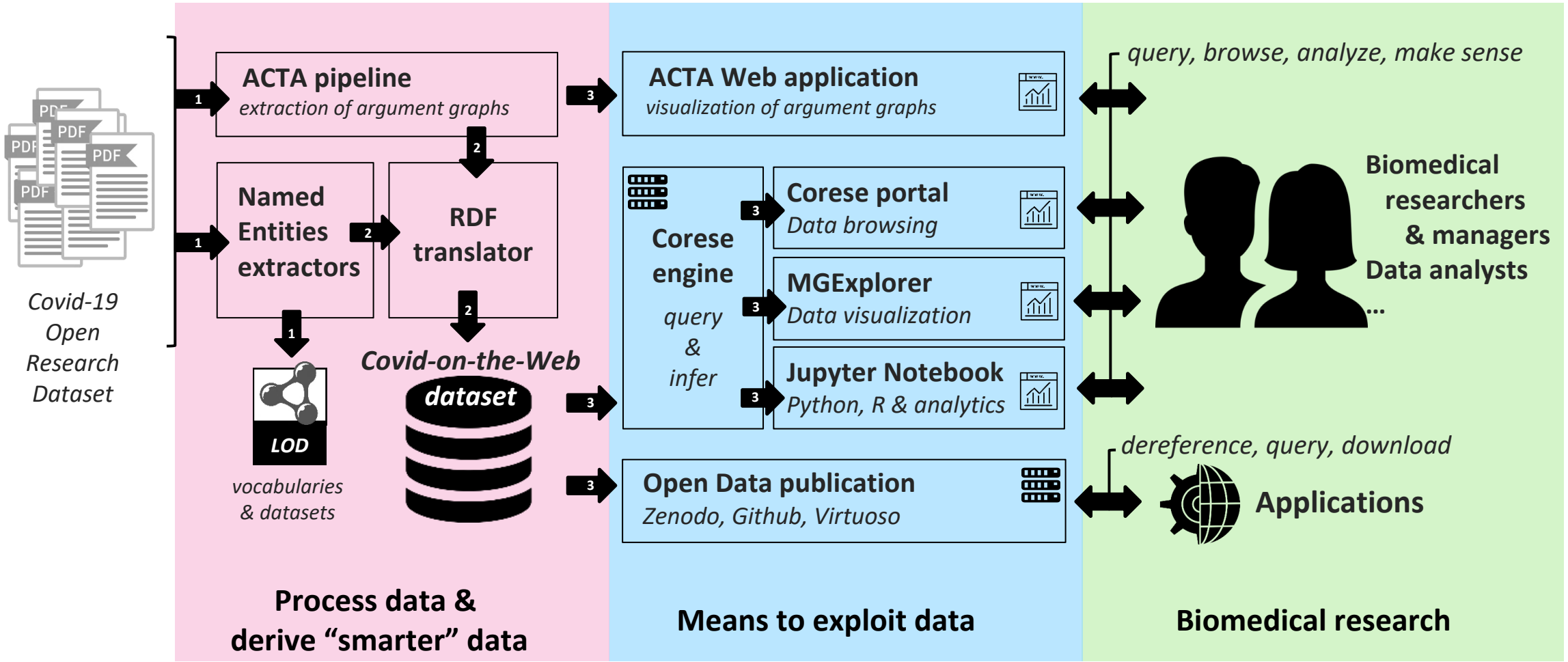
Plot stacked bar chart with default colors

```
ggplot(res,
       aes(x = dis1Label,
           fill = dis2Label)) +
  geom_bar(position = "stack") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, vjust=0)) +
  ggtitle("Co-occurrences of the diseases in the articles.") +
  xlab("cancers") +
  ylab("count") +
  labs(fill = "coronaviruses")
```



But you need to code… 

COVID ON THE WEB [ISWC 2020, IC 2021]



[Menin, Winckler, Corby, Giboin, Faron, Cadorel, Tettamanzi et al. 2020]

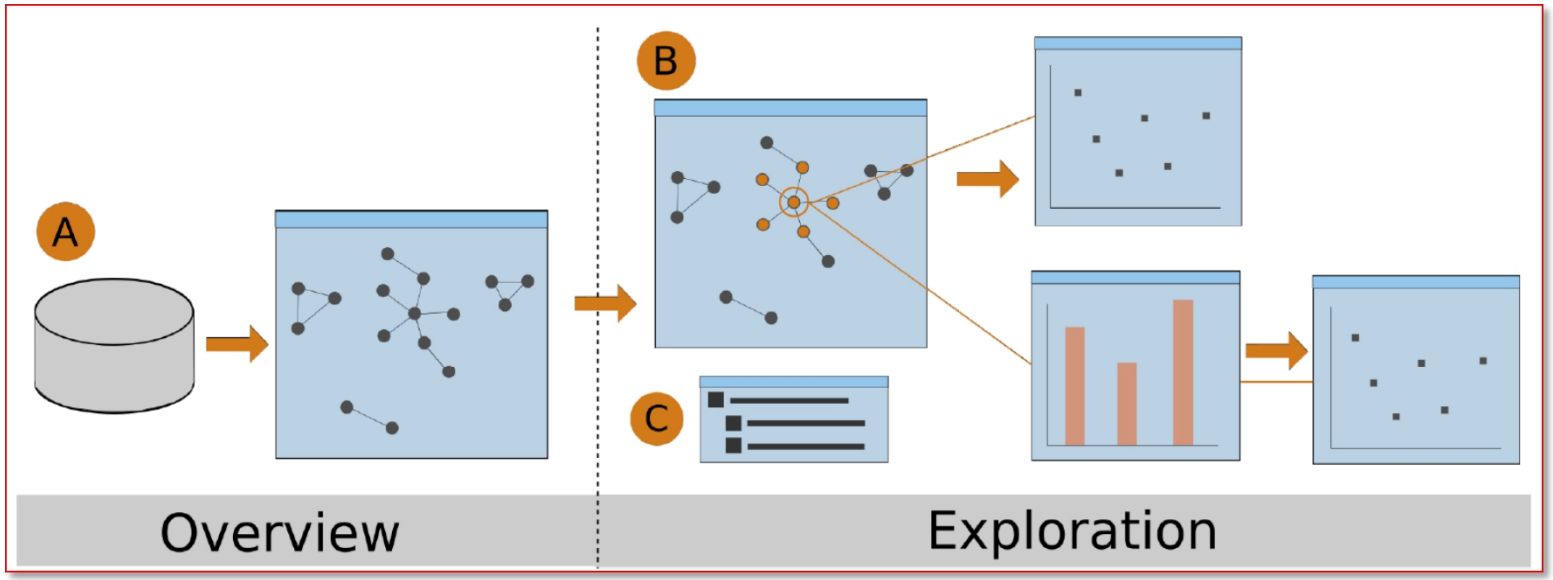
Examples of questions from [Giboin, Faron et al. 2020]

INSTITUT NATIONAL DU CANCER

- Est-ce que les Coronavirus peuvent causer des cancers ?
- Est-ce qu'ils font partie de la famille des virus oncogènes ?
- Quelles sont les séquelles des infections SARS CoV 1 et 2, et MERS ?
- Est-ce que les épidémies SARS-Cov1 et MERS sont liées à des apparitions de cancers ?
- Est-ce que les lésions causées par l'infection SARS CoV2 peuvent potentialiser une transformation tumorale? (évolution tissulaire et sensibilité au développement de cancers suite à une fibrose, ou autres lésions)
- Quelles sont les voies de signalisation intracellulaires activées par les coronavirus ? Pendant l'inflammation ? Quelles sont les adaptations métaboliques ?
- Est-ce qu'il y a des similitudes avec des virus oncogènes déjà connus tels que HPV, HBV, EBV, etc. ?…

[Menin, Winckler, Corby, Giboin, Faron et al. 2020]

Visualisation pipeline



Query Management Interface
- SPARQL Endpoint
- SPARQL SELECT Query
- Graph Style Sheet

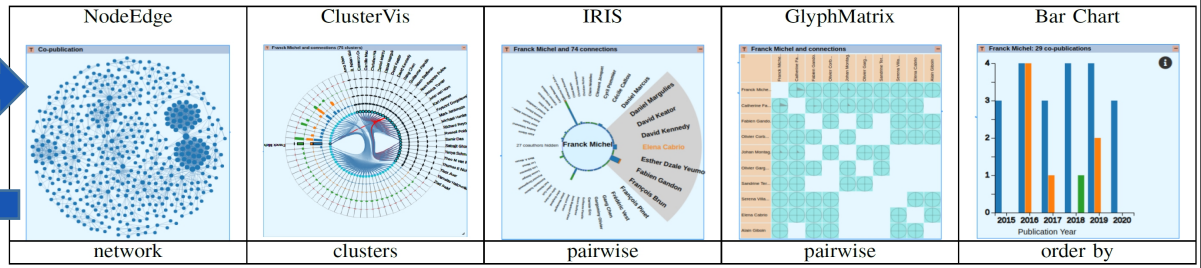
SPARQL Result Set (JSON)

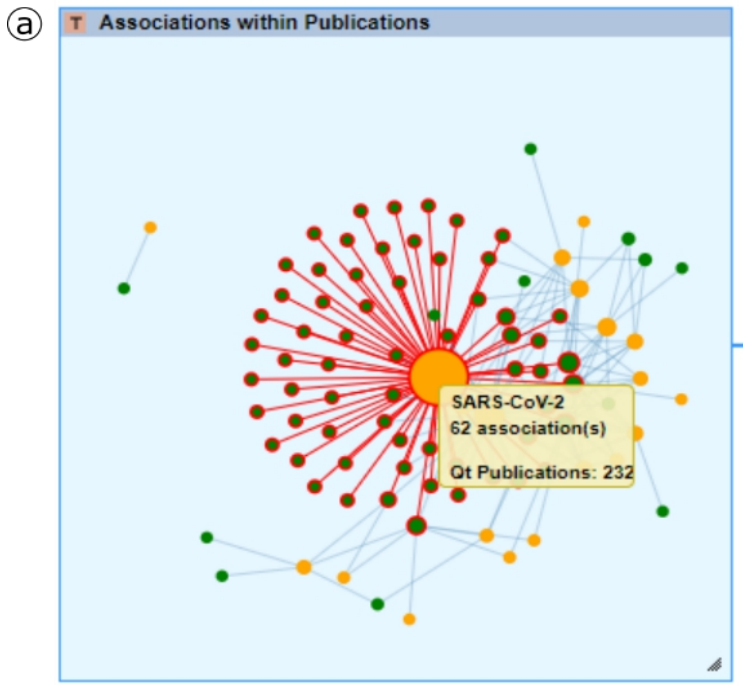
Transformation Engine

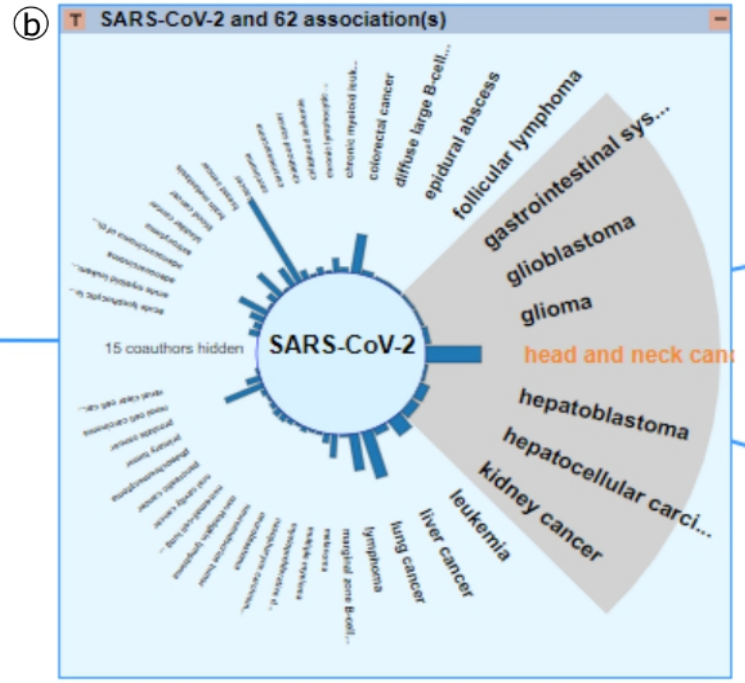
MGExplorer Data Model (Graph/JSON)

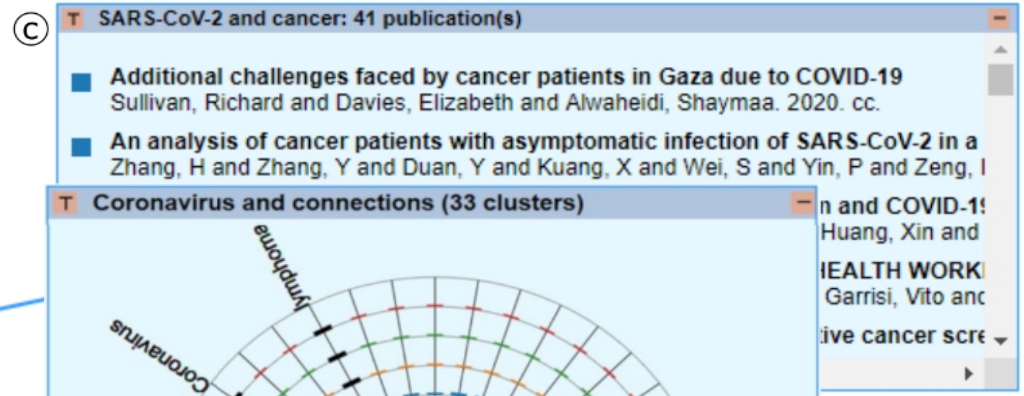
MGExplorer Visualization
- Multiple Views
- Network, clustering, pairwise relationship, temporal
- Customizable

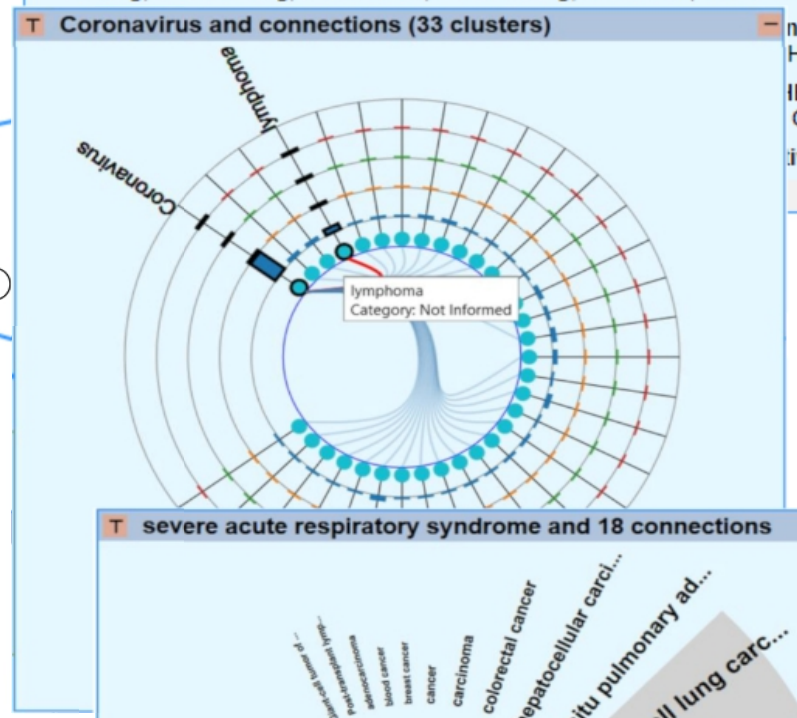
?

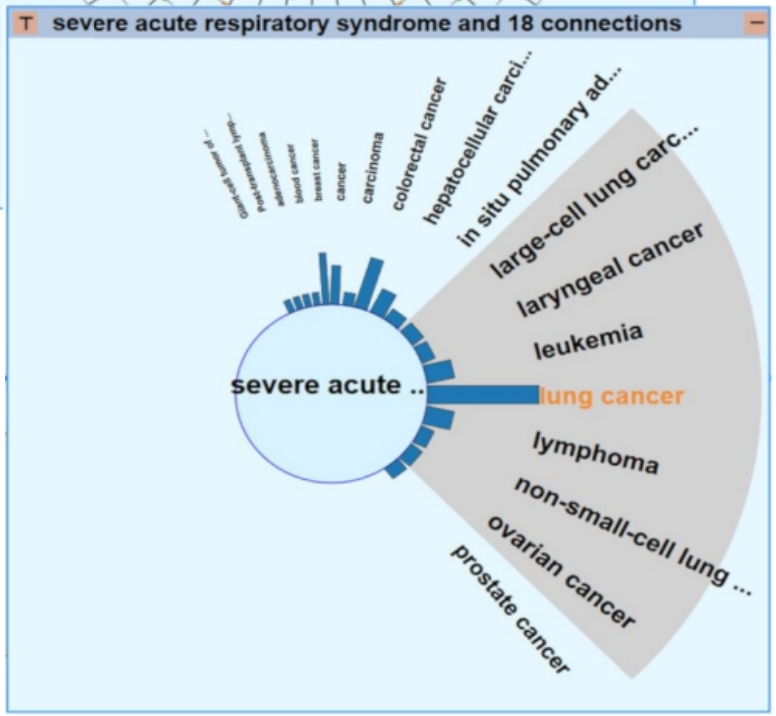












Visualisation and exploration

[Menin, Winckler, Corby, Giboin, Faron et al. 2020]

http://covid19.i3s.unice.fr:8080/

Mining interesting association rules

[Cadorel, Tettamanzi] [WI-IAT 2020]

AI methods: clustering + community detection + dimensionality reduction (auto-encoder) + Frequent Pattern Growth

- hidden patterns to enrich the dataset
- novel hypotheses for biomedical research

Antecedents	Consequents
fever, dyspnea	cough
runny nose	cough
anxiety	mental depression
surgical mask	respirator
cruise ship	diamond princess
exponential growth	basic reproduction number
liberia, western african ebola virus epidemic	guinea
people's republic of china, pneumonia	wuhan
camelus, middle east respiratory syndrome coronavirus	arabian peninsula
poultry, people's republic of china	influenza a virus subtype h7n9
tnf, cytokine	il10
eif2ak3, eif2ak2	atf6
p38 mitogen-activated protein kinases, pyrazolanthrone	sb203580
methyl, cholesterol	cyclodextrin
etiology, vasculitis	kawasaki disease
steroid, magnetic-resonance imaging	osteonecrosis
hepatitis, liver cirrhosis	hepatocellular carcinoma
pubmed	embase
facebook	twitter

Mining interesting association rules

[Cadorel, Tettamanzi] [WI-IAT 2020]

AI methods: clustering + community detection + dimensionality reduction (auto-encoder) + Frequent Pattern Growth

- hidden patterns to enrich the dataset
- novel hypotheses for biomedical research
- error detection in the dataset
- relevant clusters & communities for navigation

Antecedents	Consequents
fever, dyspnea	cough
runny nose	cough
anxiety	mental depression
surgical mask	respirator
cruise ship	diamond princess
exponential growth	basic reproduction number
liberia, western african ebola virus epidemic	guinea
people's republic of china, pneumonia	wuhan
camelus, middle east respiratory syndrome coronavirus	arabian peninsula
poultry, people's republic of china	influenza a virus subtype h7n9
tnf, cytokine	il10
eif2ak3, eif2ak2	atf6
p38 mitogen-activated protein kinases, pyrazolanthrone	sb203580
methyl, cholesterol	cyclodextrin
etiology, vasculitis	kawasaki disease
steroid, magnetic-resonance imaging	osteonecrosis
hepatitis, liver cirrhosis	hepatocellular carcinoma
pubmed	embase
facebook	twitter

Error	Acronym	Associated Named Entities	Correct Named Entity
nokia n95	n95	personal protective equipment	mask n95
íþróttabandalag vestmannaeyja	IBV	avian infectious bronchitis, respiratory tract, chicken	Infectious bronchitis virus
a59 road	a59	glycoprotein	Mouse hepatitis virus A59
international federation of basque pelota	FIPV	feline infectious peritonitis, coronavirus, transmissible gastroenteritis virus	Feline Infectious Peritonitis Virus
new international version	NIV	henipavirus, vaccine, malaysia	Nipah Virus

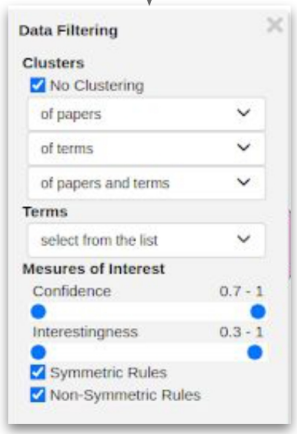
Visualization of Association found in Covid dataset [IV 2021]

[Menin, Cadorel, Tettamanzi, Winckler]

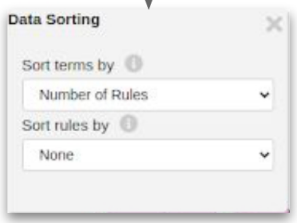
http://covid19.i3s.unice.fr:8080/arviz/explore

a) Legend

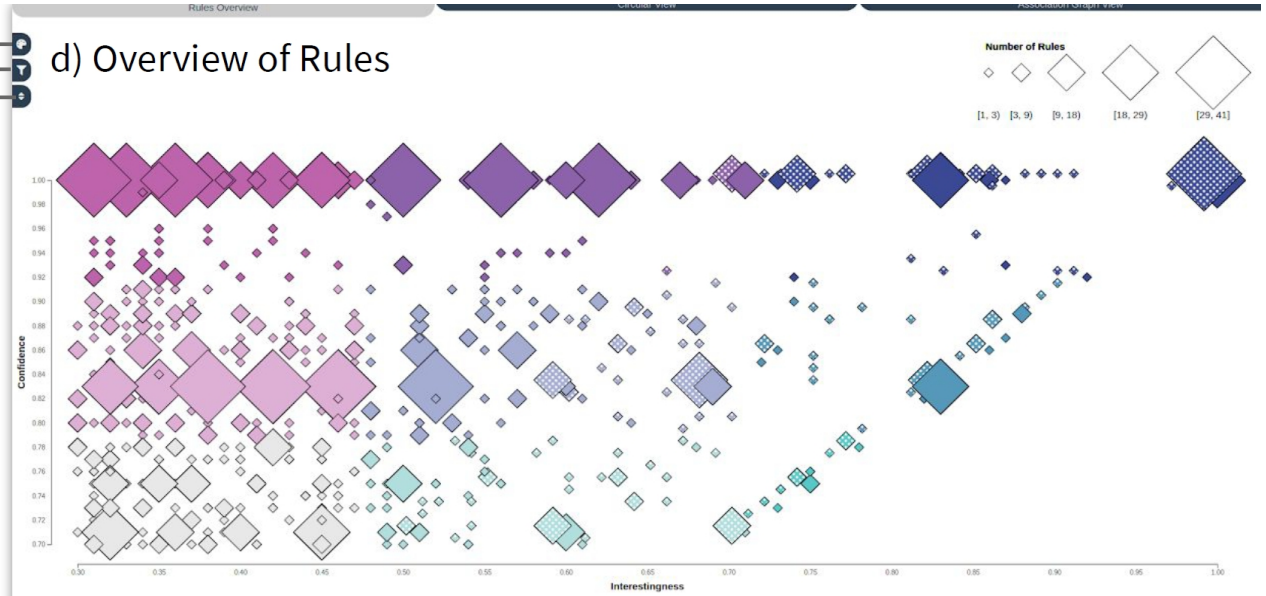
b) Data Filtering

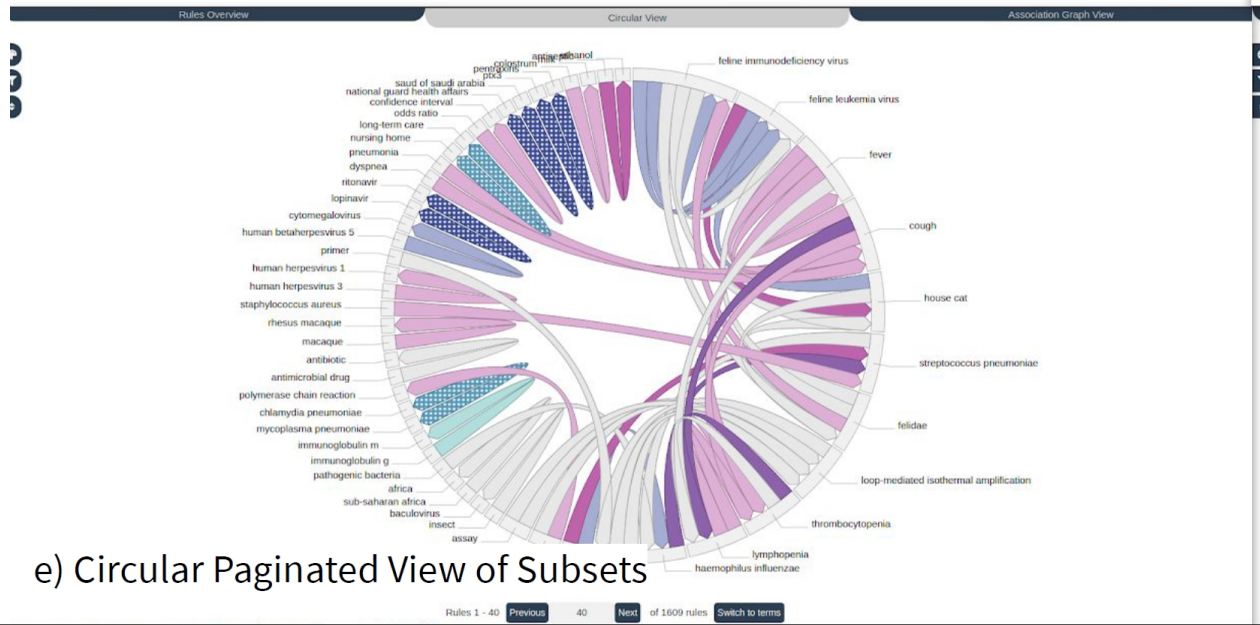


c) Data Sorting

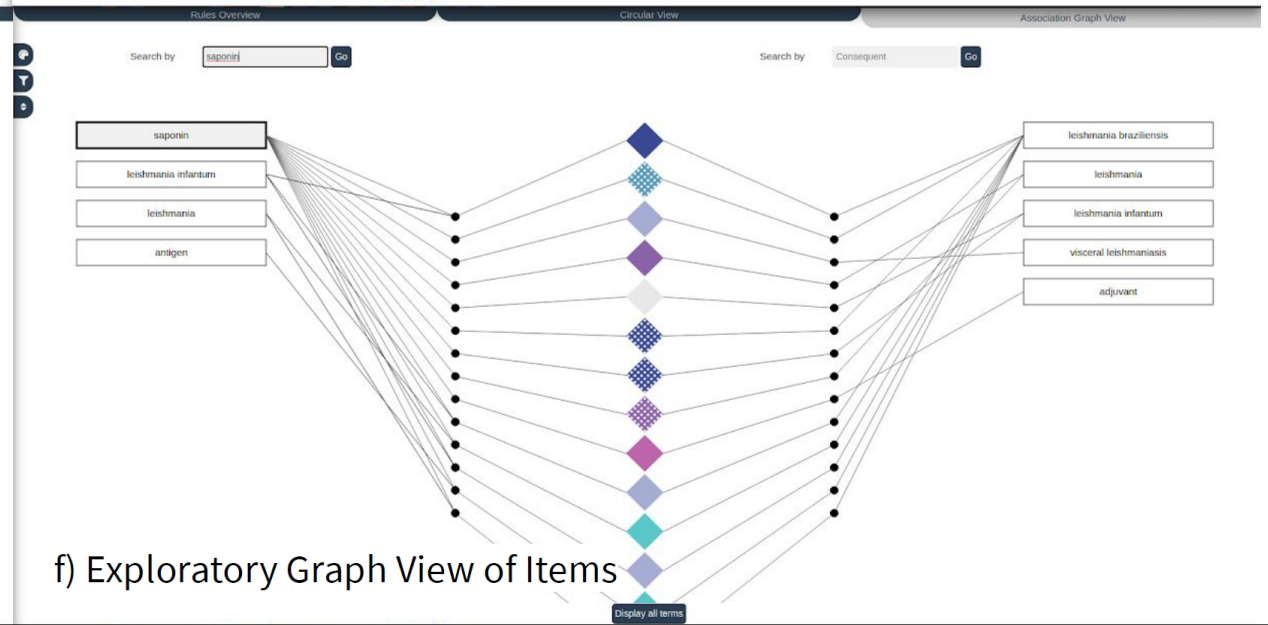


d) Overview of Rules

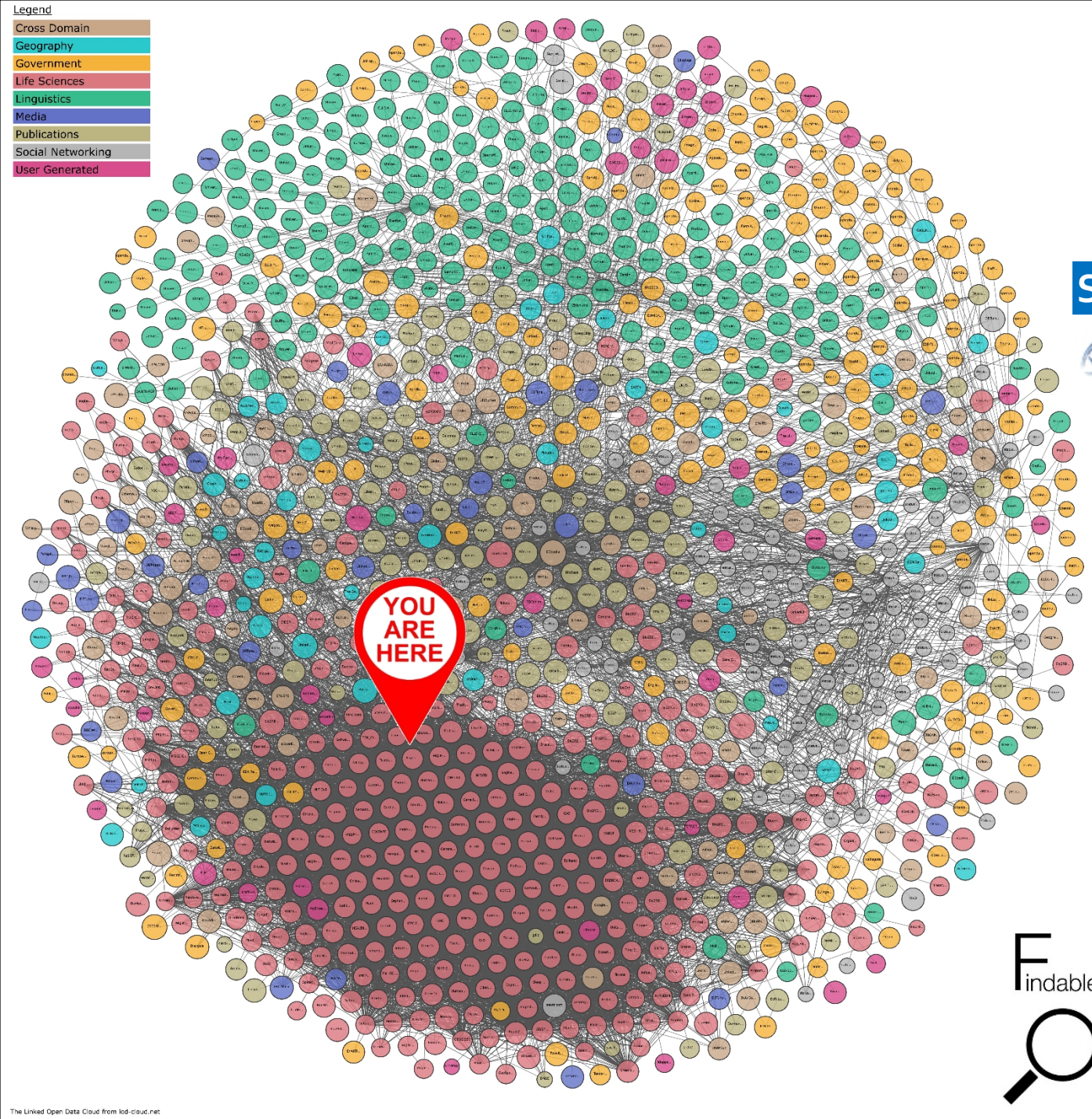




e) Circular Paginated View of Subsets



f) Exploratory Graph View of Items



CovidOnTheWeb

https://github.com/Wimmics/CovidOnTheWeb

SPARQL https://covidontheweb.inria.fr/sparql

POWERED BY VIRTUOSO https://covidontheweb.inria.fr/fct/

https://doi.org/10.5281/zenodo.3833753

Fabien Gandon, Franck Michel, Valentin Ah-Kane, Anna Bobasheva, Elena Cabrio, Olivier Corby, Catherine Faron, Raphaël Gazzotti, Alain Giboin, Santiago Marro, Tobias Mayer, Aline Menin, Mathieu Simon, Serena Villata, and Marco Winckler

Inria UNIVERSITÉ CÔTE D'AZUR cnrs i3s sophia antipolis

Findable Accessible Interoperable Reusable LOD W3C®

CovidOnTheWeb access Stats [Michel, Gazzotti]

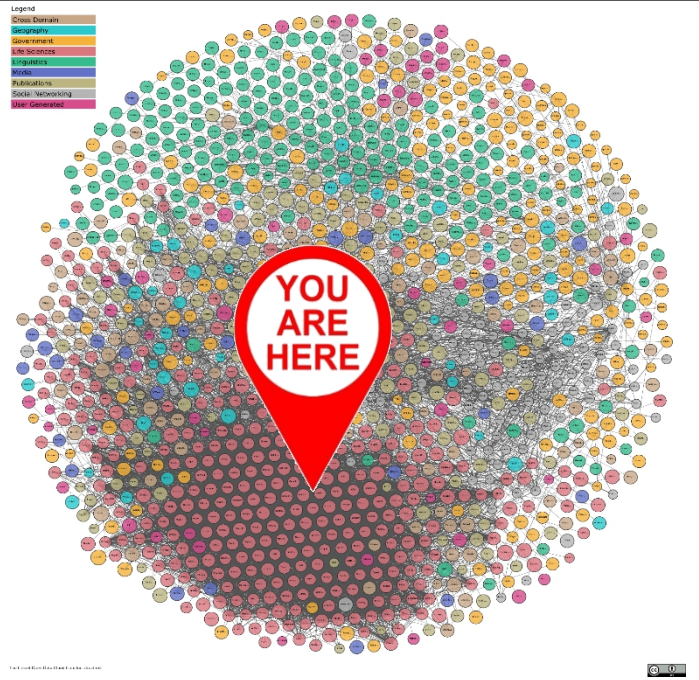
Period January-April 2021

- Total URI accesses= 48 735 hence an average of 393 URI access/day
- Total SPARQL queries 49 036 hence an average of 395 queries/day
- 300 different agent types with two major ones:
  Apache-Jena-ARQ (38 767) and Mozilla/4.0 (40 935)

Full dump of dataset on Zenodo

(end of April 2021)

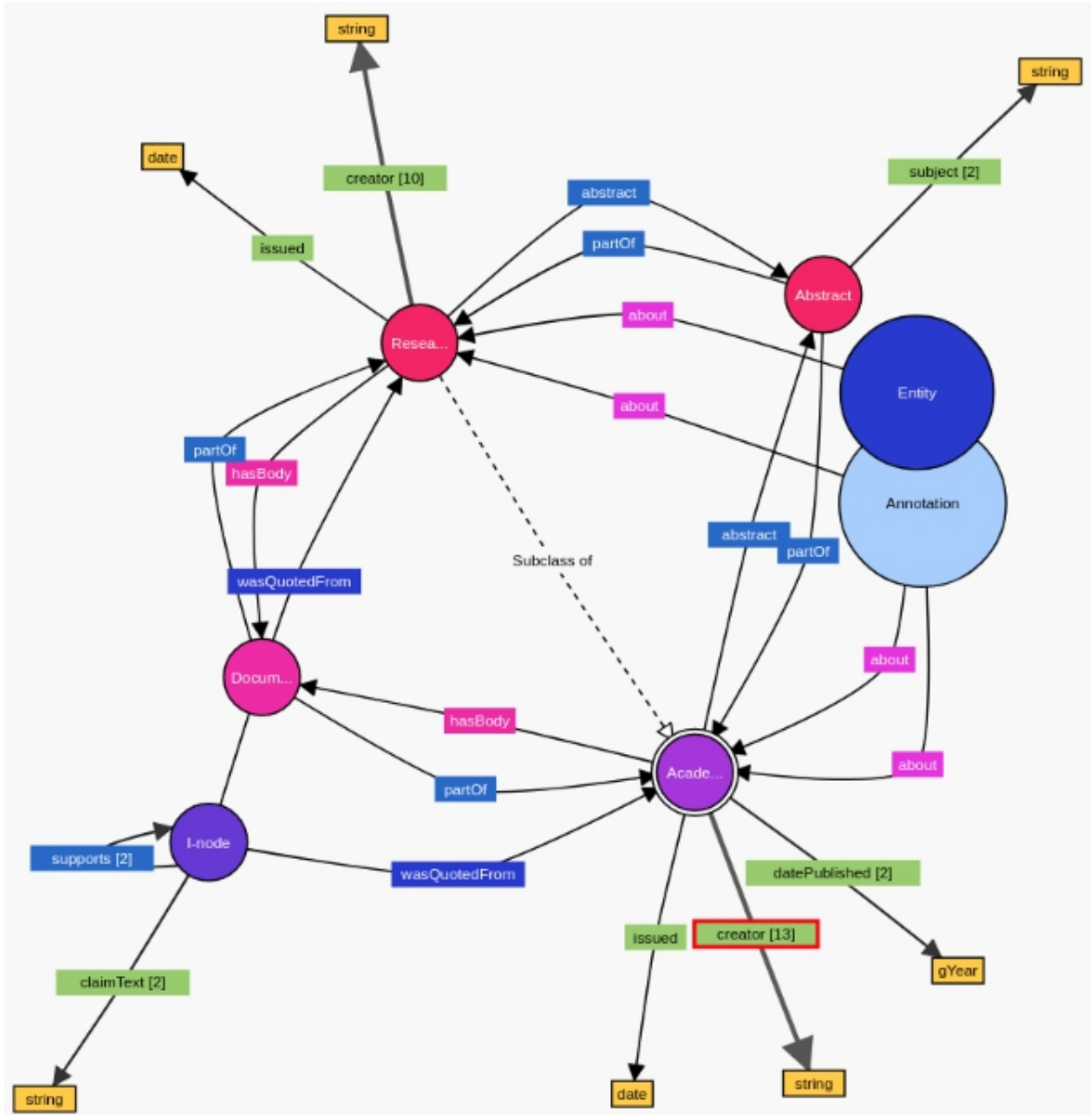
- 61 unique downloads
- 954 unique views



Type of data	JSON data	Resources produced	RDF triples
Articles metadata and textual content	7.4 GB	n.a.	1.27 M
CORD-19 Named Entities Knowledge Graph			
NEs found by DBpedia Spotlight (titles, abstracts)	35 GB	1.79 M	28.6 M
NEs found by Entity-fishing (titles, abstracts, bodies)	23 GB	30.8 M	588 M
NEs found by BioPortal Annotator (titles, abstracts)	17 GB	21.8 M	52.8 M
CORD-19 Argumentative Knowledge Graph			
Claims/evidence components (abstracts)	138 MB	53 K	545 K
PICO elements		229 K	2.56 M
Total for Covid-on-the-Web (including articles metadata and content)			
	82 GB	54 M named entities 53 K claims/evidence 229 K PICO elements	674 M

Dataset description	No. RDF triples
dataset description + definition of a few properties	170
articles metadata (title, authors, DOIs, journal etc.)	3 722 381
named entities identified by Entity-fishing in articles titles/abstracts	35 049 832
named entities identified by Entity-fishing in articles bodies	1 156 611 321
named entities identified by Bioportal Annotator in articles titles/abstracts	104 430 547
named entities identified by DBpedia Spotlight in articles titles/abstracts	65 359 664
argumentative components and PICO elements by ACTA from articles titles/abstracts	7 469 234
Total	**1 361 451 364**

Covid-on-the-Web RDF graph generated with LD-VOWL



Description of the graphical primitives and color scheme :

http://vowl.visualdataweb.org/v2/