Towards Representing Research Contributions in Scholarly Knowledge Graphs Using Knowledge Graph Cells

Presentation · August 2020

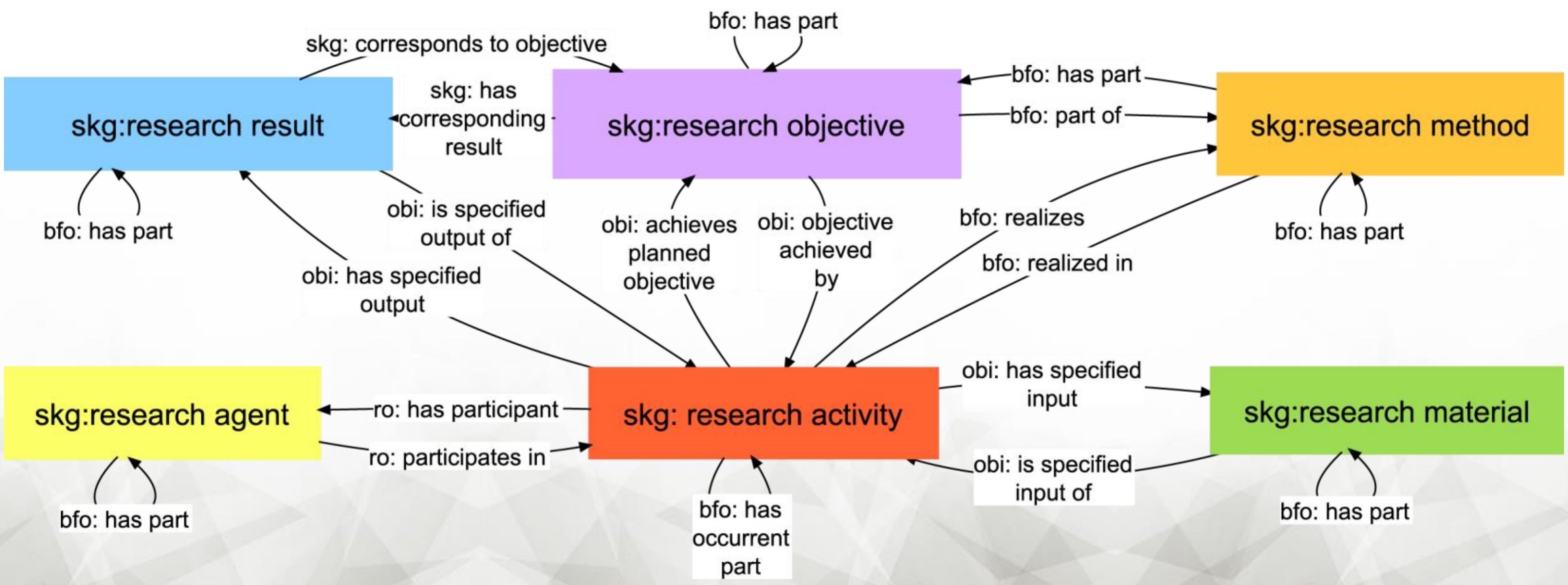
DOI: 10.13140/RG.2.2.24098.86720

Lars Vogt
Leibniz Information Centre for Science and Technology University Library

Jennifer D'Souza
Leibniz Information Centre for Science and Technology University Library

Towards Representing Research Contributions in Scholarly Knowledge Graphs Using Knowledge Graph Cells

L. Vogt, J. D'Souza, M. Stocker, S. Auer



Scholarly Communication

The Problem

Research is predominantly communicated using unstructured text. And the contents of scholarly publications are not machine-actionable.

Scholarly Communication

The Problem

The number of publications is increasing exponentially, with more than 2.5 million new publications being issued each year.

Jinha (2010) Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing* 23(3), 258–263.

Scholarly Communication

The Problem

As a result, researchers often have to search the contents of potentially relevant publications themselves to find those that are actually relevant to them. This becomes increasingly impossible with an exponentially increasing number of publications.

Landhuis (2016) Scientific literature: information overload. *Nature* 535(7612), 457–458.

Scholarly Communication

The Solution

A promising solution to this knowledge management problem is to make these contents machine-actionable.

Scholarly Communication

The Context – The FAIR Guiding Principles

Research communities and public stake holders have identified an increasing demand for the Findability, Accessibility, Interoperability, and Reusability of data and metadata (FAIR Guiding Principles).

Wilkinson *et al.* (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1): 160018.

Scholarly Communication

The Context – The FAIR Guiding Principles

The use of ontologies, knowledge graphs, the Resource Description Framework (RDF), and the Web Ontology Language (OWL) have proven to provide the semantic technologies required for implementing the FAIR Guiding Principles.

Jacobsen *et al.* (2019) FAIR principles: interpretations and implementation considerations. *Data Intelligence* 2(1-2), 10–29.
Vogt *et al.* (2019) SOCCOMAS: a FAIR web content management system that uses knowledge graphs and that is based on semantic programming. *Database* 2019: baz067.

Scholarly Communication

The Goal – A knowledge graph of scholarly publications

Our goal is to develop a knowledge graph of scholarly publications that makes available the contents of the publications in a form that complies with the FAIR Principles.

Challenges

Machine-actionability requires common data model

In order to build a FAIR knowledge graph of scholarly publications, some challenges have to be overcome. First, machine-actionability and FAIRness requires data to be formalized and standardized to some degree. This requires the development of a basic data model for integrating scholarly contents within the graph (Research Contribution Model).

Challenges

Machine-actionable graphs are often too complex for human readers

Another challenge results from the discrepancy between the requirements for machine-actionability as opposed to human-readability of data, with the former usually resulting in overly complex graphs with lots of information that is irrelevant for a human reader. We developed the idea of Knowledge Graphs Cells to tackle this problem.

The Research Contribution Model (RCM)

The Research Contribution Model (RCM)

Challenge

Modelling the contents of scholarly publications is very challenging, as they cover multiple domains, each with its own set of established standards and terminological conventions.

The Research Contribution Model (RCM)

Scholarly publications contain assertions

Scholarly publications contain assertions. An assertion is a proposition which is asserted or denied to be true. Independent of a particular domain, we can distinguish different basic types of assertions.

The Research Contribution Model (RCM)

Types of assertions

- empirical data (descriptive, phenomenon, observation, measurement)
- activity of producing data (process of observing, measuring, etc.)
- method of producing data (plan for a data production activity)

The Research Contribution Model (RCM)

Types of assertions

- result (ontological, hypothesis, theory, test-result value, explanation, etc.)
- activity of analyzing data (process of analyzing, testing, etc.)
- method of analyzing data (plan for a data analysis activity)

The Research Contribution Model (RCM)

Types of assertions

Data (phenomena), results (theories), and methods (plans for research activities) are the key elements of scholarly discourse.

Szostak (2007) Interdisciplinarity and the Classification of Scholarly Documents by Phenomena, Theories, and Methods. *Interdisciplinarity and transdisciplinarity in the organization of scientific knowledge: Actas del VIII Congreso ISKO-España*, 469–477.

The Research Contribution Model (RCM)

Types of assertions

For each type of a scholarly assertion, we specify a graph pattern/template.

The Research Contribution Model (RCM)

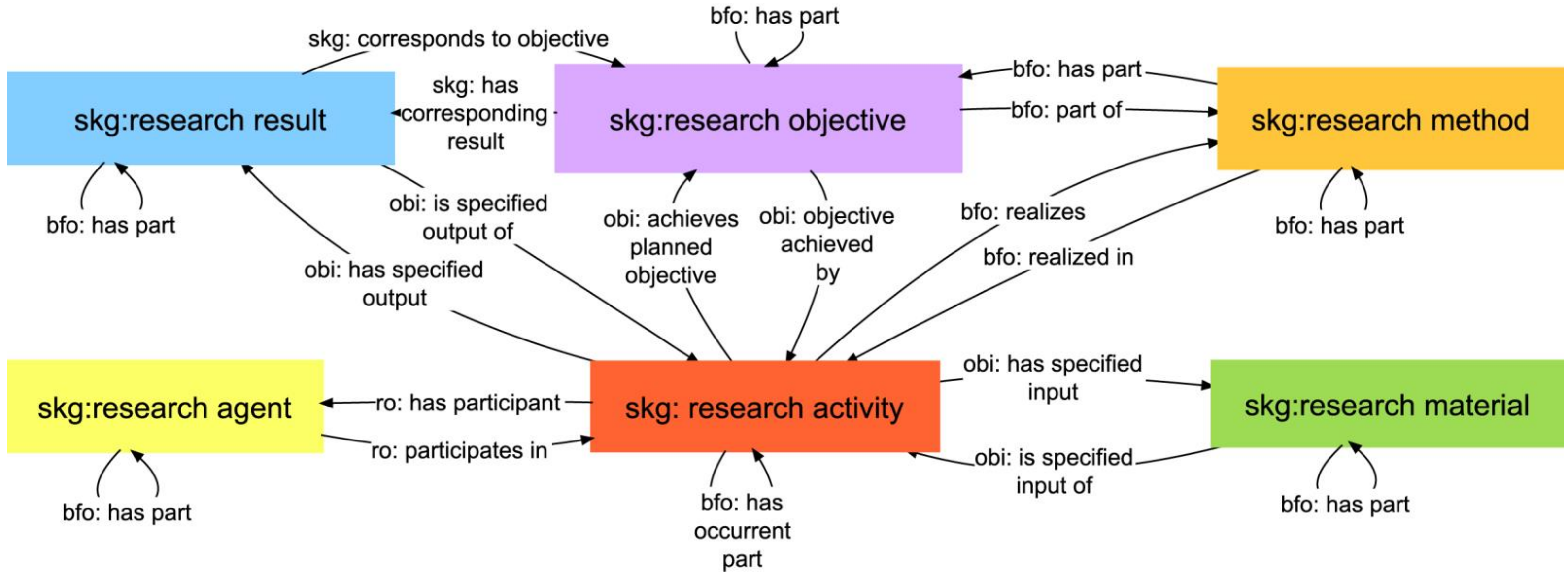
Types of assertions

- result (ontological, hypothesis, theory, test-result value, ex
- ac
  ar
- m
  analysis activity)

And the RCM relates them to one another.

The Research Contribution Model (RCM)

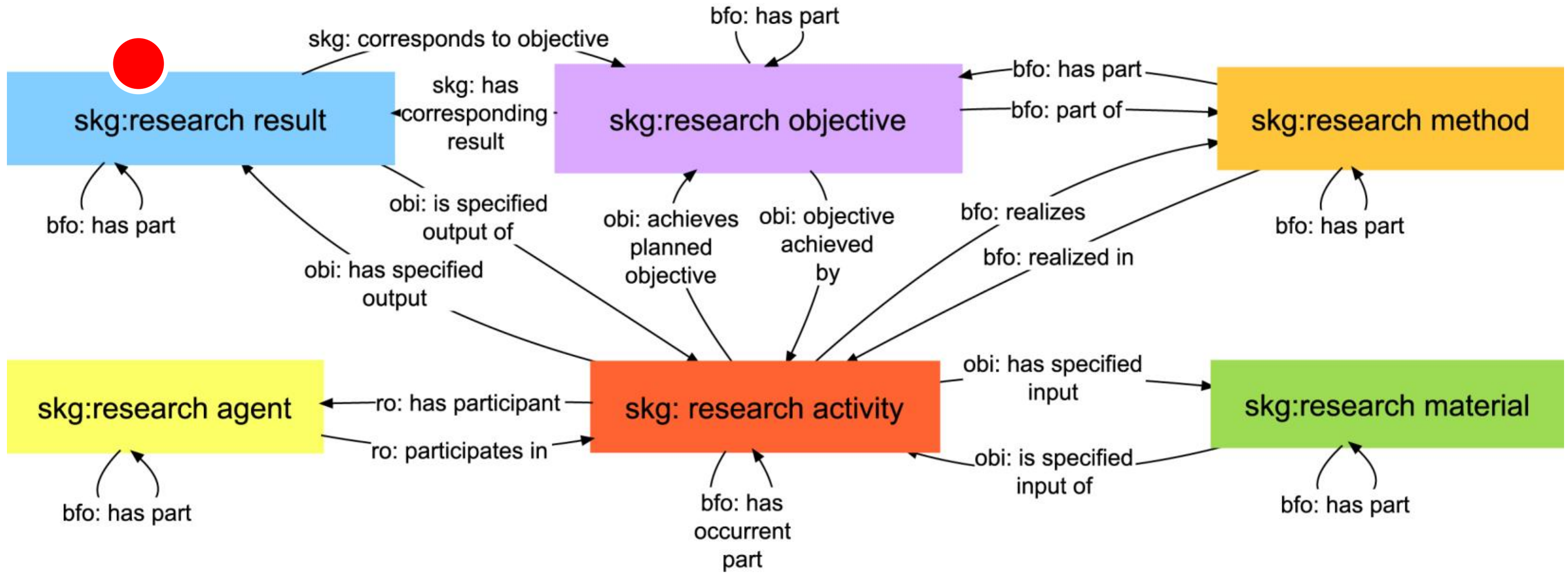
Integrating different types of assertions



The Research Contribution Model (RCM)

Integrating different types of assertions

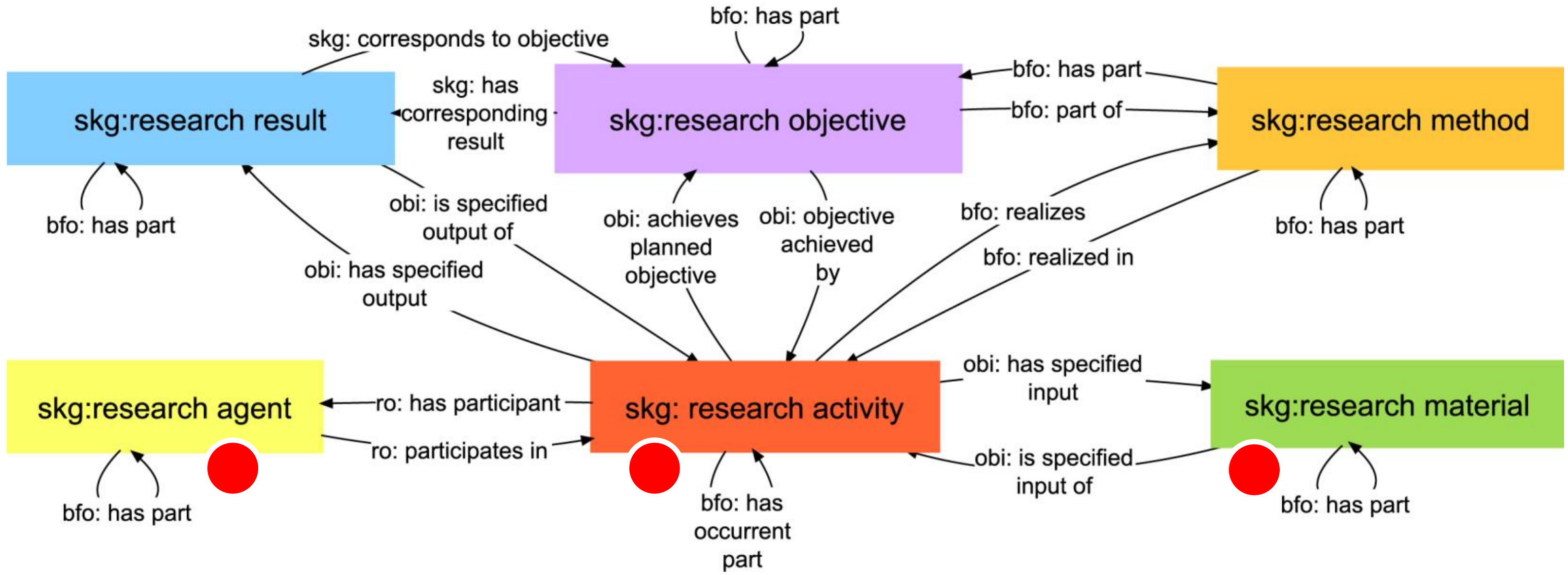
empirical data



The Research Contribution Model (RCM)

Integrating different types of assertions

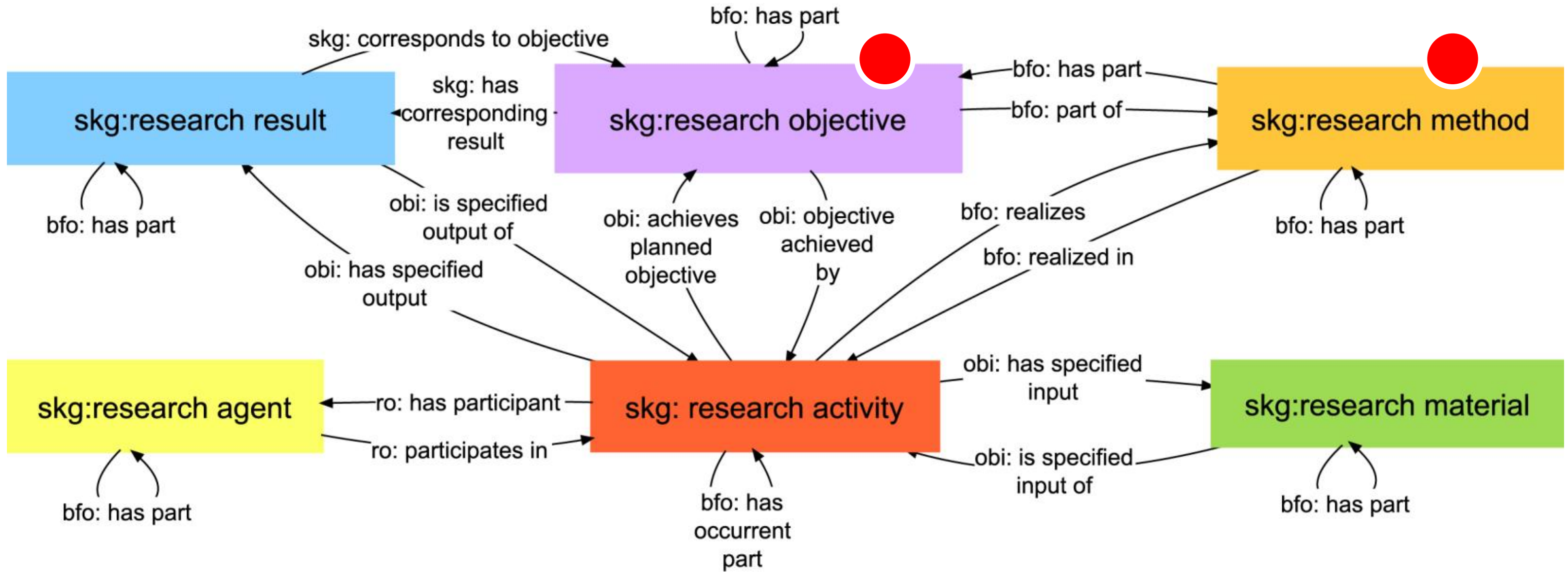
activity of producing data



The Research Contribution Model (RCM)

Integrating different types of assertions

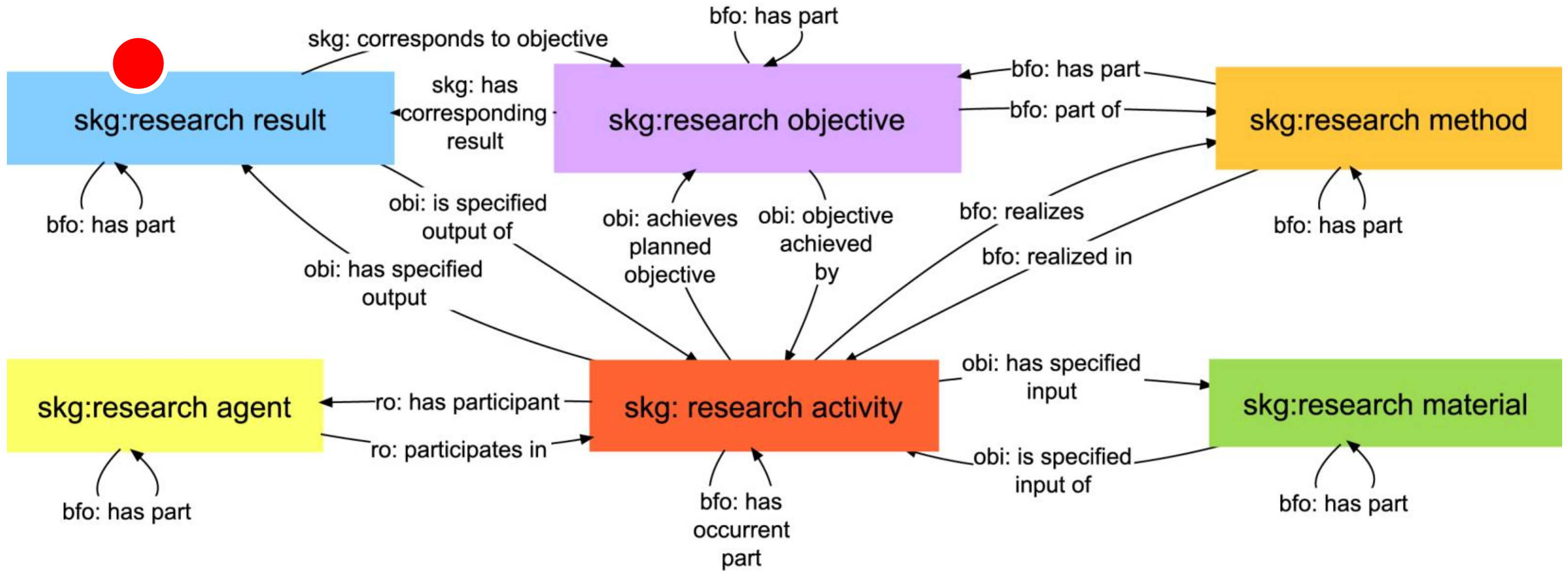
method of producing data



The Research Contribution Model (RCM)

Integrating different types of assertions

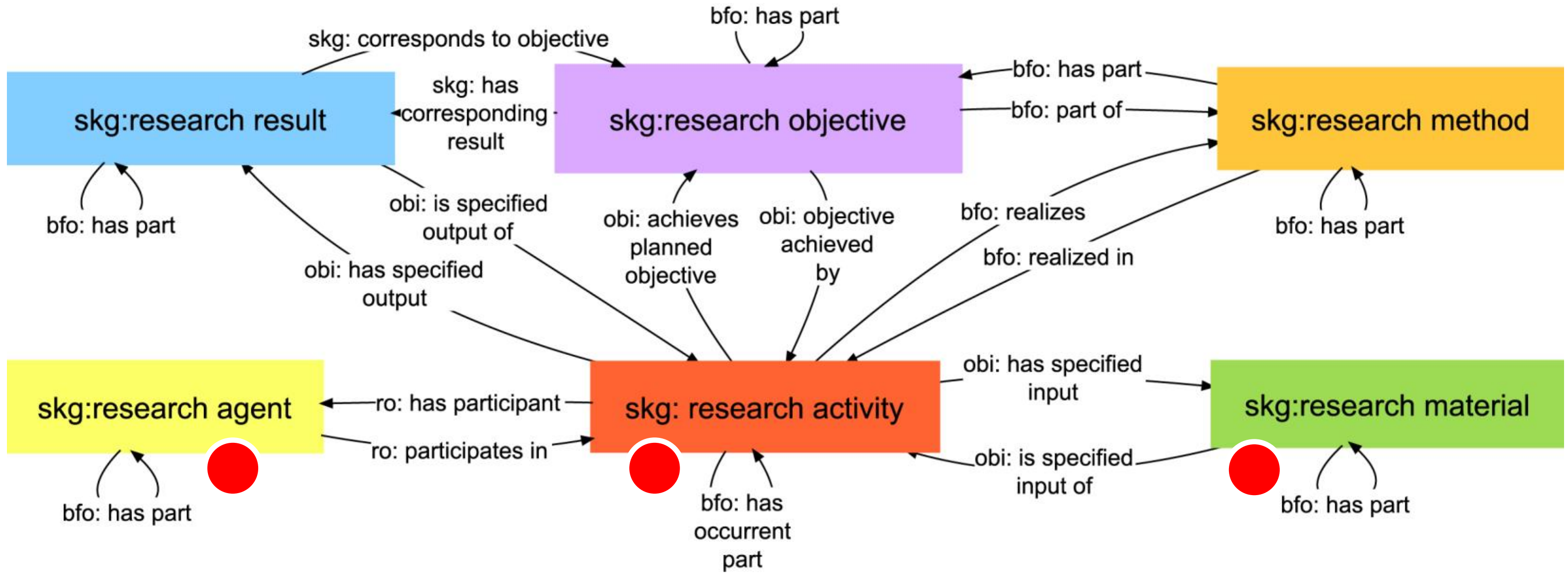
result



The Research Contribution Model (RCM)

Integrating different types of assertions

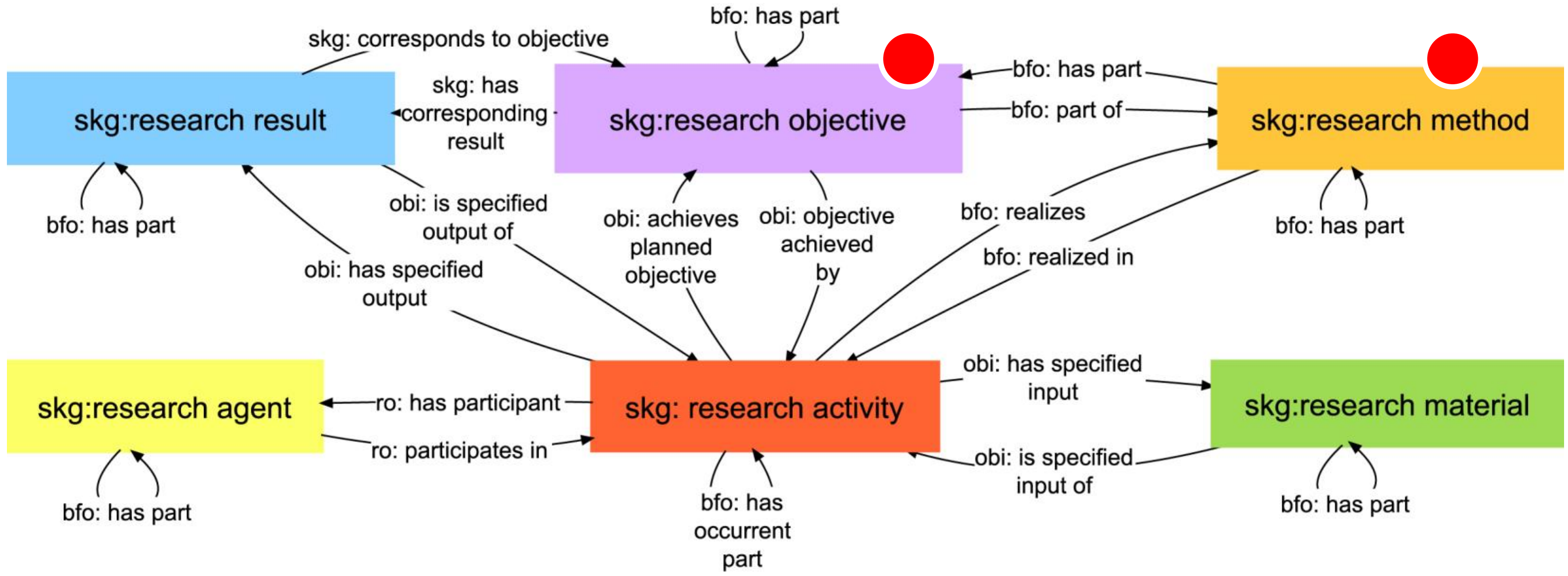
activity of analyzing data



The Research Contribution Model (RCM)

Integrating different types of assertions

method of analyzing data



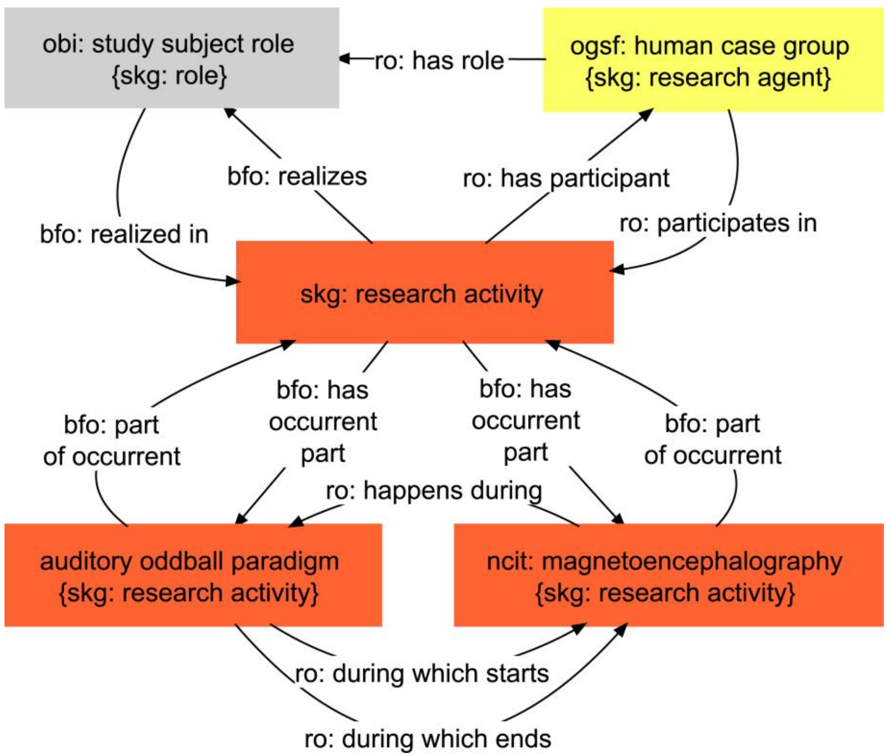
The Research Contribution Model (RCM)

Inherent nested hierarchy

Because the world is granular and because every whole can be partitioned into finer grains—a research result consists of sets of results, a method of sets of steps, an activity of sets of events and material entities of sets of parts—the RCM can be applied like a Matryoshka.

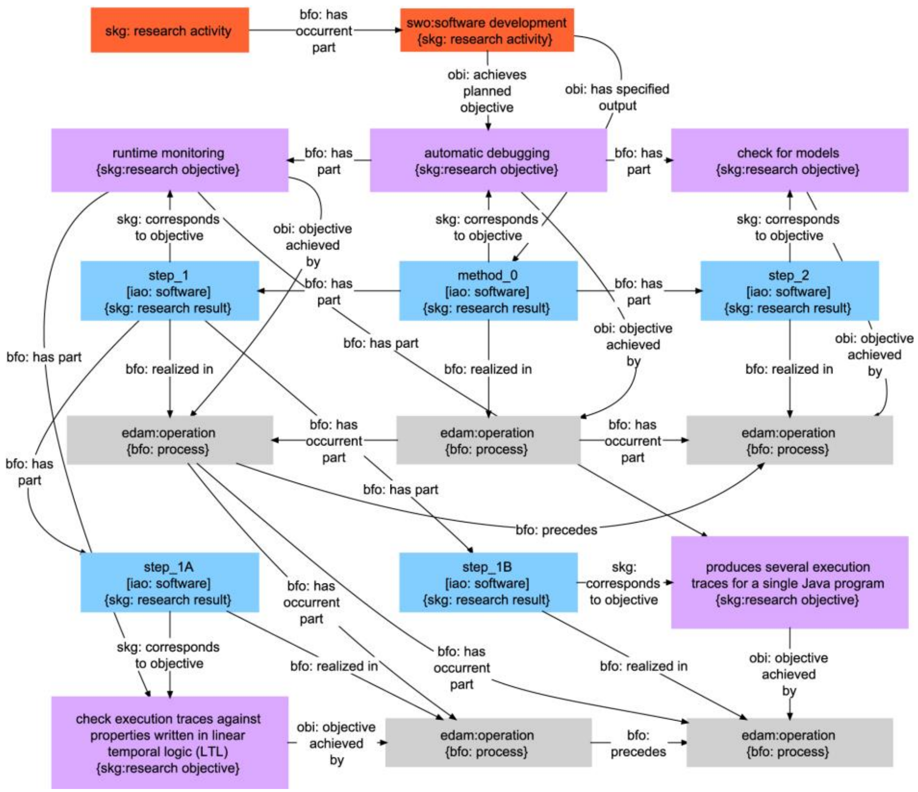
The Research Contribution Model (RCM)

Magnetoencephalography during auditory oddball paradigm



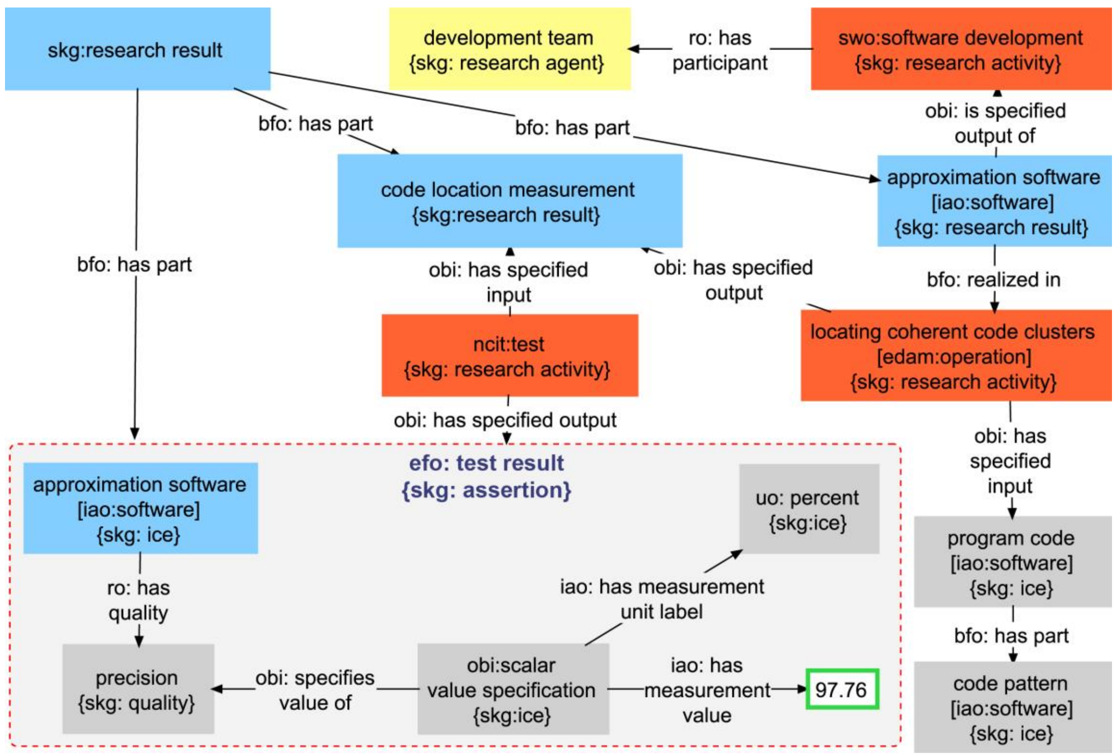
The Research Contribution Model (RCM)

Description of a software development



The Research Contribution Model (RCM)

Measurement of precision



Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

Europe PMC Funders Group
Author Manuscript
J Neurol. Author manuscript; available in PMC 2019 June 11.

Published in final edited form as:
J Neurol. 2015 October 1; 262(10): 2232–2240. doi:10.1007/s00415-015-7832-2.

Voxel based analysis in neuroferritinopathy expands the phenotype and determines radiological correlates of disease severity

MJ Keogh[1], **BS Aribisala**[2], **J He**[2], **E Tulip**[3], **D Butteriss**[4], **C Morris**[5], **G Gorman**[6], **R Horvath**[1], **PF Chinnery**[1,6], and **AM Blamire**[2]

[1]Institute of Genetic Medicine, Newcastle University, Central Parkway, Newcastle upon Tyne, NE1 3BZ

[2]Institute of Cellular Medicine & Newcastle Magnetic Resonance Centre, Newcastle University, NE4 5PL

[3]School of Biomedical Sciences, Newcastle University, NE2 4HH

[4]Department of Neuroradiology, Royal Victoria Infirmary, Newcastle Upon Tyne, NE1 4LP

[5]Medical Toxicology Centre, Wolfson Building, Claremont Place, Newcastle University, Newcastle upon Tyne, NE2 4AA

Example taken from:

Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

PDF Viewer

doi:10.1007/s00415-015-7832-2

Annotation

Neuroferritinopathy is an autosomal dominant adult-onset movement disorder which occurs due to mutations in the ferritin light chain gene (FTL). Extensive iron deposition and cavitation are observed post-mortem in the basal ganglia, but whether more widespread pathological changes occur, and whether they correlate with disease severity is unknown.

3D-T1w and quantitative T2 whole brain MRI scans were performed in 10 clinically symptomatic patients with the 460InsA FTL mutation and 10 age-matched controls. Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed. Clinical assessment using the Unified Dystonia Rating Scale (UDRS) and Unified Huntington's Disease Rating Scale (UHDRS) was undertaken in all patients.

VBM detected significant tissue changes within the substantia nigra, midbrain and dentate together with significant cerebellar atrophy in patients (FWE, $p < 0.05$). Iron deposition in the caudate head and cavitation in the lateral globus pallidus correlated with UDRS score ($p < 0.001$). There were no differences between groups with VBR.

Our data show that progressive iron accumulation in the caudate nucleus, and cavitation of the globus pallidus correlate with disease severity in neuroferritinopathy. We also confirm sub-clinical

Text: *"Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed."*

mark assertion in text

Example taken from:
Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

PDF Viewer

doi:10.1007/s00415-015-7832-2

Annotation

Neuroferritinopathy is an autosomal dominant adult-onset movement disorder which occurs due to mutations in the ferritin light chain gene (FTL). Extensive iron deposition and cavitation are observed post-mortem in the basal ganglia, but whether more widespread pathological changes occur, and whether they correlate with disease severity is unknown.

3D-T1w and quantitative T2 whole brain MRI scans were performed in 10 clinically symptomatic patients with the 460InsA FTL mutation and 10 age-matched controls. Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed. Clinical assessment using the Unified Dystonia Rating Scale (UDRS) and Unified Huntington's Disease Rating Scale (UHDRS) was undertaken in all patients.

VBM detected significant tissue changes within the substantia nigra, midbrain and dentate together with significant cerebellar atrophy in patients (FWE, $p < 0.05$). Iron deposition in the caudate head and cavitation in the lateral globus pallidus correlated with UDRS score ($p < 0.001$). There were no differences between groups with VBR.

Our data show that progressive iron accumulation in the caudate nucleus, and cavitation of the globus pallidus correlate with disease severity in neuroferritinopathy. We also confirm sub-clinical

Data Analysis Method ▼

Text: *"Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed."*

classify the assertion

Example taken from:

Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

PDF Viewer

doi:10.1007/s00415-015-7832-2

Annotation

Neuroferritinopathy is an autosomal dominant adult-onset movement disorder which occurs due to mutations in the ferritin light chain gene (FTL). Extensive iron deposition and cavitation are observed post-mortem in the basal ganglia, but whether more widespread pathological changes occur, and whether they correlate with disease severity is unknown.

3D-T1w and quantitative T2 whole brain MRI scans were performed in 10 clinically symptomatic patients with the 460InsA FTL mutation and 10 age-matched controls. Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed. Clinical assessment using the Unified Dystonia Rating Scale (UDRS) and Unified Huntington's Disease Rating Scale (UHDRS) was undertaken in all patients.

VBM detected significant tissue changes within the substantia nigra, midbrain and dentate together with significant cerebellar atrophy in patients (FWE, $p < 0.05$). Iron deposition in the caudate head and cavitation in the lateral globus pallidus correlated with UDRS score ($p < 0.001$). There were no differences between groups with VBR.

Our data show that progressive iron accumulation in the caudate nucleus, and cavitation of the globus pallidus correlate with disease severity in neuroferritinopathy. We also confirm sub-clinical

Data Analysis Method ▼

Text: *"Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed."*

x ero:morphometry

x chmo:relaxometry

semantic enrichment

Example taken from:

Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

PDF Viewer

doi:10.1007/s00415-015-7832-2

Annotation

Neuroferritinopathy is an autosomal dominant adult-onset movement disorder which occurs due to mutations in the ferritin light chain gene (FTL). Extensive iron deposition and cavitation are observed post-mortem in the basal ganglia, but whether more widespread pathological changes

...ptomatic ...hometry ...ssment ...ng Scale ...te ...n the ...< 0.001). ...of the ...b-clinical

using **templates** for semantifying the assertion

Data Analysis Method ▼

Text: *"Voxel-based morphometry (VBM) and voxel-based relaxometry (VBR) were subsequently performed."*

x ero:morphometry

x chmo:relaxometry

has part:

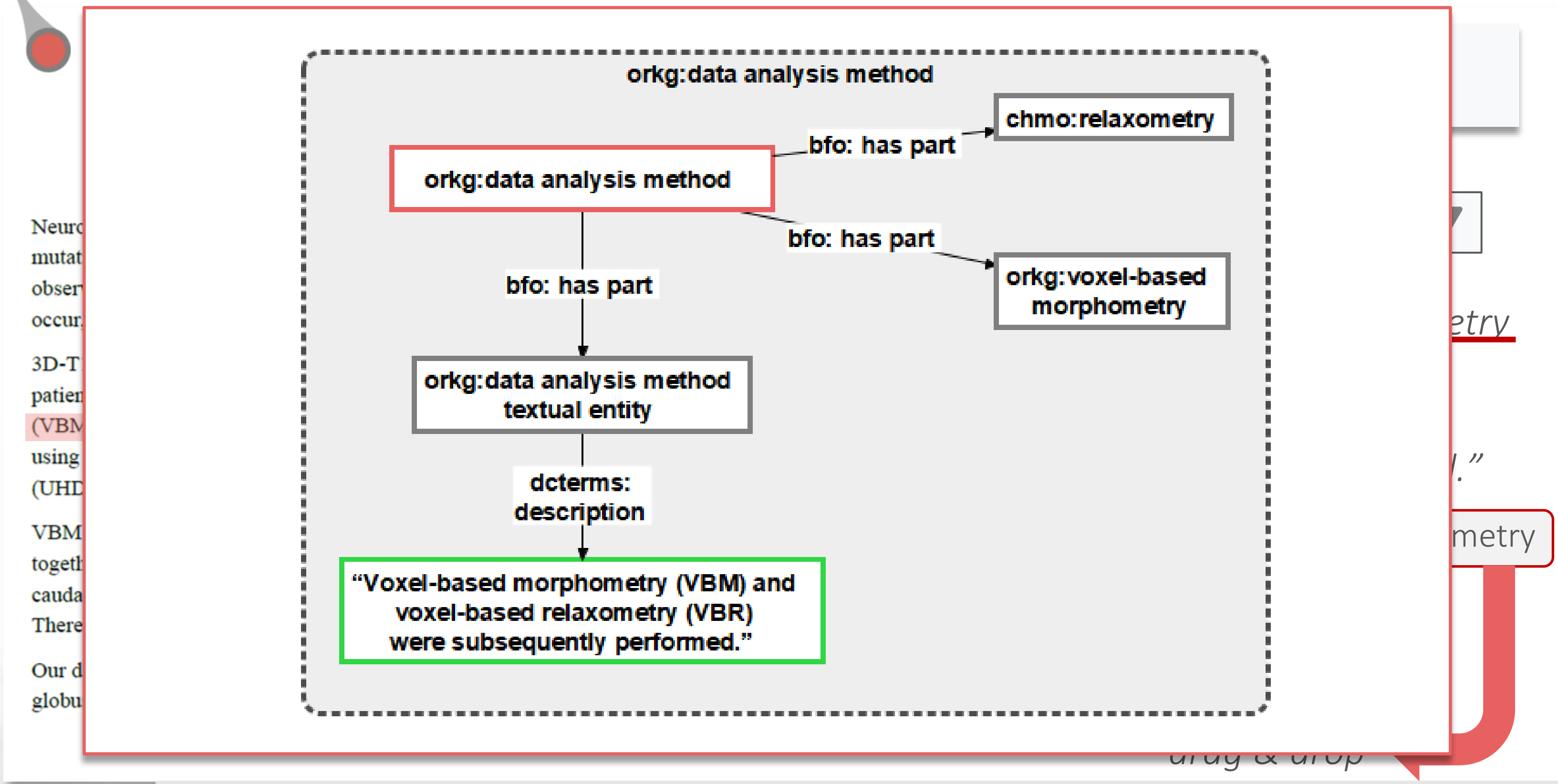
x ero:voxel-based morphometry

drag & drop or automated

...ic papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions



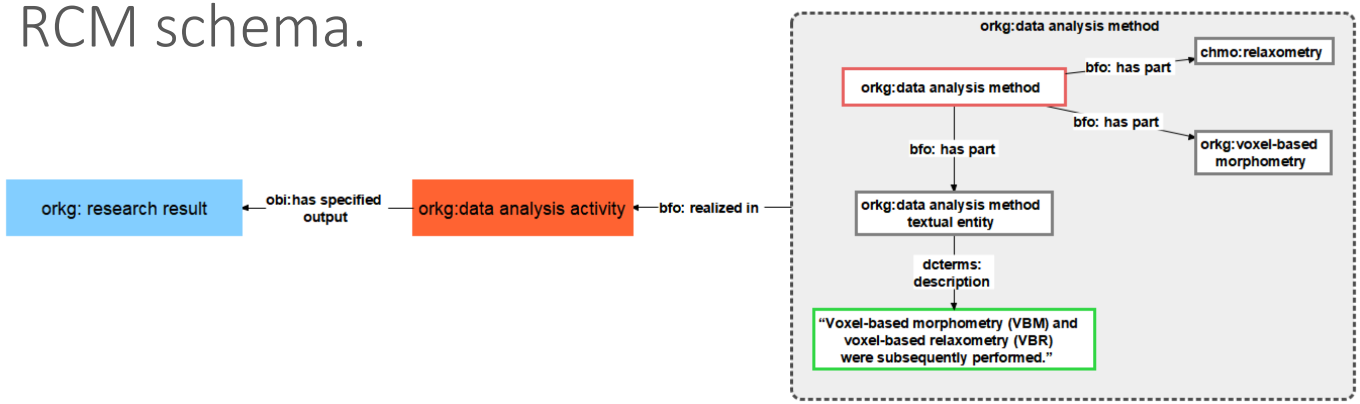
Example taken from:

Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Bottom-up data entry—at the level of individual assertions

This semantified scholarly assertion can be automatically linked to and integrated with the general RCM schema.

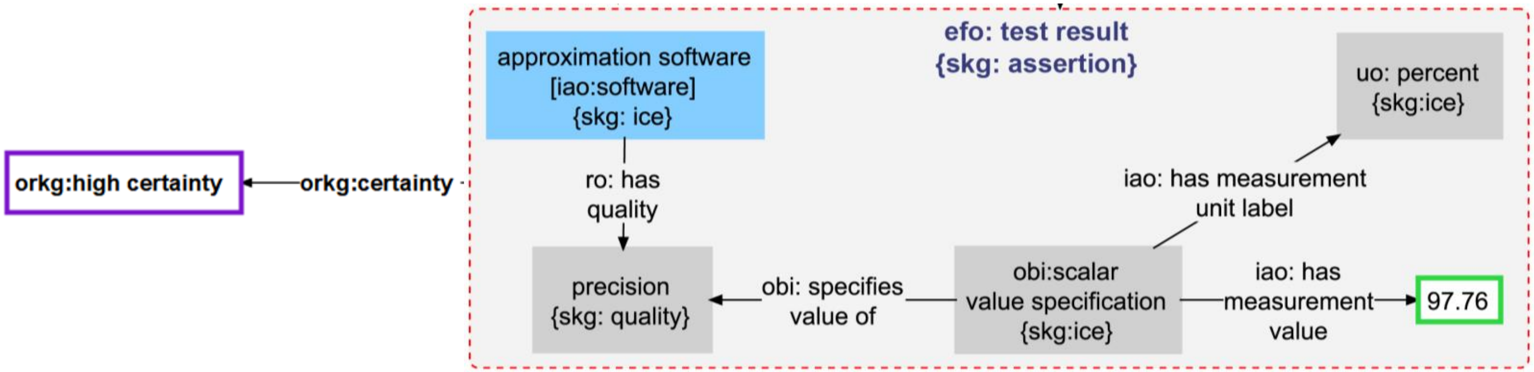


Example taken from:
Udell (2015) Annotating to extract findings from scientific papers. *hypothesis.is*, https://web.hypothesis.is/blog/annotating-to-...

Employing the RCM in a knowledge graph

Use of named graphs allows statements about statements

Because each assertion is organized in its own named graph, we can e.g. specify its degree of certainty, ...

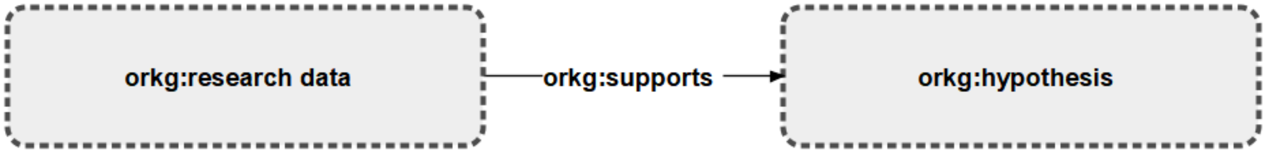


Prieto *et al.* (2019) Data-driven classification of the certainty of scholarly assertions. *PeerJ* 8, e8871.

Employing the RCM in a knowledge graph

Use of named graphs allows statements about statements

…or relate it to another assertion (supporting, contradicting, neutral), …



Prieto *et al.* (2019) Data-driven classification of the certainty of scholarly assertions. *PeerJ* 8, e8871.

Employing the RCM in a knowledge graph

Use of named graphs allows statements about statements

… or provide personal opinions about it, for instance the data provider wants to point out that they conceived a logical inconsistency, or a wrong statistical method being applied.

Prieto *et al.* (2019) Data-driven classification of the certainty of scholarly assertions. *PeerJ* 8, e8871.

Knowledge Graph Cell (KGC)

Knowledge Graph Cell (KGC)

Goal

Provide a means to have the contents of scholarly publications in a machine-actionable format but at the same time present the contents in human-readable UI-pages.

Knowledge Graph Cell (KGC)

Basic idea

The representation of the contents of a scholarly publication in a knowledge graph usually comprises several UI-pages.

Knowledge Graph Cell (KGC)

Basic idea

We partition the overall data graph into integrated sub-graphs, with each sub-graph comprising information that should be represented on one such UI-page.

Knowledge Graph Cell (KGC)

Basic idea

We specify a KGC for each sub-graph so that each UI-page of a scholarly knowledge graph has a specific KGC assigned to it.

Knowledge Graph Cell (KGC)

Basic idea

A KGC can be instantiated through user input, resulting in the generation of

1) a sub-graph that is based on the graph-template specified in its KGCs and
2) a corresponding UI-page.

Knowledge Graph Cell (KGC)

Basic idea

The user input for a given KGC can trigger the instantiation of another KGC. This way, KGCs can relate to one another and we can specify the entire organization of the data of a knowledge graph application as a set of interlinked KGCs.

Knowledge Graph Cell (KGC)

Basic idea

A middleware reads the information contained in the KGCs and in the overall data graph and executes it, thereby generating the knowledge graph application.

Knowledge Graph Cell (KGC)

Organization of an RDF-based KGC into sets of Quad Templates

An RDF-based KGC is organized into Quads:

Subject (URI) Property (URI) Object (URI/Literal) :NG(URI),

with each Quad being described in a KGC as a Quad Template (QT).

Knowledge Graph Cell (KGC)

Organization of an RDF-based KGC into sets of Quad Templates

An RDF-b

Subject (U   al)   :NG(URI),

with eac   Quad
Templat

described in the

forthcoming paper

Knowledge Graph Cell (KGC)

Organization of a Neo4j-based KGC

Currently, we are developing a Neo4j-based implementation of the KGC idea.

Case Study: COVID-19 Data

Case description

We want to model information about a population that has been infected with the SARS-CoV-2 Coronavirus. We want to cover Basic Reproduction Number (R0) and Case Fatality Rate (CFR) measurements, each with 95% Confidence Intervals. All this information should be represented on a single UI-page.

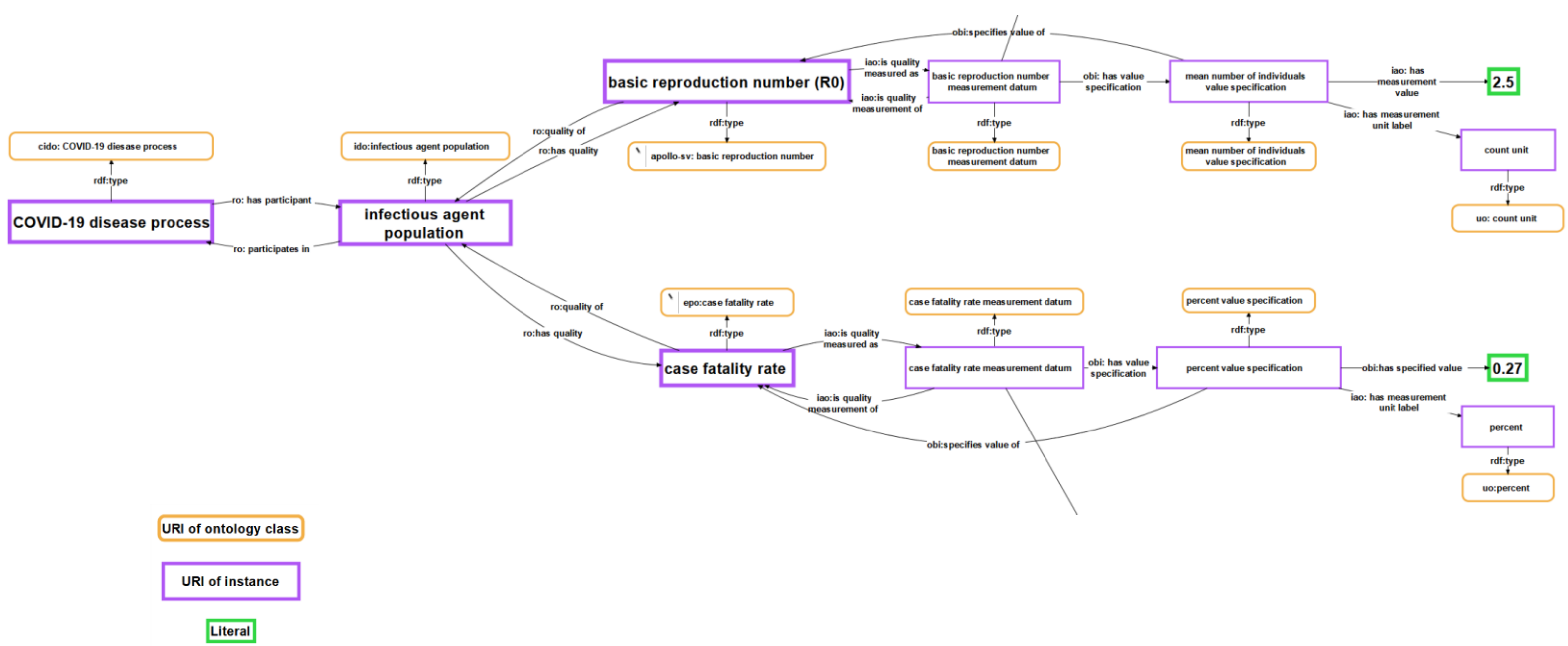
Case Study: COVID-19 Data

Infectious Agent Population KGC

We want to model this information with a general Infectious Agent Population KGC, which can be reused for other types of infections.

Case Study: COVID-19 Data

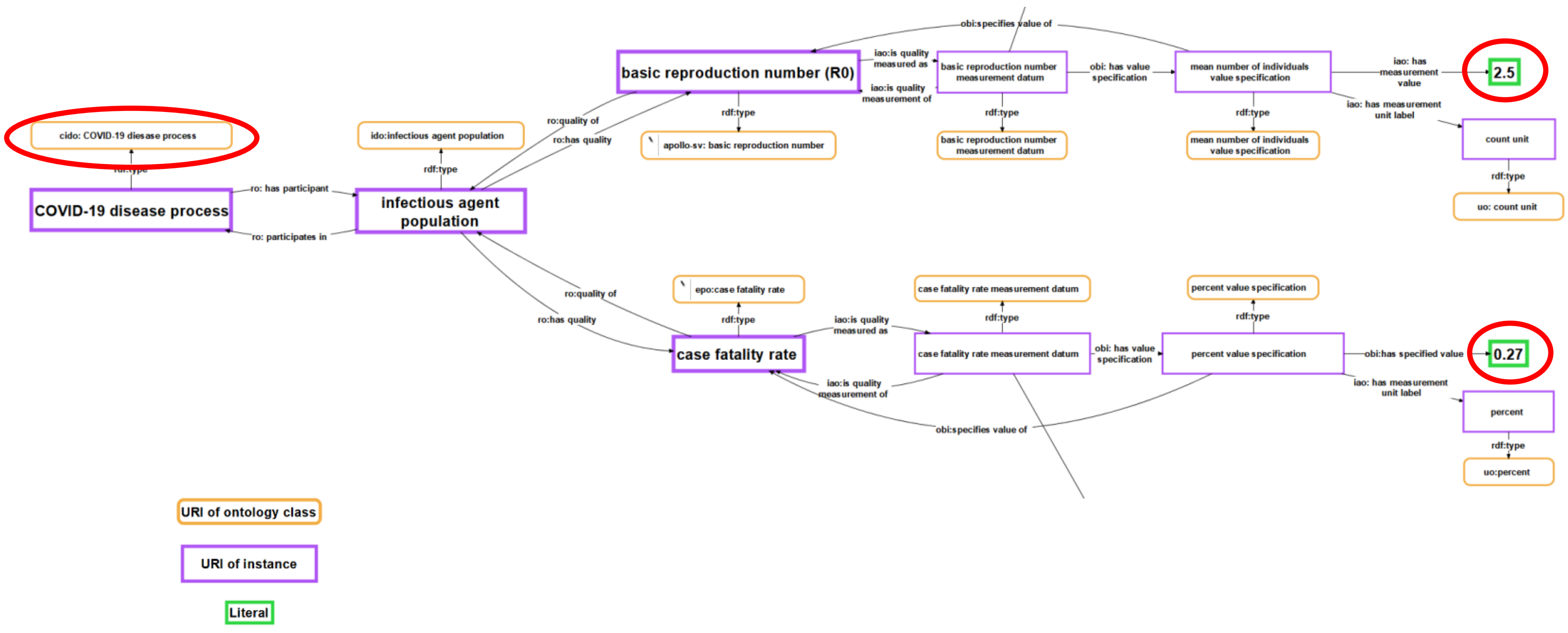
Underlying OWL graph pattern



Case Study: COVID-19 Data

Infectious Agent Population KGC

user input

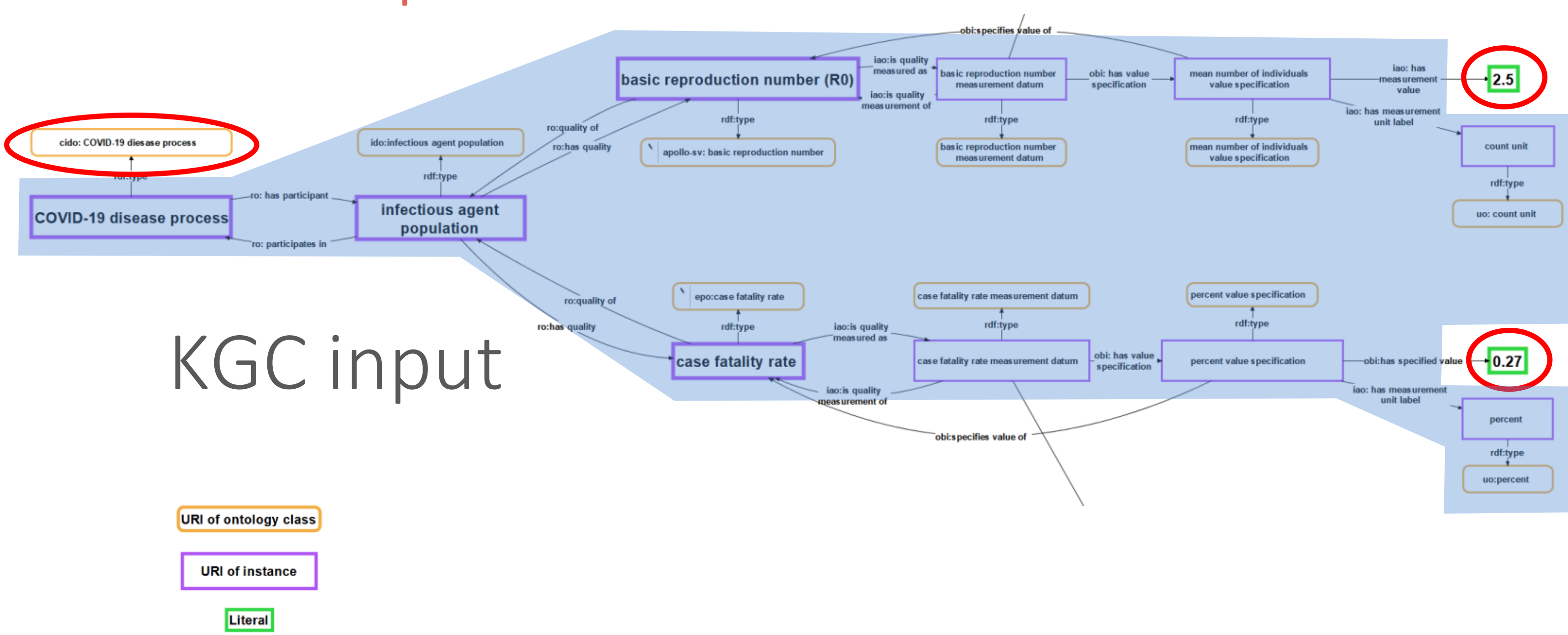


Case Study: COVID-19 Data

Infectious Agent Population KGC

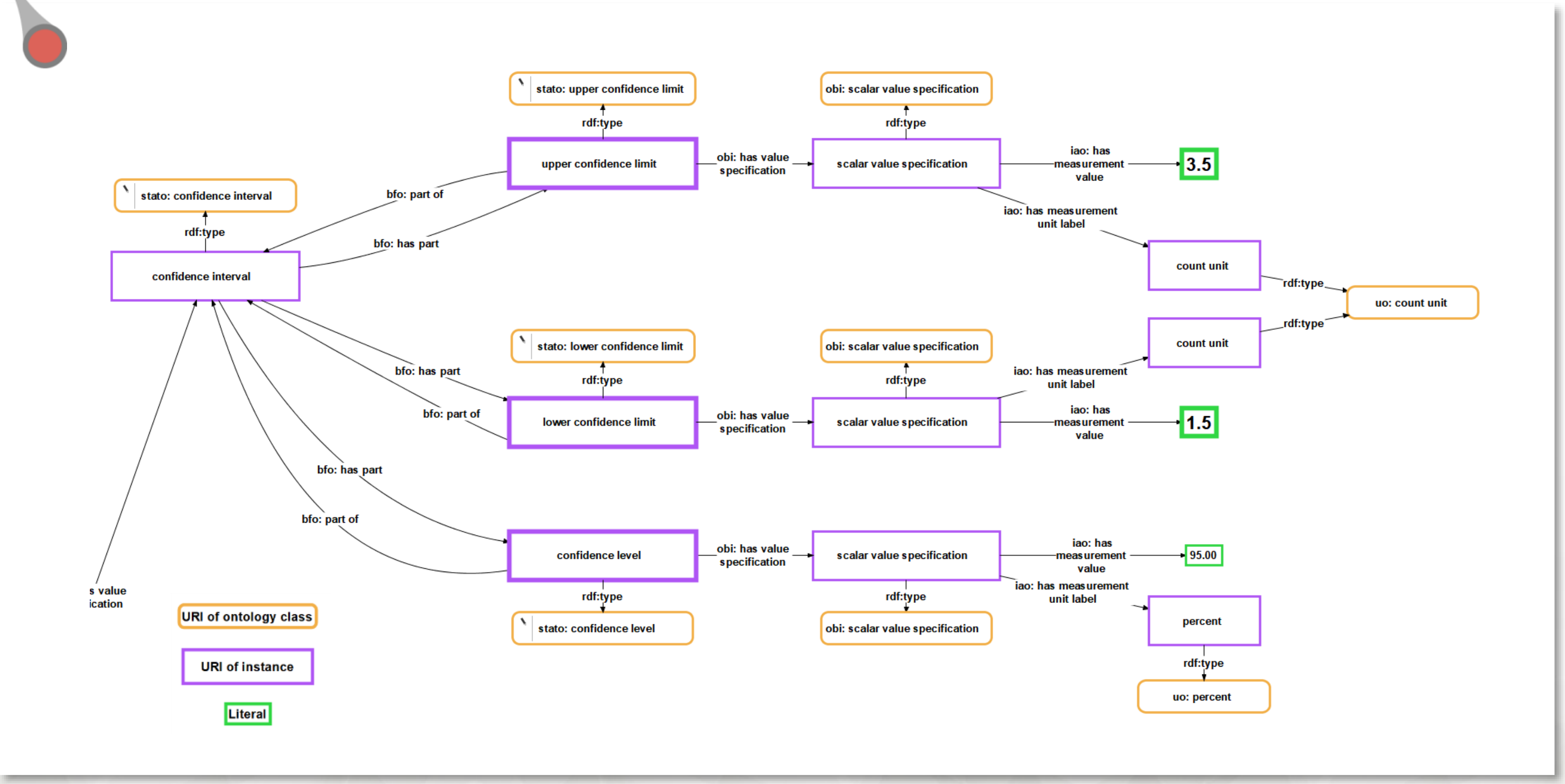
user input

KGC input



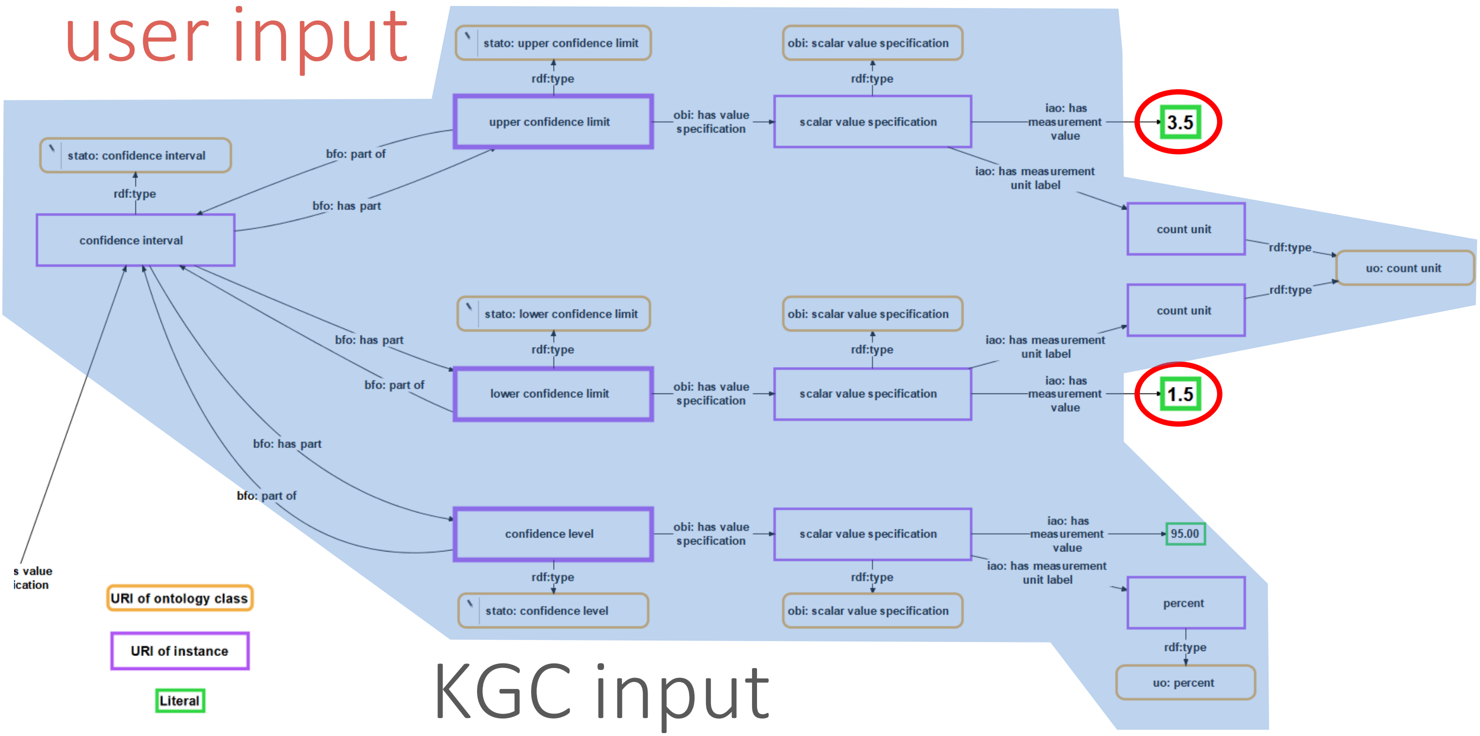
Case Study: COVID-19 Data

Underlying OWL graph pattern



Case Study: COVID-19 Data

Infectious Agent Population KGC



Case Study: COVID-19 Data

Infectious Agent Population KGC

Next, we described the Infectious Agent Population KGC as a set of Cypher scripts.

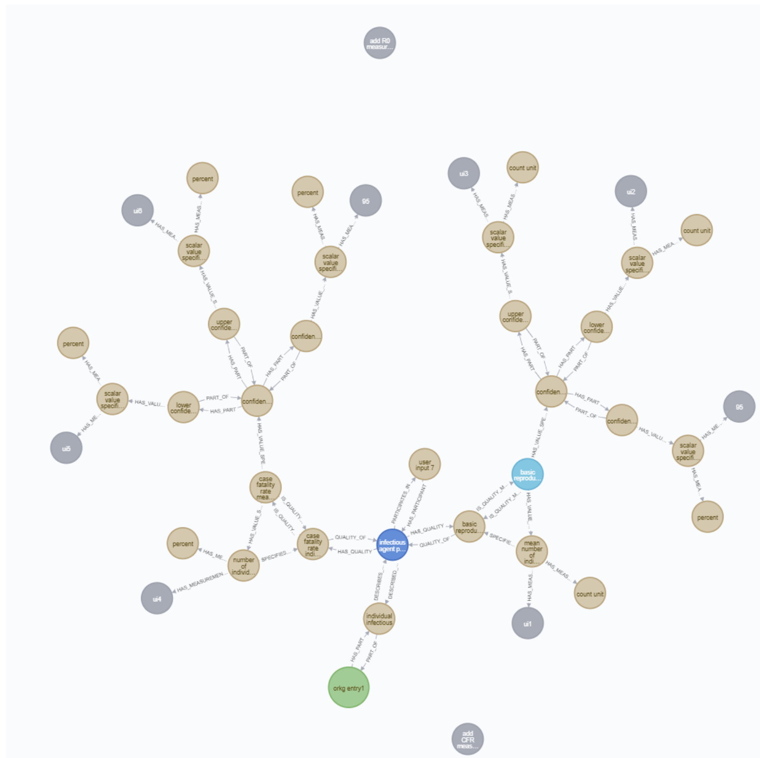
Case Study: COVID-19 Data

Infectious Agent Population KGC

Instantiation Cypher script

Case Study: COVID-19 Data

Infectious Agent Population KGC



Case Study: COVID-19 Data

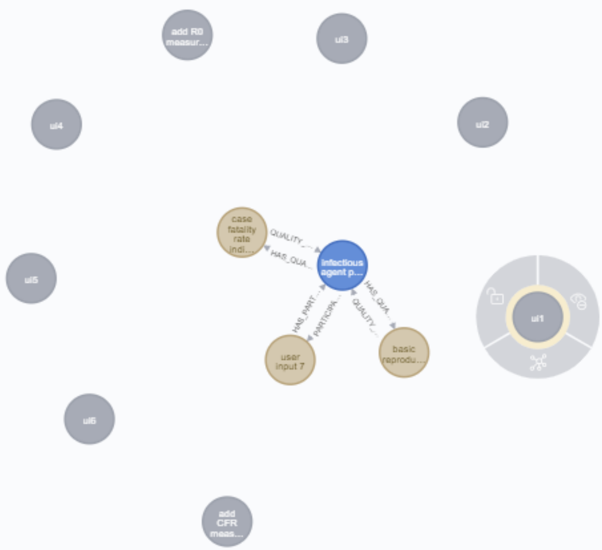
Infectious Agent Population KGC

UI-page Cypher script

```
// search for all nodes belonging to ORKG entry1 kgc data graph data_graph_entry1_kgc_1
that contain gui information

MATCH (n)
        WHERE n.data_OrkgEntry="entry1" AND
        n.kgc_kgc_data_graph="data_graph_entry1_kgc_1" AND
        n.gui_information="true"
RETURN n
```

Case Study: COVID-19 Data

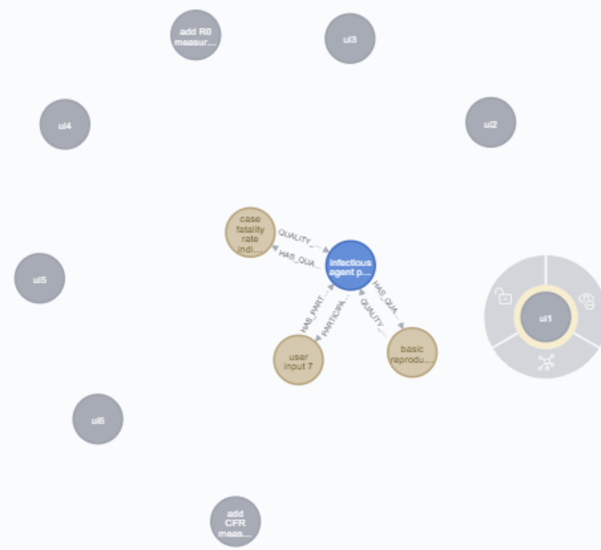
Infectious Agent Population KGC



Literal Value **<id>:** 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

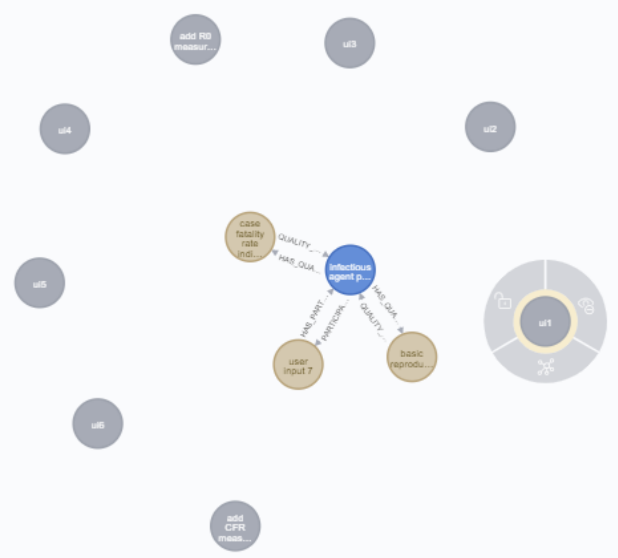


scholarly publication ID

Literal Value **<id>:** 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

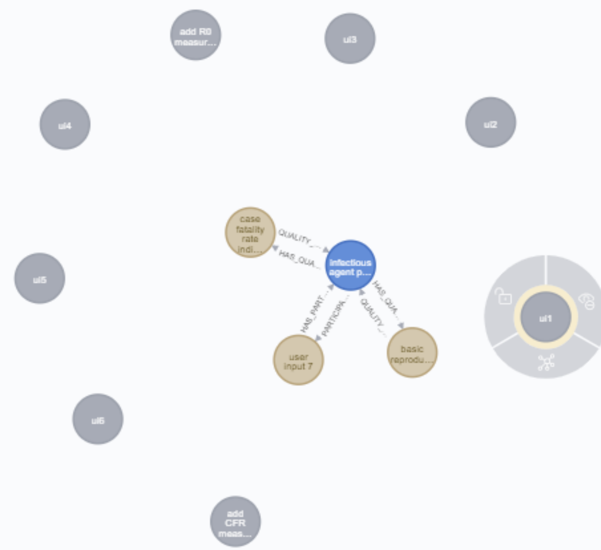


named graph information / UI-page ID

Literal Value **<id>:** 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

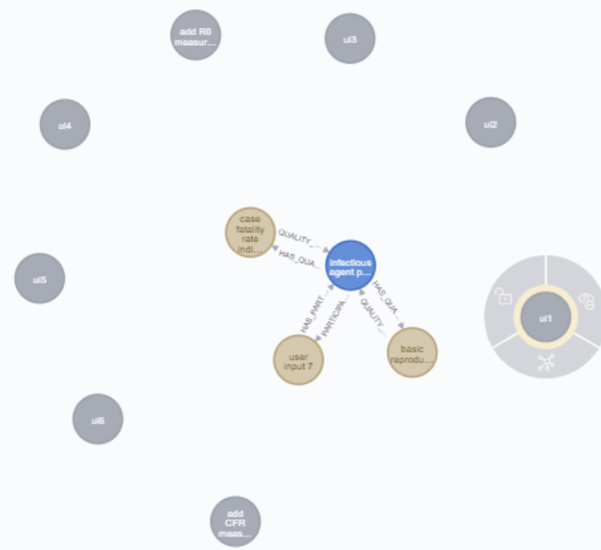


data type specification: float

Literal Value **<id>:** 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

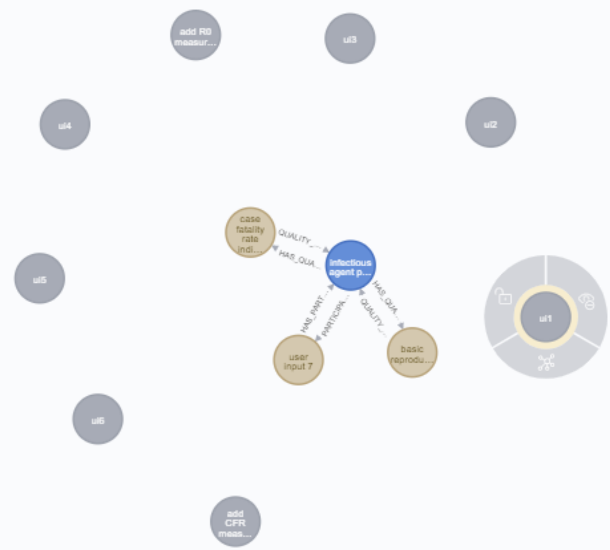


UI-page information

Literal Value <id>: 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction numbe **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

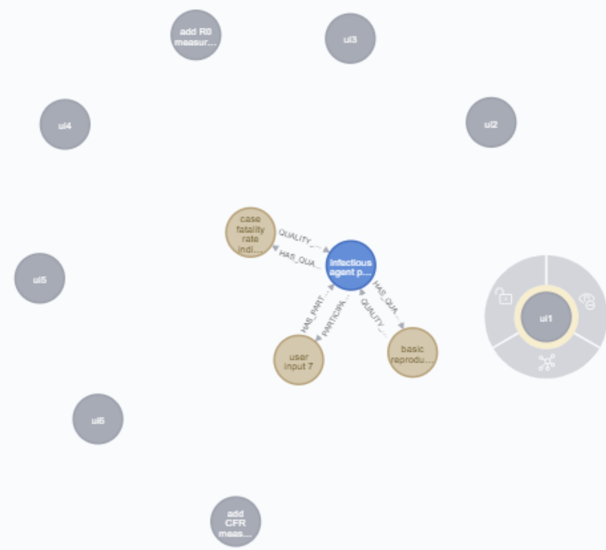


edit Cypher Script

Literal Value <id>: 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC



value: yet unspecified

Literal **Value** **<id>:** 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** ui1

Case Study: COVID-19 Data

Infectious Agent Population KGC

update Cypher script for R0 mean value

```
// update the values for the R0 mean value [ui1] of entry1 and kgc data graph
data_graph_entry1_kgc_1

:param [{entry, dataGraph, guiGroupOrder, guiSectionOrder, ui1, guiOrder1, guiOrder2,
guiOrder3}] => {RETURN 0 AS guiSectionOrder, "entry1" AS entry, 1 AS guiOrder1, 2 AS
guiOrder2, 3 AS guiOrder3, "data_graph_entry1_kgc_1" AS dataGraph, 0 AS
guiGroupOrder, "2.5" AS ui1};

MATCH (n {data_OrkgEntry:$entry, gui_section_order:$guiSectionOrder,
          gui_group_order:$guiGroupOrder, gui_order:$guiOrder1,
          kgc_kgc_data_graph:$dataGraph})
SET n.value=$ui1
```

Case Study: COVID-19 Data

Infectious Agent Population KGC

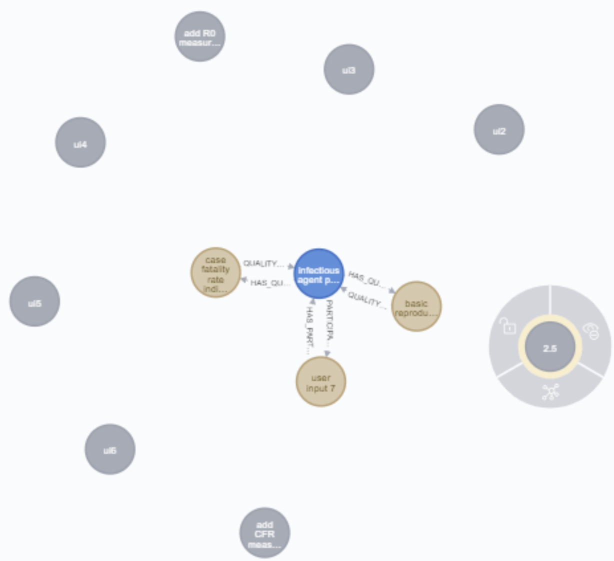
update Cypher script for R0 mean value

```
// update the values for the R0 mean value [ui1] of entry1 and kgc data graph data_graph_entry1_kgc_1

:param [{entry, dataGraph, guiGroupOrder, guiSectionOrder, ui1, guiOrder1, guiOrder2, guiOrder3}] => {RETURN 0 AS guiSectionOrder, "entry1" AS entry, 1 AS guiOrder1, 2 AS guiOrder2, 3 AS guiOrder3, "data_graph_entry1_kgc_1" AS dataGraph, 0 AS guiGroupOrder, "2.5" AS ui1};

MATCH

SET n.value=$ui1
```

provided by frontend/middleware and user-input

Case Study: COVID-19 Data

Infectious Agent Population KGC

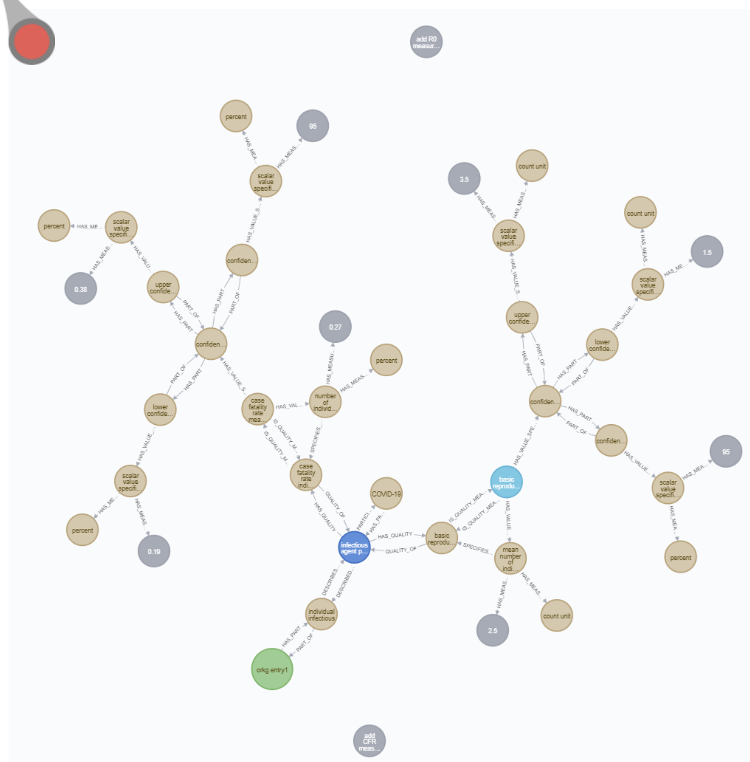


Value: 2.5

Literal Value <id>: 36 **data_OrkgEntry:** entry1 **gui_data_type:** xsd:float **gui_edit_cypher_address:** infectiousAgentPopulationKGC_R0_UserInput1_URI **gui_element:** inputfieldX **gui_group_order:** 0 **gui_information:** true **gui_information_text:** mean value **gui_label2:** key:value **gui_order:** 1 **gui_section_order:** 0 **gui_tooltip:** The mean basic reproduction number **kgc_kgc_data_graph:** data_graph_entry1_kgc_1 **name:** user input 1 **value:** 2.5

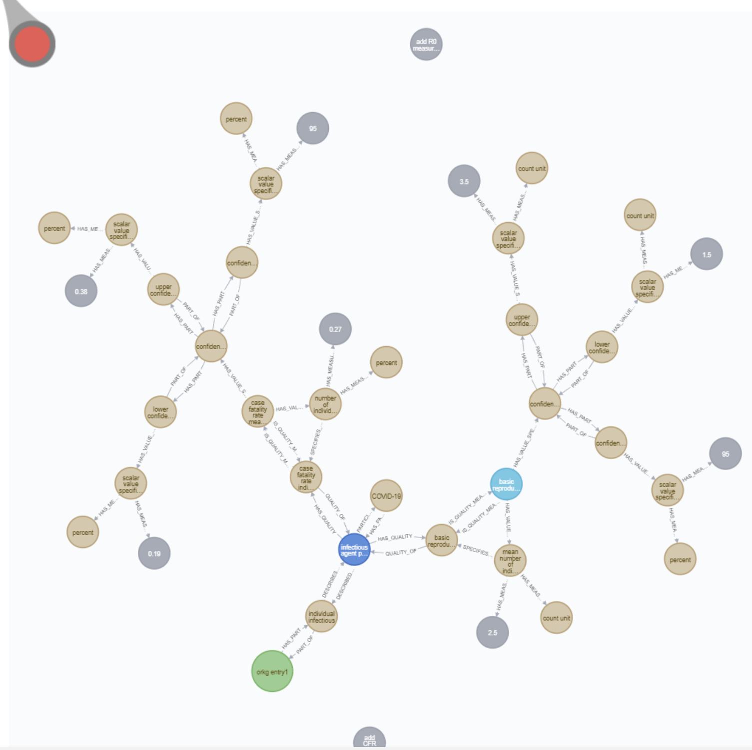
Case Study: COVID-19 Data

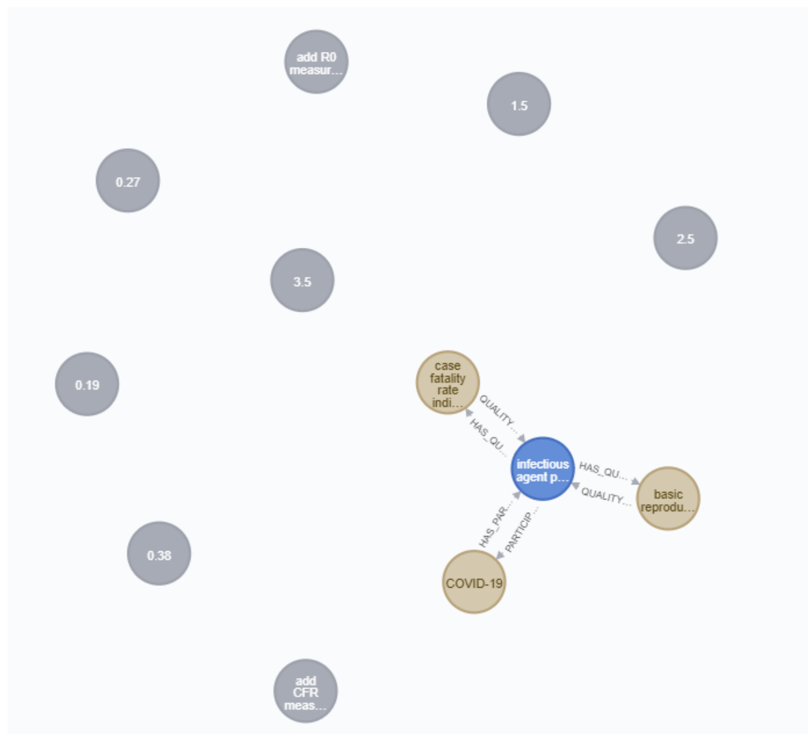
Infectious Agent Population KGC



Case Study: COVID-19 Data

Infectious Agent Population KGC



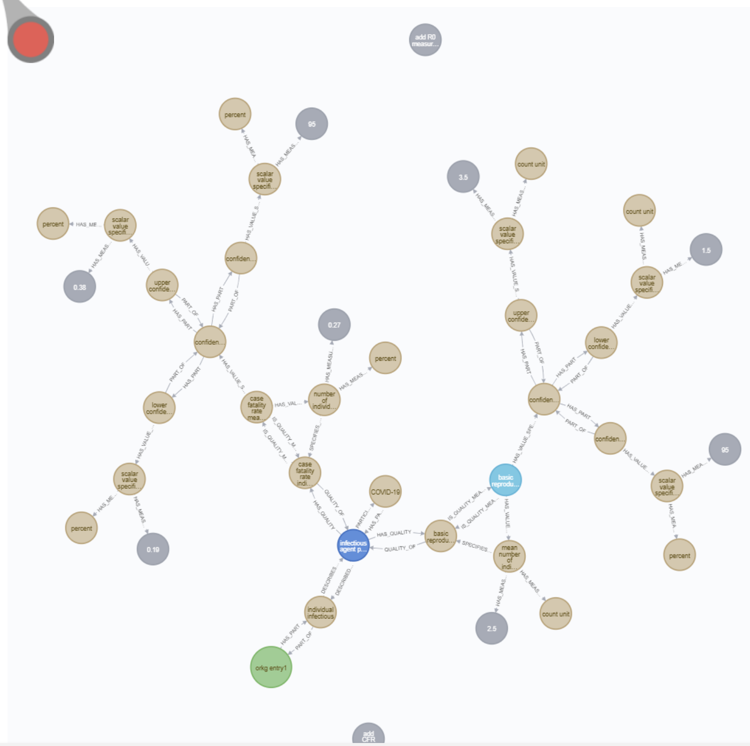


machine-actionable graph

human-readable UI-page

Case Study: COVID-19 Data

Infectious Agent Population KGC



Infectious Agent Population

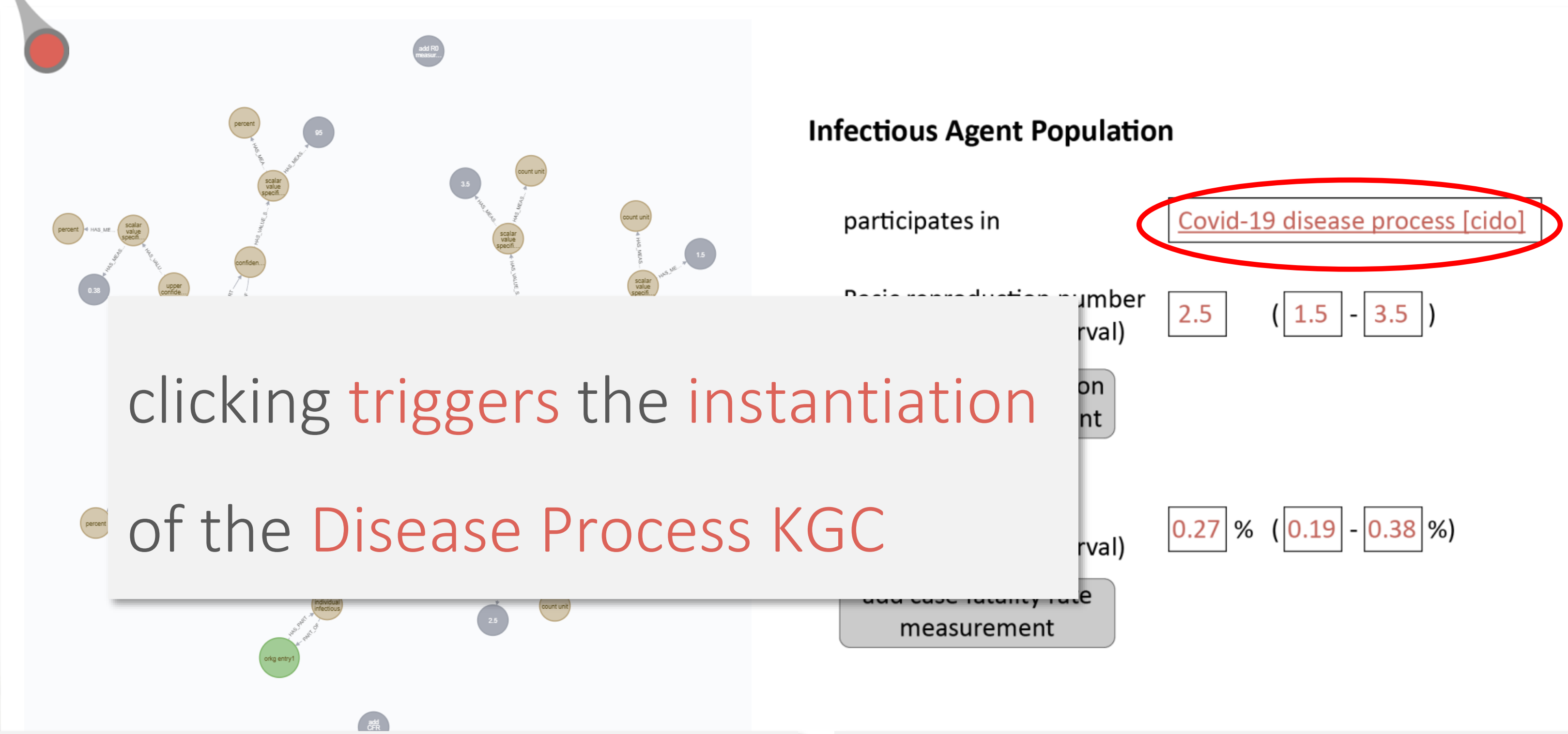
participates in	Covid-19 disease process [cido]	
Basic reproduction number (95% confidence interval)	2.5	(1.5 - 3.5)
add basic reproduction number measurement		
Case fatality rate (95% confidence interval)	0.27 %	(0.19 - 0.38 %)
add case fatality rate measurement		

machine-actionable graph

human-readable UI-page

Case Study: COVID-19 Data

Infectious Agent Population KGC



clicking triggers the instantiation of the Disease Process KGC

machine-actionable graph

human-readable UI-page

Benefits of using Knowledge Graph Cell (KGC)

Uniform modeling scheme

KGCs can be used for implementing a uniform modeling scheme and a specific data model (e.g. RCM) in a knowledge graph application, with all user-input being enforced to comply with the model. This results in semantically consistent and FAIR contents without the users having to be experts in semantics and understanding graphs.

Benefits of using Knowledge Graph Cell (KGC)

Data consistency

KGCs enforce a certain degree of data consistency through (i) input-control that structures and restricts user input and (ii) by employing KGCs in tools and procedures for checking the contents of the knowledge graph for consistency without having to rely on DL-based reasoning.

Benefits of using Knowledge Graph Cell (KGC)

Querying

With predefined SPARQL/Cypher queries, they provide means to interact with the data graph at a fine-grained level without requiring users to understand the query languages.

Benefits of using Knowledge Graph Cell (KGC)

Human-readability

KGCs filters information contained in the data graph to what is relevant to a human reader and makes this information available to the frontend.

Benefits of using Knowledge Graph Cell (KGC)

Agility, flexibility

A knowledge graph application can be developed and specified as a set of interlinked KGCs. The middleware runs the application. New KGCs can be added and existing ones be modified.

RCM and KGCs combined in a Knowledge Graph for Scholarly Publications

FAIRness

When combining the KGC idea with the RCM data model for scholarly publications, we can build a Knowledge Graph for Scholarly Publications that provides data and metadata that are compliant with the FAIR Guiding Principles.

Towards Representing Research Contributions in Scholarly Knowledge Graphs Using Knowledge Graph Cells

L. Vogt, J. D'Souza, M. Stocker, S. Auer