

24/10/2021



DaMaLOS 2021 @ISWC

THE SURVEY ONTOLOGY: PACKAGING SURVEY RESEARCH AS RESEARCH OBJECTS

Mario Scrocca, Damiano Scandolari, Gloria
Re Calegari, Ilaria Baroni and Irene Celino
name.surname@cefriel.com

INNOVATE > GROW > REPEAT

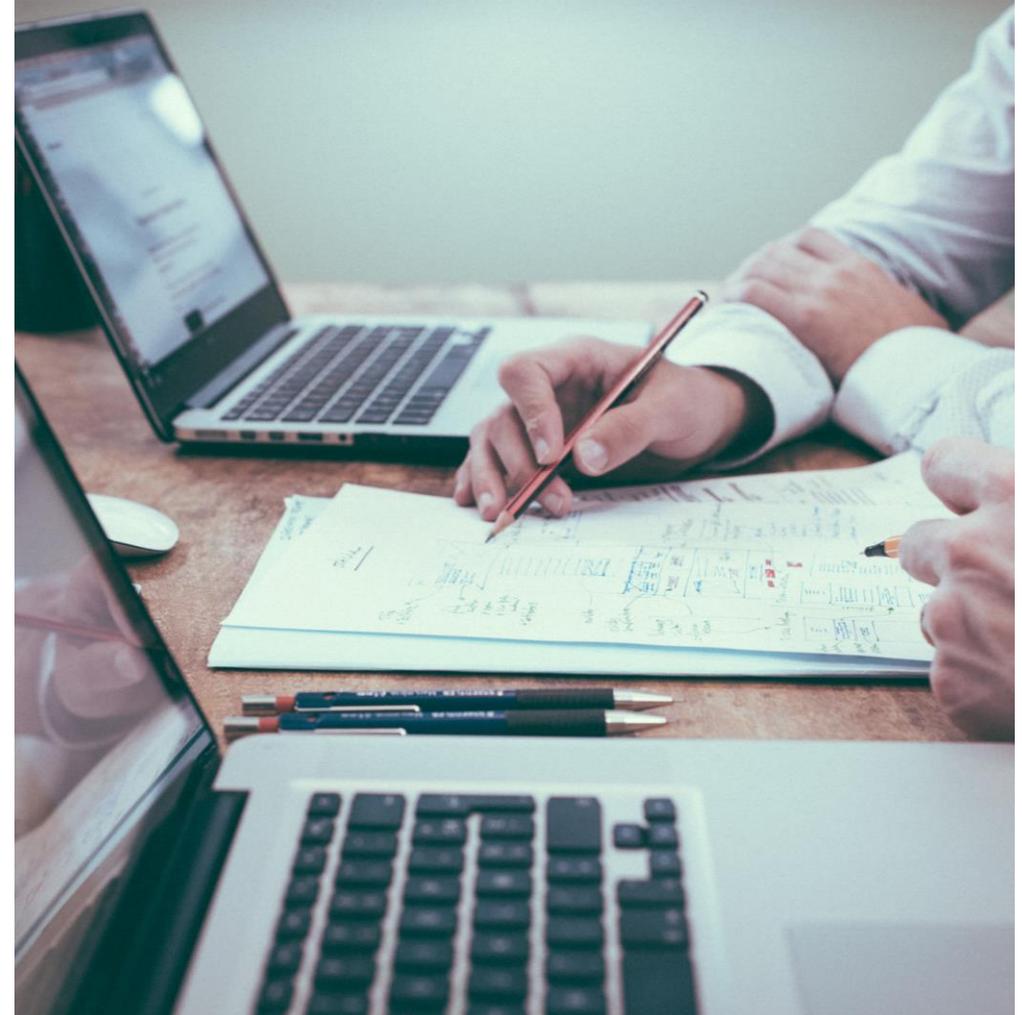
Cefriel
POLITECNICO DI MILANO

ISWC 2021



Sharing survey research

- **Questionnaire design** is a well explored discipline defining different and robust methodologies
- **Open issue:** how to simplify and support data, methods and result sharing
- The **Survey Ontology**, that we designed to empower our **CONEY** toolkit for conversational surveys, **embraces the research object principles**, and defines an **open vocabulary** to represent, annotate, and share a representation of the questionnaire structure and the gathered responses of a survey.





Coney

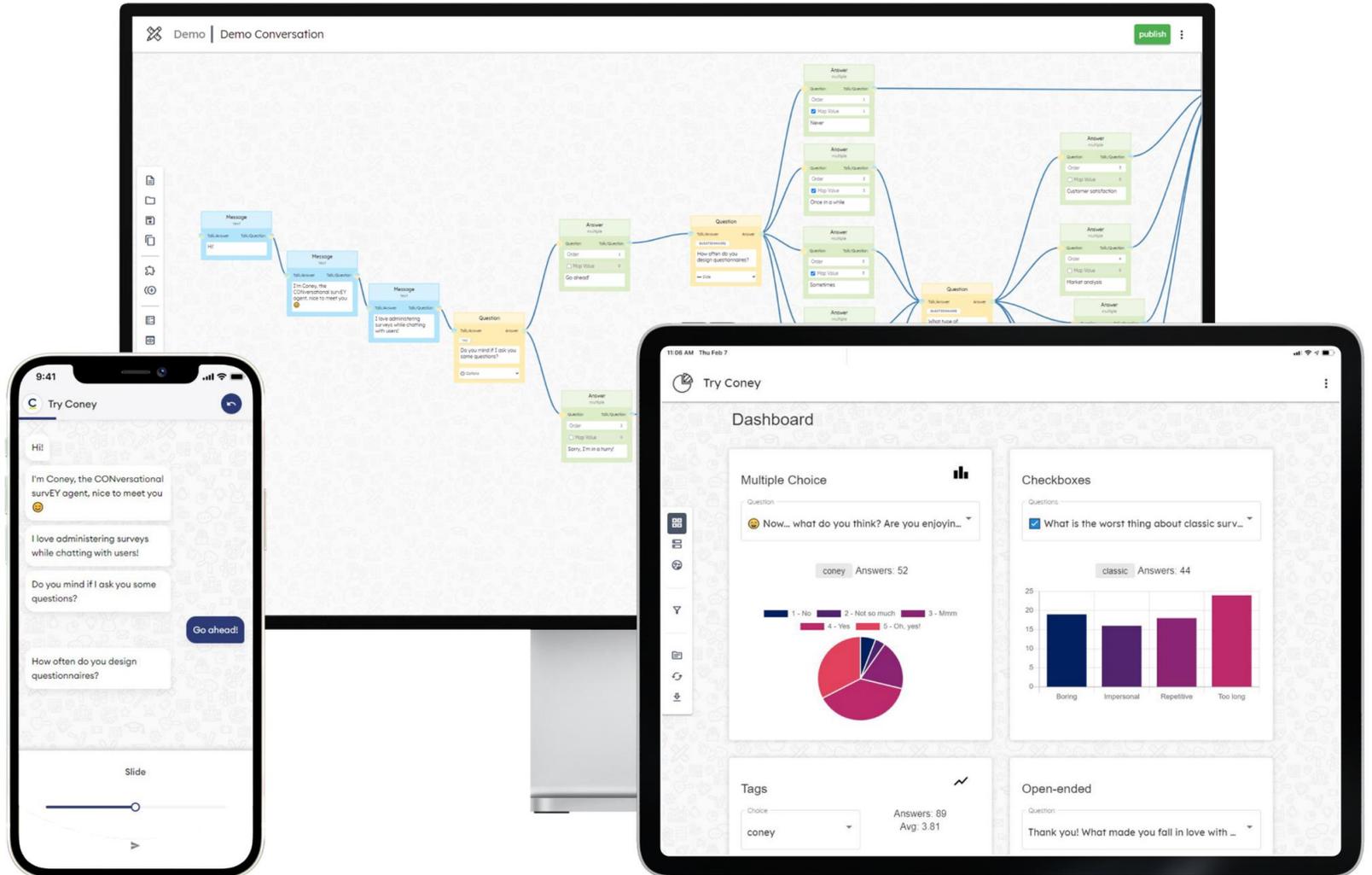
Coney is a toolkit for designing, administering and analyzing results of **conversational surveys**.

Coney enhances user experience and user engagement thanks to:

- a chat interaction pattern for **interactive "storytelling"** (colloquial and multimedia content)
- conversation flow with **multiple branches** according to the respondent's answers

Coney's data model is based on the **Survey Ontology**, and the tool allows **exporting** both the survey structure and the collected answers exploiting the ontology.

Survey design



Survey administration

Survey results

More information: coney.cefriel.com
Try Coney: bit.ly/try-coney

A practical case: Studying Motivation of Citizen Scientists

- The H2020 ACTION project aimed at studying motivation and investigating the factors influencing people's participation in citizen science projects
- The **survey design** process requires different steps for the definition of an effective questionnaire:
 - define the **research question**
 - define **investigated factors**
 - **formulate questions**
 - **set-up** the questionnaire
 - **test the survey** with some user
- As a best practice, we analysed **related work** on the topic trying to rely on **existing questionnaires and surveys** used to evaluate the level of motivations of participants in Citizen Science projects.
- **Several issues arise** for a survey designer trying to access related survey research. We will discuss the related challenges addressed by the **Survey Ontology** to foster survey research packaging and sharing.



Challenges

C1 Make the survey structure available as structured data: to avoid the risk of “burying” the survey semantics in documents (e.g., PDF), we aim at providing away to export a survey as a dataset by itself.

		Not at all 1	2	3	4	Very much 5	Not relevant
1	I want to learn						
2	I am interested in the topic of this project						
3	I am interested in science and/or technology						
4	I participate out of curiosity						
5	I want to improve my skills						
6	It's an opportunity to explore new things						
7	It enables me to be creative						
8	I want to do something new						

Image from “Questionnaire to measure citizen scientist’s motivation” (Levontin, Gilad and Chako)
https://cs-eu.net/sites/default/files/media/2019/03/for%20researchers_0.pdf

Challenges

C1 Make the survey structure available as structured data: to avoid the risk of “burying” the survey semantics in documents (e.g., PDF), we aim at providing away to export a survey as a dataset by itself.

C2 Annotate questions with the respective investigated variables: to make it possible to analyse survey results more easily, as well as to enable the comparison between different studies.

		Not at all 1	2	3	4	Very much 5	Not relevant
1	I want to learn						
2	I am interested in the topic of this project						
3	I am interested in science and/or technology						
4	I participate out of curiosity						
5	I want to improve my skills						
6	It's an opportunity to explore new things						
7	It enables me to be creative						
8	I want to do something new						

Category	Item #	item
Self-direction	1	I want to learn
	2	I am interested in the topic of this project
	3	I am interested in science and/or technology
	4	I participate out of curiosity
	5	I want to improve my skills
	6	It's an opportunity to explore new things
	7	It enables me to be creative
Stimulation	8	I want to do something new
	9	I want to break away from my routine
	10	I strive to challenge myself
	11	This activity is related to another hobby I have

Image from “Questionnaire to measure citizen scientist’s motivation” (Levontin, Gilad and Chako)
https://cs-eu.net/sites/default/files/media/2019/03/for%20researchers_0.pdf

Challenges

C1 Make the survey structure available as structured data: to avoid the risk of “burying” the survey semantics in documents (e.g., PDF), we aim at providing away to export a survey as a dataset by itself.

C2 Annotate questions with the respective investigated variables: to make it possible to analyse survey results more easily, as well as to enable the comparison between different studies.

C3 Annotate answers with their numerical coding: to ease the result analysis, we aim to allow the survey designer to annotate also the questions’ pre-defined answers with their numerical value for subsequent computation of mean, median, variance, etc.

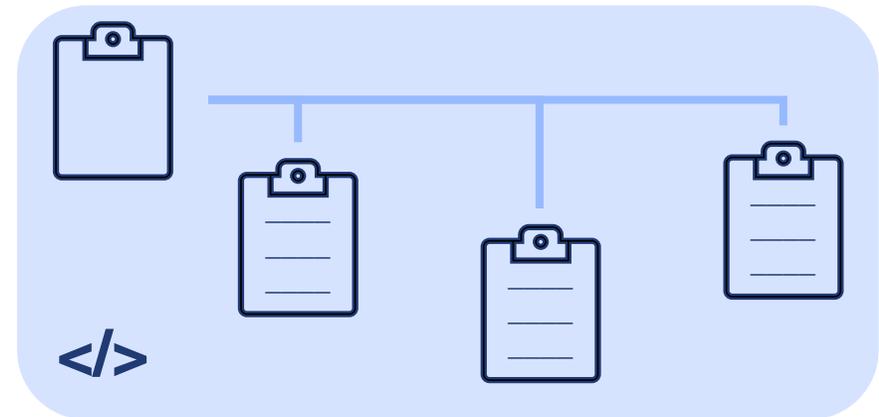
		Not at all 1	2	3	4	Very much 5	Not relevant
1	I want to learn						
2	I am interested in the topic of this project						
3	I am interested in science and/or technology						
4	I participate out of curiosity						
5	I want to improve my skills						
6	It's an opportunity to explore new things						
7	It enables me to be creative						
8	I want to do something new						

Category	Item #	item
Self-direction	1	I want to learn
	2	I am interested in the topic of this project
	3	I am interested in science and/or technology
	4	I participate out of curiosity
	5	I want to improve my skills
	6	It's an opportunity to explore new things
	7	It enables me to be creative
Stimulation	8	I want to do something new
	9	I want to break away from my routine
	10	I strive to challenge myself
	11	This activity is related to another hobby I have

Image from “Questionnaire to measure citizen scientist’s motivation” (Levontin, Gilad and Chako)
https://cs-eu.net/sites/default/files/media/2019/03/for%20researchers_0.pdf

Challenges

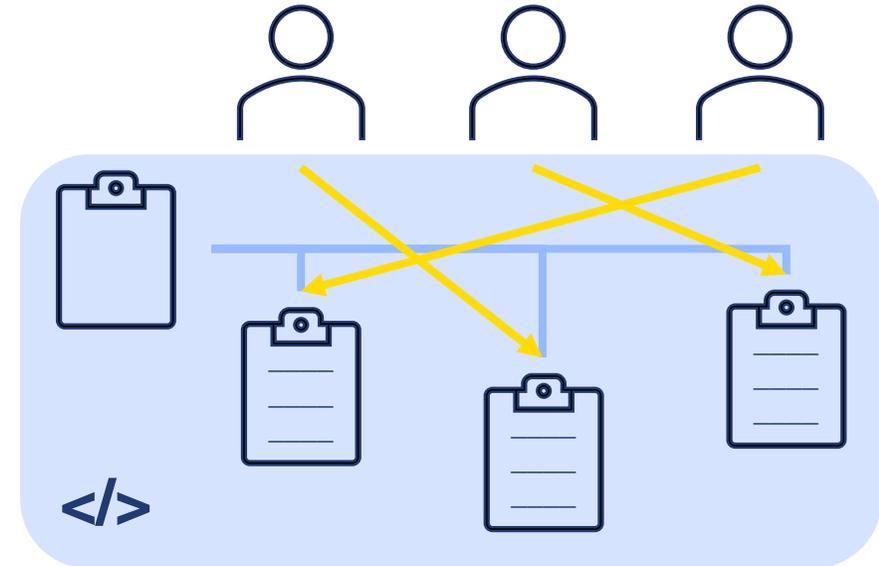
C4 Make the collected answers available as structure data: to facilitate the analysis, we aim to allow for result export by employing the same data model (cf. C1); this allows crosslinking between the survey structure and its results, as well as between different compiling campaigns of the same survey



Challenges

C4 Make the collected answers available as structure data: to facilitate the analysis, we aim to allow for result export by employing the same data model (cf. C1); this allows crosslinking between the survey structure and its results, as well as between different compiling campaigns of the same survey

C5 Keep provenance of answers: to track the link between respondents and their answers, we aim at using provenance; this also helps in cross-study assessment, if the respondents are uniquely identified.

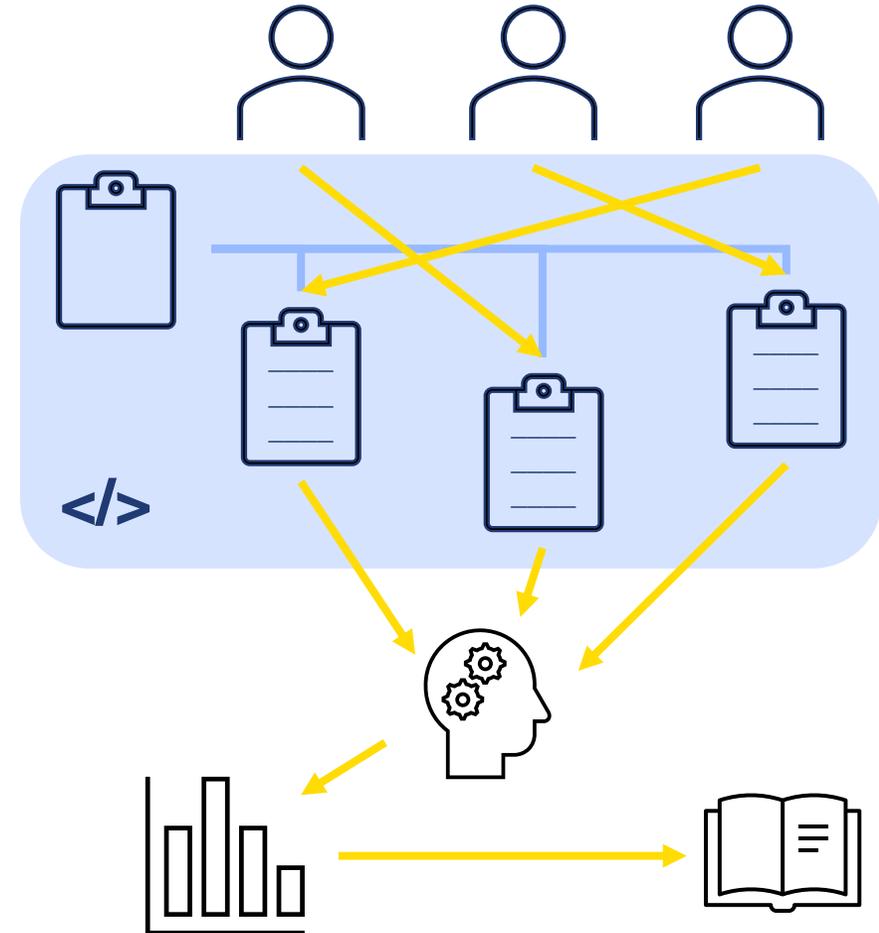


Challenges

C4 Make the collected answers available as structure data: to facilitate the analysis, we aim to allow for result export by employing the same data model (cf. C1); this allows crosslinking between the survey structure and its results, as well as between different compiling campaigns of the same survey

C5 Keep provenance of answers: to track the link between respondents and their answers, we aim at using provenance; this also helps in cross-study assessment, if the respondents are uniquely identified.

C6 Share the survey methodology: to foster repeatability and reproducibility of research, we aim at facilitating to share not only questions and collected answers, but also the scientific method behind it, like the hypothesis for correlation, causality and other interplay between the investigated variables, or the actual analysis processes and techniques



The Survey Ontology

Following the **Linked Open Terms** methodology¹, we identified use cases, user stories and inputs from a domain analysis to define requirements in the ontology engineering process.

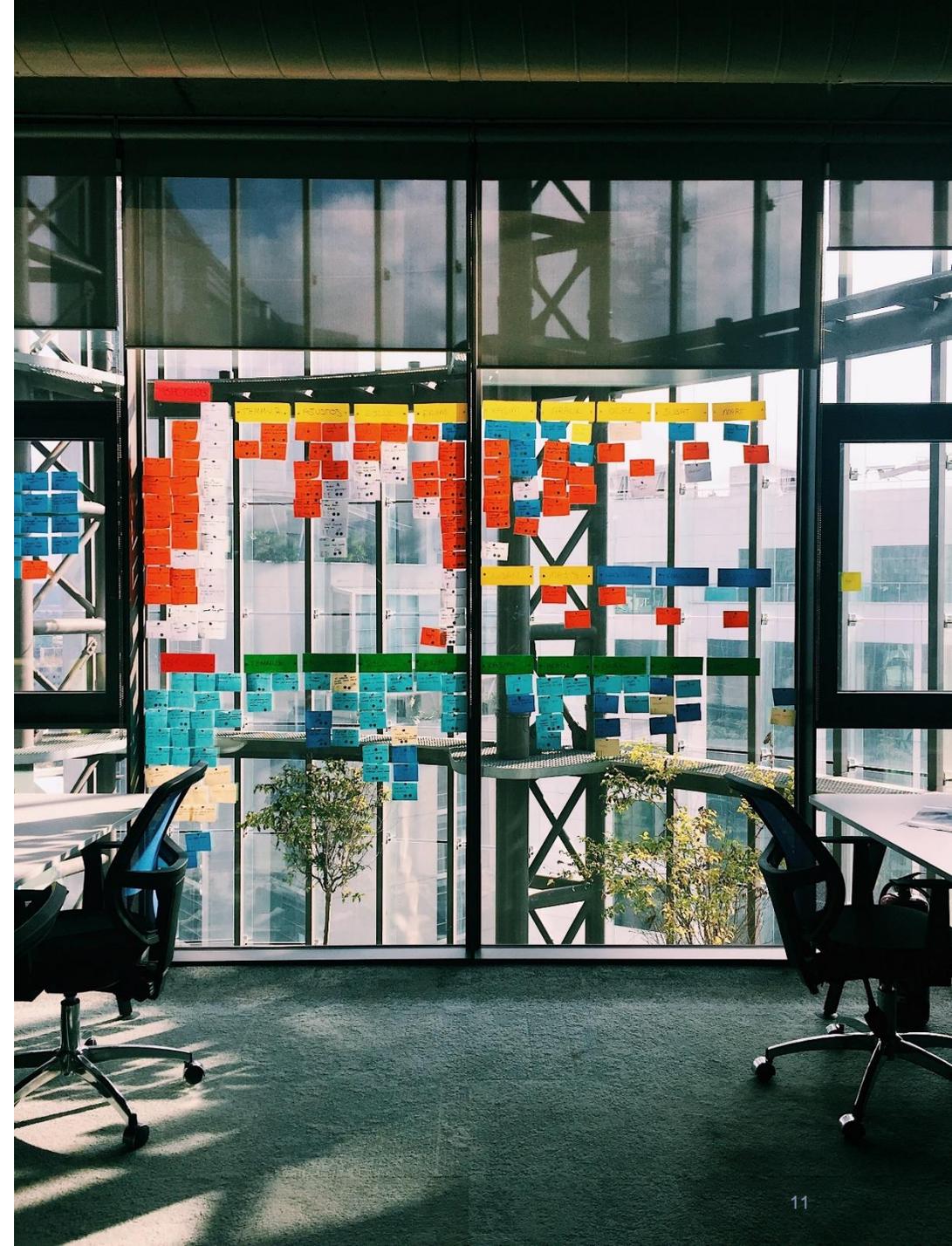
- A set of use cases and user stories associated with the presented challenges is available in the Wiki on GitHub
<https://github.com/cefriel/survey-ontology/wiki/Use-Cases>
- The domain analysis considered: (i) data models of existing survey tools, (ii) already defined ontologies in the same domain (e.g., DDI-RDF), and (iii) the experience/feedback gained in designing the CONEY tool for conversational surveys

Ontology published at: <https://w3id.org/survey-ontology> (**sur:**)

The repository (<https://github.com/cefriel/survey-ontology>):

- hosts the ontology, the documentation, the validation report and a set of SHACL shapes defined to validate knowledge graphs modelled using the ontology;
- is used for the ontology maintenance activities (e.g., issues, additional requirements, etc.)

¹<https://lot.linkeddata.es/>



Modelling Survey Research as a Workflow-centric Research Object

The *Research Object Suite of Ontologies* [1] focuses on the principles of identity, aggregation and provenance annotations, defining a set of workflow-centric ontologies to represent research objects.

The **Survey Ontology** extends the proposed approach by interpreting:

- a complete **survey research study** as a **research object**,
- the **survey procedure** as a **scientific workflow** (*wfdesc* module), and
- the **survey's collected answers** as **provenance traces** of its execution (*wfprov* module).

A **survey research object** aggregates two main resources identified for our ontology: (i) a **survey procedure**, describing the survey structure, and (ii) a **survey dataset**, containing collected answers.

Moreover, a **survey research object** can **aggregate any other additional resource**, e.g., representing study hypotheses, investigated variables, models produced from the result analysis, related publications, etc.

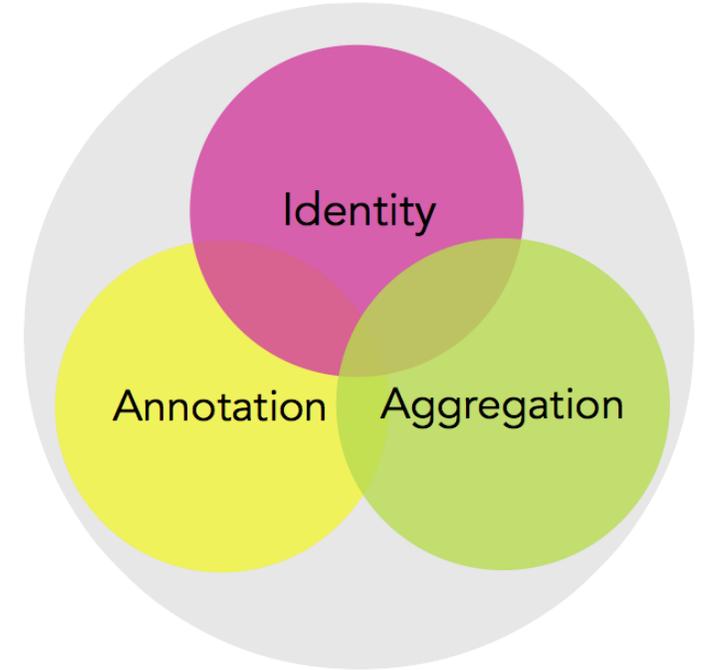
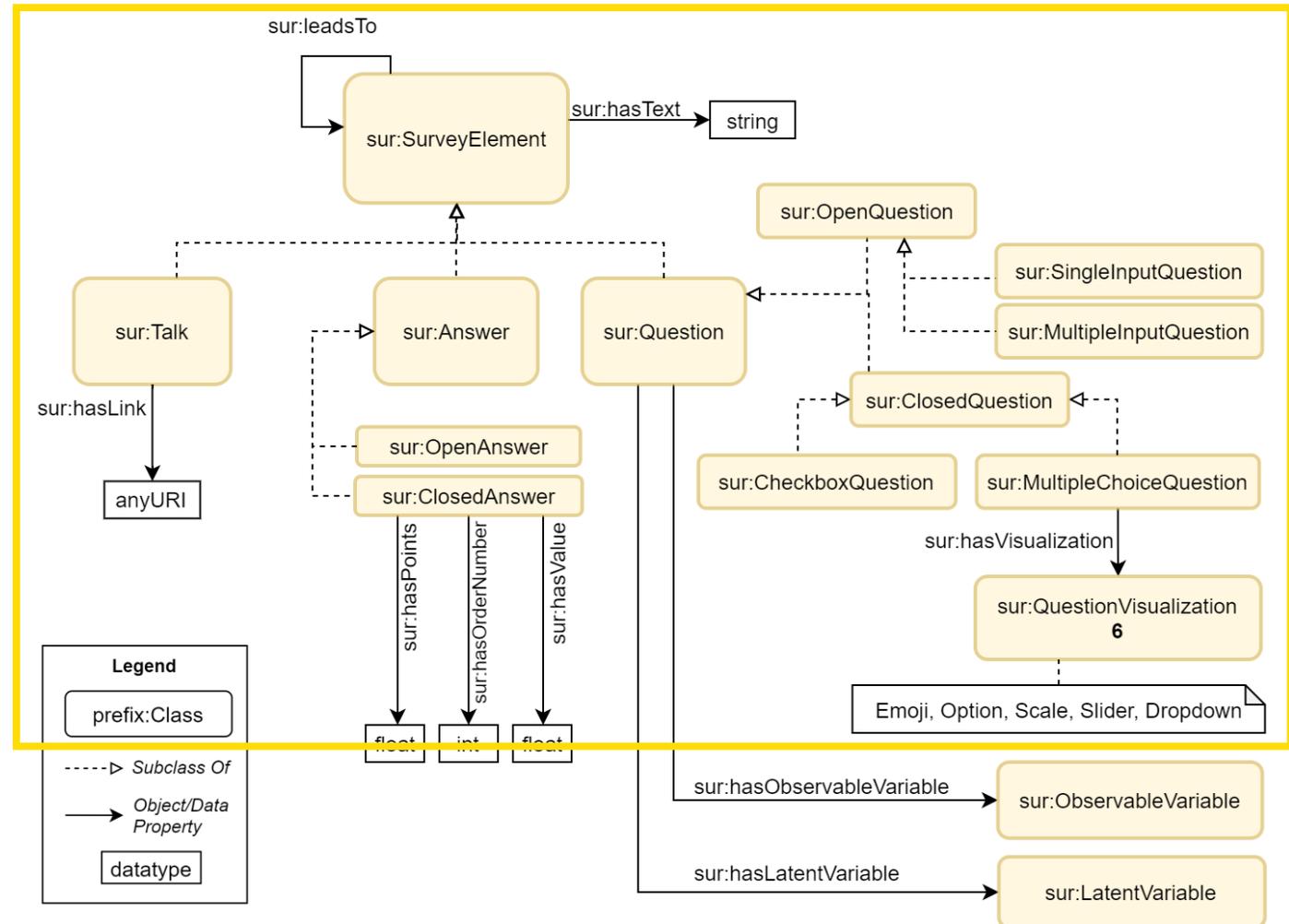


Image rights <https://www.researchobject.org/>

[1] Belhajjame, K., et al. *Using a suite of ontologies for preserving workflow-centric research objects*. <https://doi.org/10.1016/j.websem.2015.01.003>

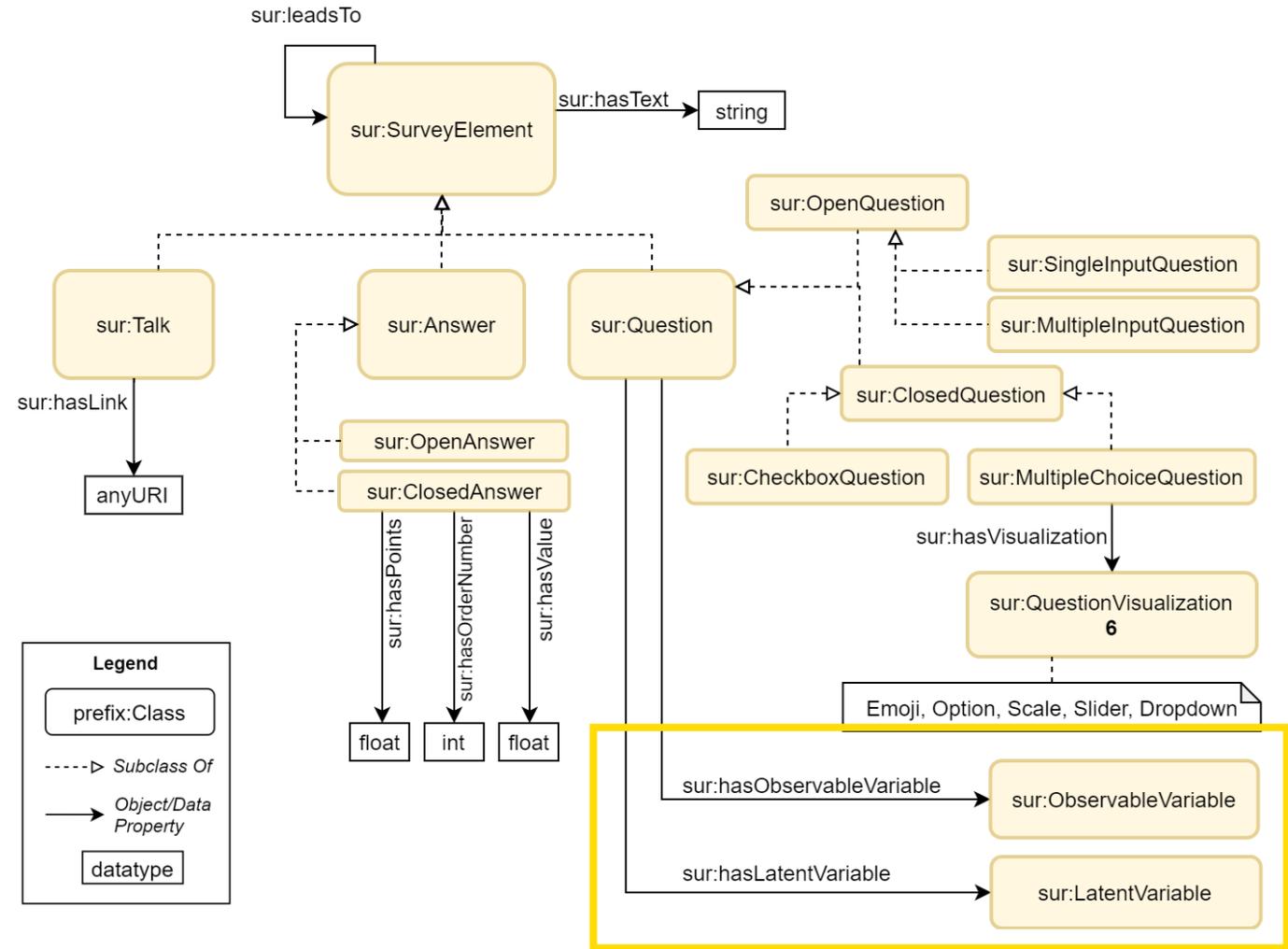
Survey Ontology

- C1 Make the survey structure available as structured data



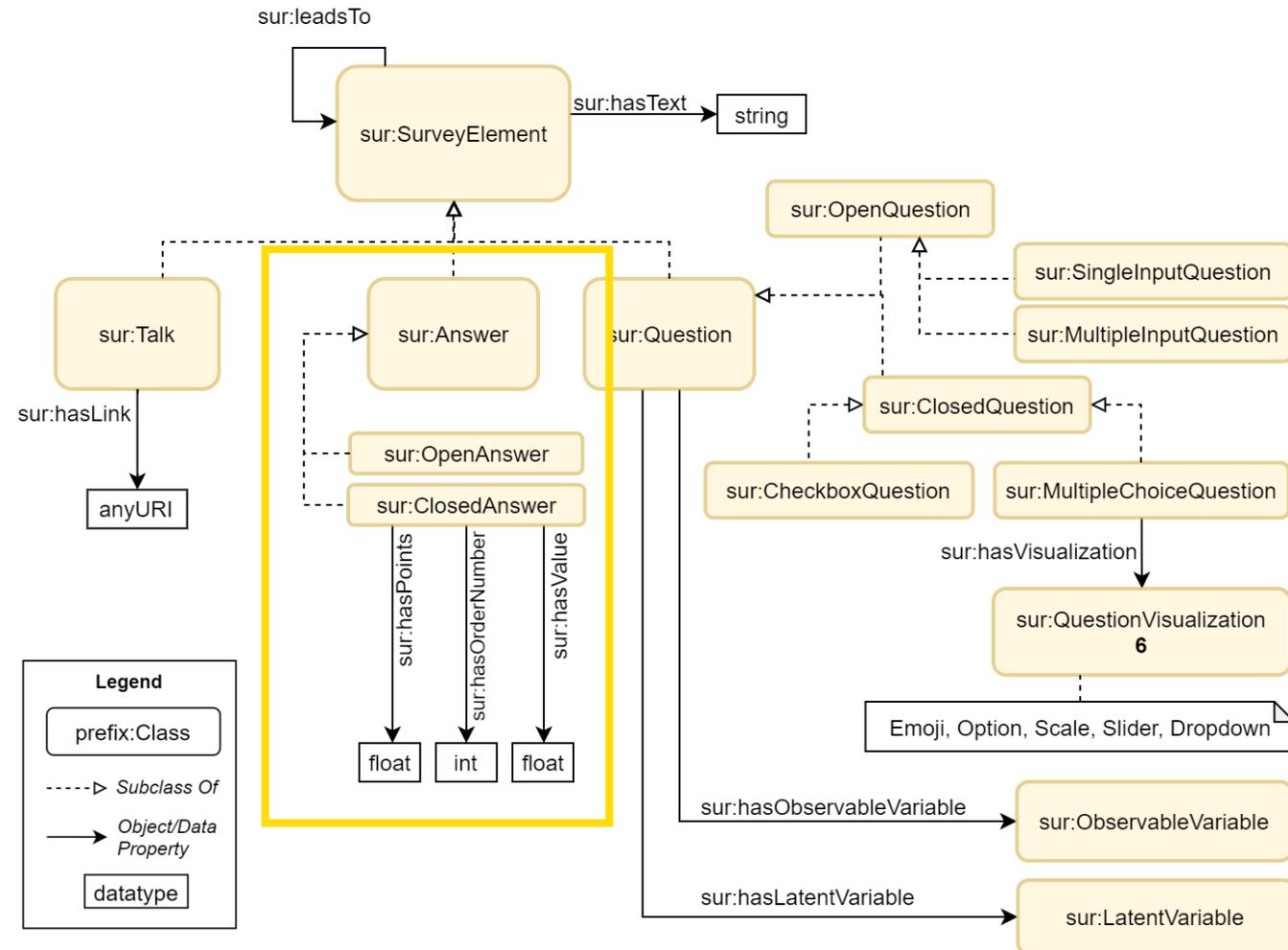
Survey Ontology

- **C1** Make the survey structure available as structured data
- **C2** Annotate questions with the respective investigated variables



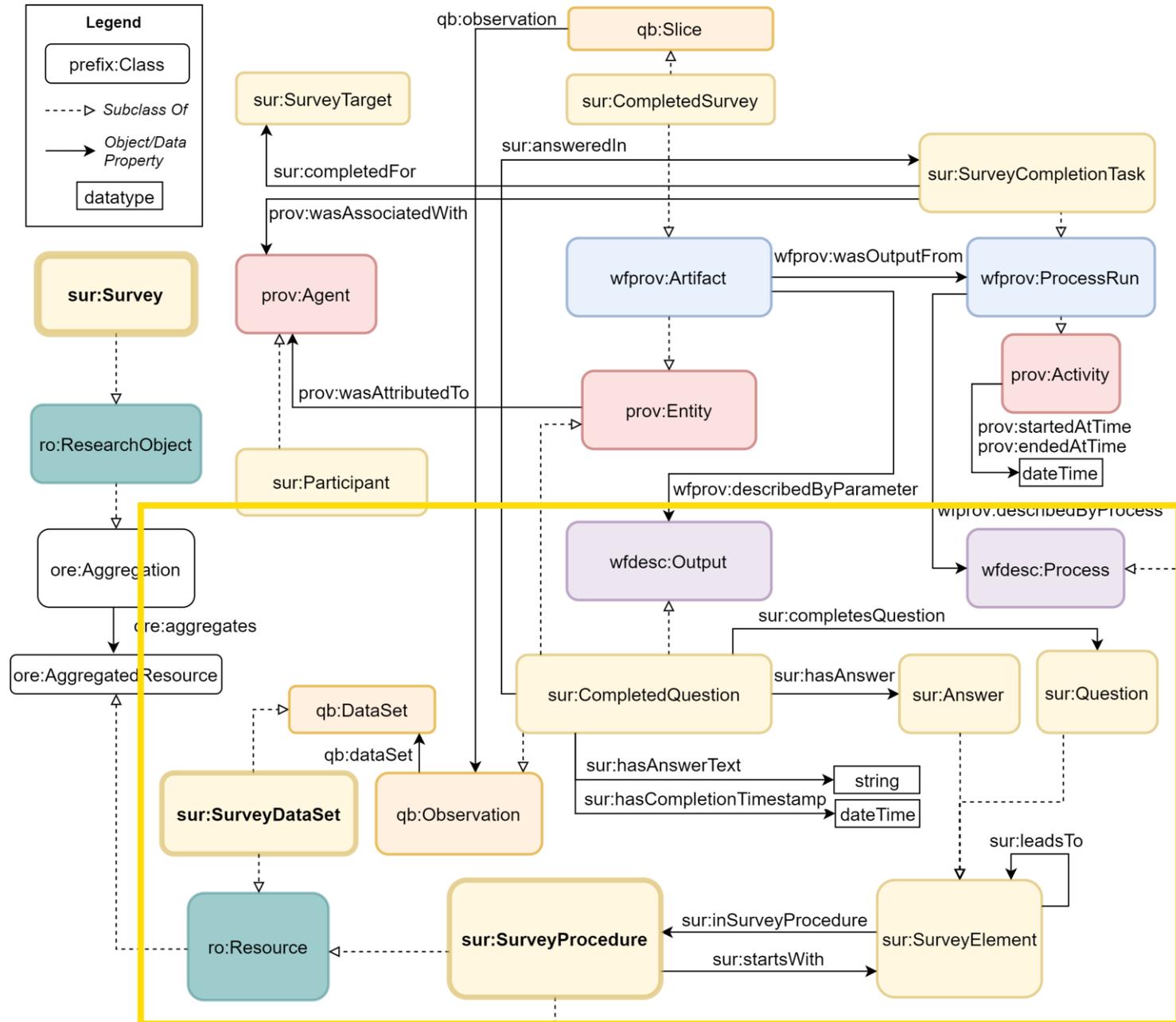
Survey Ontology

- **C1** Make the survey structure available as structured data
- **C2** Annotate questions with the respective investigated variables
- **C3** Annotate answers with their numerical coding



Survey Ontology

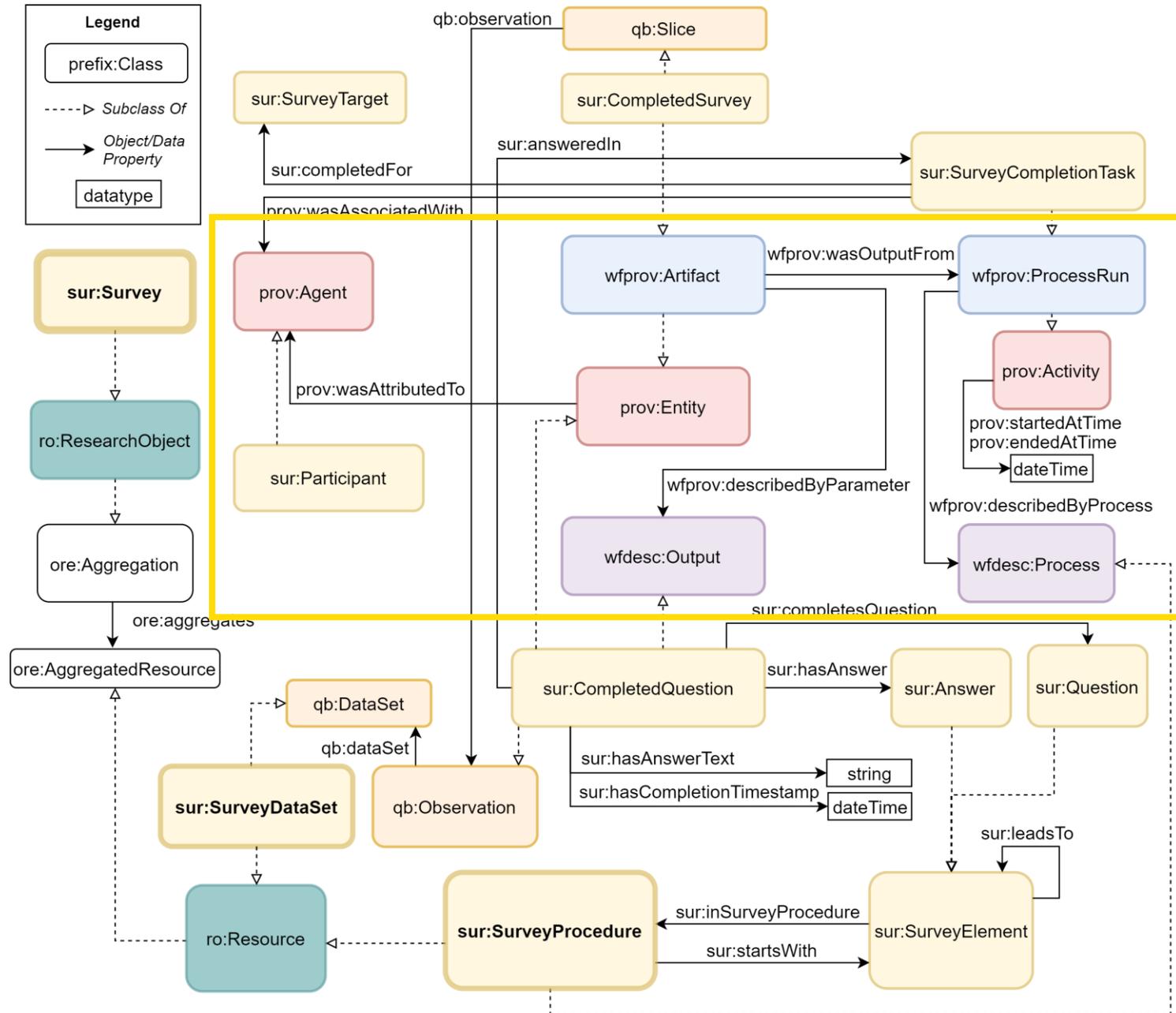
- C4 Make the collected answers available as structure data



sur <https://w3id.org/survey-ontology#>
ore <http://www.openarchives.org/ore/terms/>
qb <http://purl.org/linked-data/cube#>
prov <http://www.w3.org/ns/prov#>
ro <http://purl.org/wf4ever/ro#>
wfprov <http://purl.org/wf4ever/wfprov#>
wfdesc <http://purl.org/wf4ever/wfdesc#>

Survey Ontology

- **C4** Make the collected answers available as structure data
- **C5** Keep provenance of answers

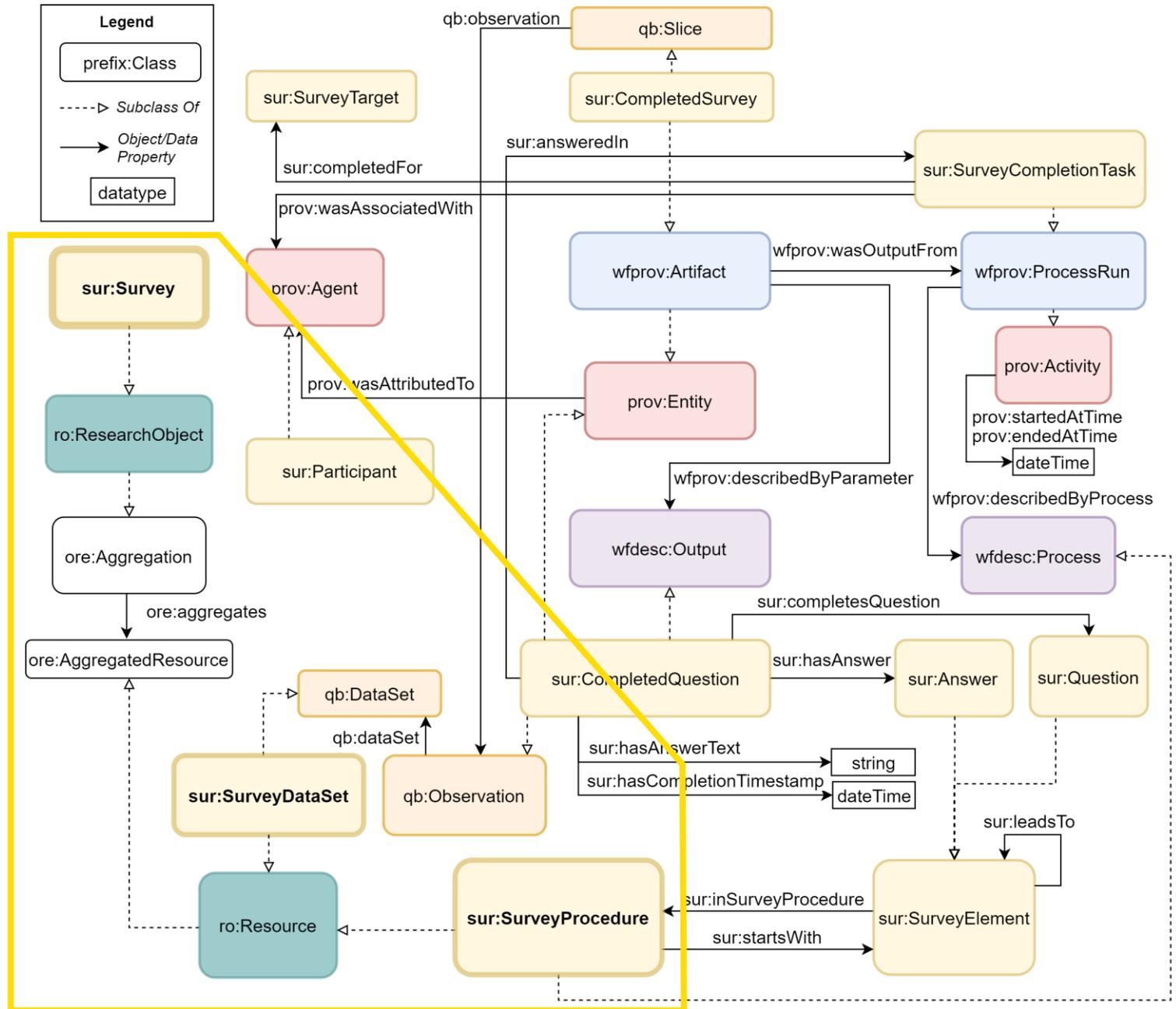


sur <https://w3id.org/survey-ontology#>
ore <http://www.openarchives.org/ore/terms/>
qb <http://purl.org/linked-data/cube#>
prov <http://www.w3.org/ns/prov#>
ro <http://purl.org/wf4ever/ro#>
wfprov <http://purl.org/wf4ever/wfprov#>
wfdesc <http://purl.org/wf4ever/wfdesc#>

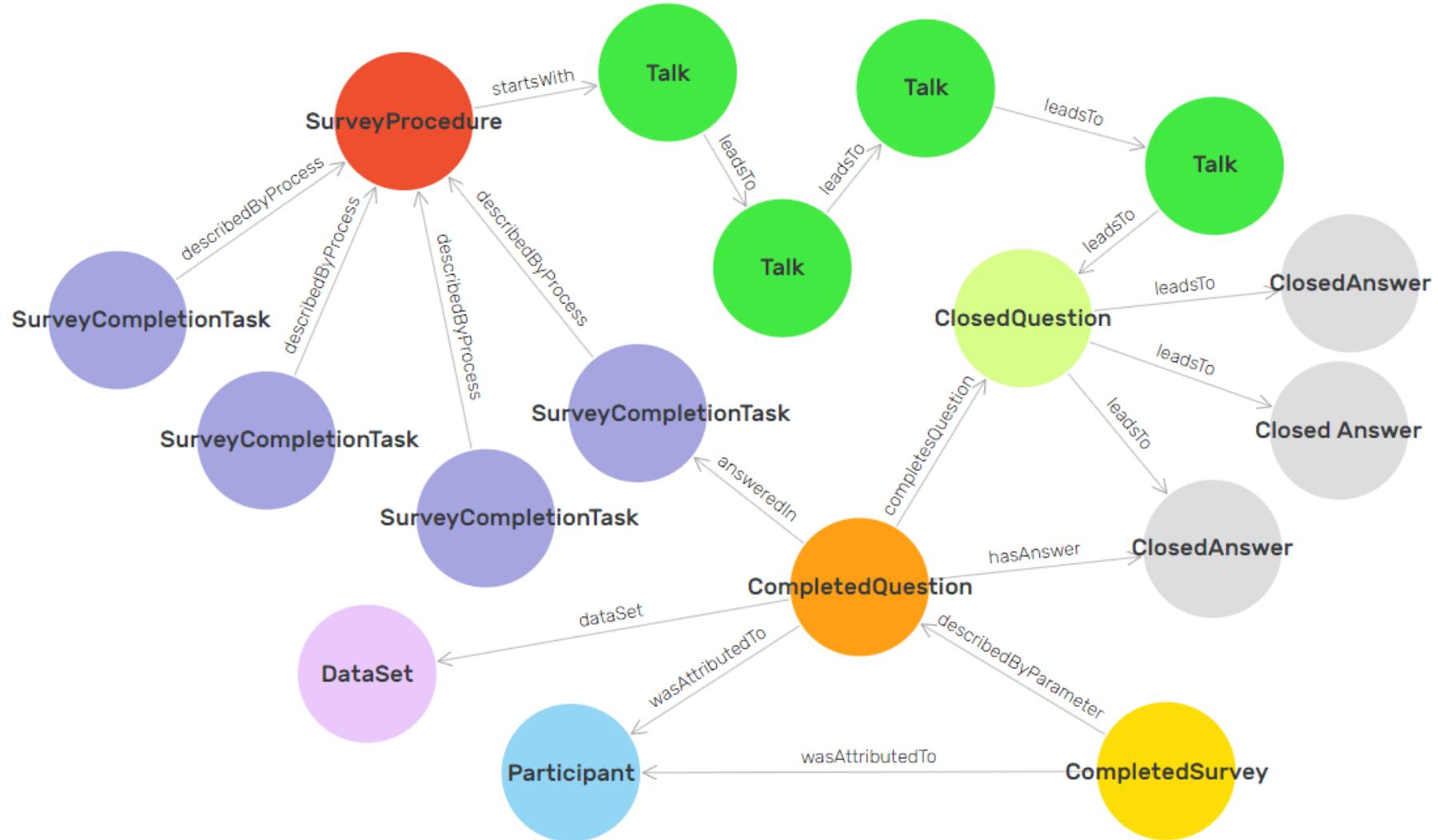
Survey Ontology

- **C4** Make the collected answers available as structure data
- **C5** Keep provenance of answers
- **C6** Share the survey methodology

sur <https://w3id.org/survey-ontology#>
ore <http://www.openarchives.org/ore/terms/>
qb <http://purl.org/linked-data/cube#>
prov <http://www.w3.org/ns/prov#>
ro <http://purl.org/wf4ever/ro#>
wfprov <http://purl.org/wf4ever/wfprov#>
wfdesc <http://purl.org/wf4ever/wfdesc#>



A conversational survey modelled using the Survey Ontology



Studying Motivation of Citizen Scientists: TESS Network Motivation Survey

- Considering a specific survey study, we discuss how we exploited the **Survey Ontology** and Coney to **package and share a research object** describing it to foster reproducibility.
- Within the [ACTION](#) project, we performed a **survey study to analyse the motivation** to participate, in the effort of fighting light pollution, of the TESS citizen science community (<https://tess.stars4all.eu/>).
- Considering this study, we describe how the proposed approach **addresses challenges identified** for a survey researcher interested in sharing the study

TESS Network

TESS

To start I have some questions about you and your involvement in the TESS network

How old are you?

25-34

Which of the following categories identifies you the most?

- Astronomical outreach (museum,...)
- Amateur astronomers
- Professional astronomers
- Public administrations
- Light pollution fighters
- Astrotourism
- Other

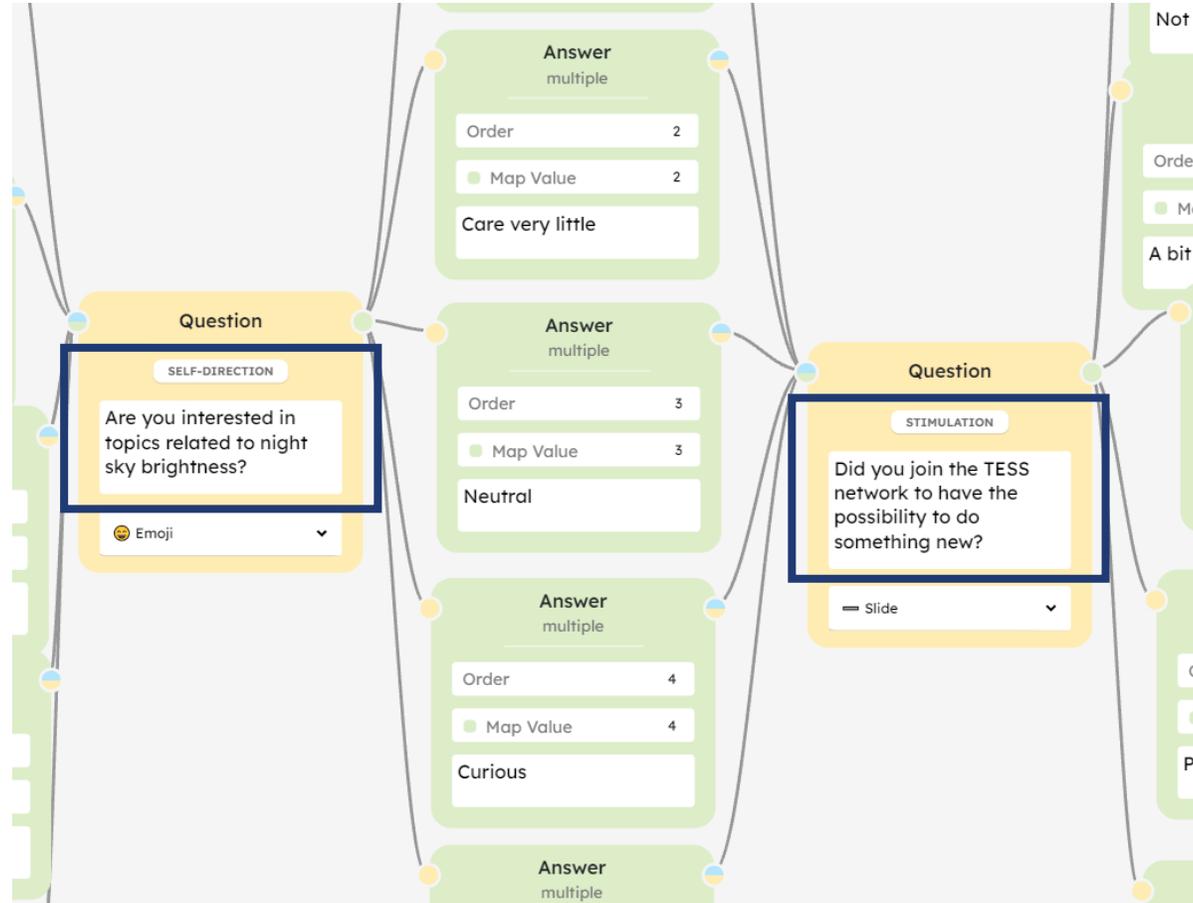
SEND

Irene Celino et al. (2021). *Participant motivation to engage in a citizen science campaign: the case of the TESS network*. JCOM 20 (06), A03. <https://doi.org/10.22323/2.20060203>.

Studying Motivation of Citizen Scientists: TESS Network Motivation Survey

In the defined **survey structure**:

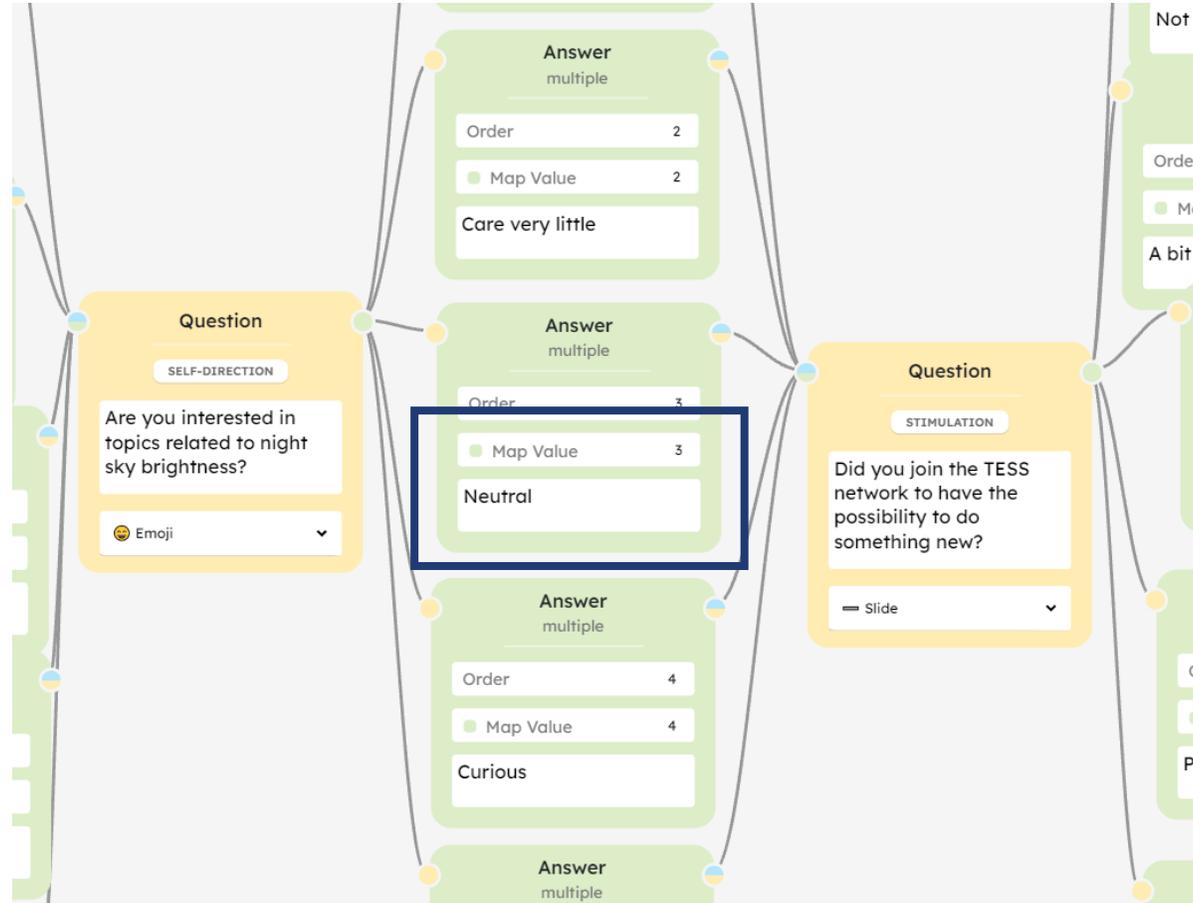
- All the questions were designed to have closed answers and have been annotated with the respective latent variable investigated (**C2**);



Studying Motivation of Citizen Scientists: TESS Network Motivation Survey

In the defined **survey structure**:

- All the questions were designed to have closed answers and have been annotated with the respective latent variable investigated (**C2**);
- the answers are associated with both a qualitative value, a textual label to be displayed in the chat, and a quantitative value, the numerical coding for results analysis (**C3**).



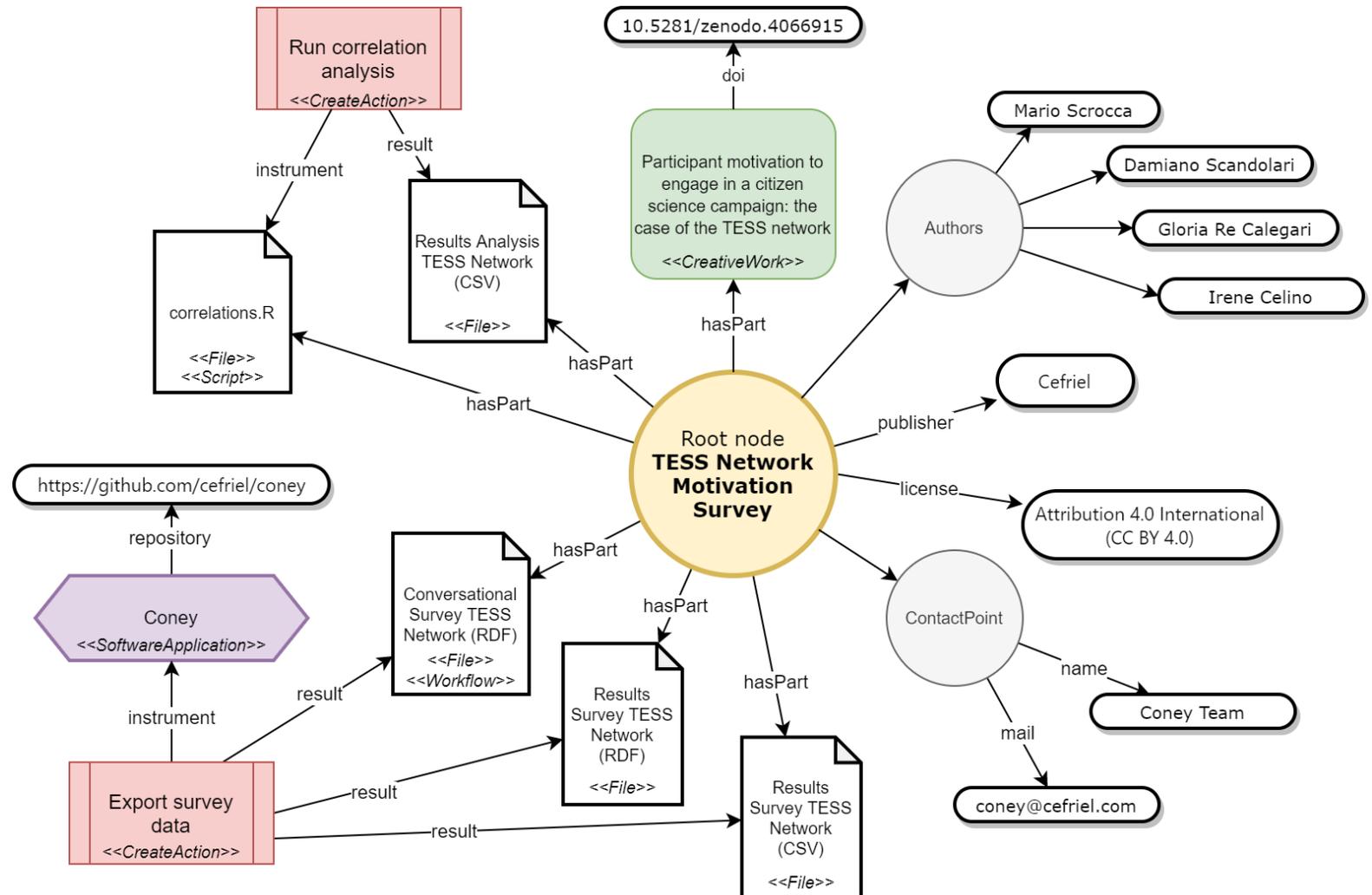
TESS Network Motivation Survey Research Object

The **RO-Crate** specification (<https://w3id.org/ro/crate>) defines a lightweight approach to publish research objects.

The survey data, exported from **Coney** and modelled using the **Survey Ontology**, are packaged in a comprehensive RO-Crate research object including:

- the representation of the survey structure (**C1**),
- the collected answers (**C4**) with provenance information (**C5**),
- the script and results of the analysis, and related publications (**C6**).

RO-Crate published on Zenodo:
<https://doi.org/10.5281/zenodo.5140351>



Conclusions and Future Work

- We presented the **Survey Ontology**, which is the conceptual data model behind our [Coney toolkit](#), but it is also a **generic and comprehensive open vocabulary to describe any kind of survey**
- The Survey Ontology:
 - appeared as a **missing element** in the panorama of available ontologies;
 - we developed it **reusing and interlinking existing complementary vocabularies** like PROV-O, Data Cube, and the Research Objects suite of ontologies;
 - **fosters the packaging and share of survey research as research objects** (e.g., survey structure, the investigated variables and the results of the analysis).
- To support our claims, we published and described a comprehensive research object for the **TESS Network Motivation** study performed using Coney and exploiting the survey ontology to describe the relevant resources.
- As **future work**, we plan to:
 - investigate the **alignment and reuse of other available ontologies**, e.g., to represent hypotheses and the results of the data analysis (e.g., [Research Variable Ontology](#))
 - to extend Coney to **automate the generation of research objects for surveys** compliant with the RO-Crate specification.

Thank you for your attention!



Mario Scrocca
mario.scrocca@cefriel.com