**Introduction**

.

# Survey LOT4KG Methodology

Thank you for considering to participate in this survey, in which we intend to validate and refine the LOT4KG methodology [link]. This survey collects information on Knowledge Graph (KG) projects focusing on the methodology around KG construction as well as Ontology and KG updates. The collected methodology data will be made available online together with the KG project but will NOT include the contact information of the respondent.

With proceeding, you consent to the collection of this information and its sharing for research purposes. Please contact us (email: r.pernisch@vu.nl) if you have any concerns with regards to sharing the collected data about your KG. We included an additional consent box for the relevant information later in the survey.

**We thank you for your time and effort!**

Romana Pernisch, Vrije Universiteit Amsterdam
Maria Poveda-Villalon, Universidad Politecnica Madrid
Lise Stork, University of Amsterdam
Diego Conde-Herreros, Universidad Politecnica Madrid
David Chaves-Fraga, Universidade de Santiago de Compostela

**Intake**

Q1.1. The following information will only be presented in a summarised manner and will not be put online when sharing the collected data.

Contact Details (email)

Affiliation

Role

Experience with Semantic Web technologies, in years

Q1.2. Level of education (select the highest with applies)

- High School / Matura
- Bachelor (BSc/BA)
- Masters (MSc/MA)
- Doctorate (PhD, Dr. of Science, etc.)
- Other

Q1.3. Number of KGs, ontologies or shapes for KG constructed or involved in construction in the past:

KG (focus on Abox)

Ontology or Schema (focus on Tbox, Terminology, Vocabulary)

Shapes

**KG Intake**

. Pick a KG you constructed or have published, which, if possible, involved

more than one interaction. With this we mean that the KG was not only constructed and published but also interacted with again e.g. to change certain aspects or add new data after its initial construction.

Q2.1. Please prodive a reference to the KG in the form of a link to the KG itself, resource identifier or publication:

Q2.2. We want to know about the purpose of this KG, please select from the list below anything that fits.

**Reference/Querying**

- Search
- Exploration
- Other:

**Inference**

- KG Completion
- KG Validation
- Creation of new knowledge (e.g. hypothesis generation)
- Other:

**Machine Learning Task Research**

- Recommender System
- KG Completion, Link Prediction
- Question Answering
- Anomaly Detection
- Other:

**User-facing task** - API, not direct access to graph

- Recommender System
- Search (document or information)
- Decision Support
- Exploration
- Other:

None

Q2.3. Other task that you was not covered in the above list:

Q2.4. What kind of data sources were integrated in this Knowledge Graph?

- Structured
  - Relational database
  - Graph database
  - CSV
  - Other:
- Linked Data
  - Wikidata
  - DBpedia
  - Yago
  - Other:
- Unstructured
  - Documents
  - Webpages
  - Other, smaller pieces:

Q2.5. I consent to the sharing of the above-provided information, in this block, about the KG:

- I consent
- I do not consent

**Ontology Engineering**

.

## Ontology and Schema Engineering

First, we want to know about the ontology or schema engineering process for your Knowledge Graph. With ontology and schema we loosly refer to the **Tbox** of the KG. This can range from OWL to SKOS and anything in between.

General Remark:
Even if statements are generally written from the a first person singular perspective, you can assume them as true even if someone from your team or even multiple people has done a specified task, rather than yourself. This is true for the **remainder of the survey**.

Q3.1. Choose one of the following that you most agree with. The statements are in first person, but we do not make a difference here if this was done by one person, multiple or not by you specifically but someone from your team.

- I engineered my own ontology.
- I reused and changed an existing ontology or ontologies.
- I use (without changes) an already existing ontology. I consider this ontology as external to the KG being developed and under someone else's authority.
- Other:

Q3.1.1. Did you follow an Ontology Engineering Methodology?

- No
- Yes

Q3.1.2. If yes, which methodology did you follow?

- Ontology 101
- Linked Open Terms (LOT)
- SAMOD
- AMOD
- eXtreme Design methodology (XD)
- NeOn
- METHONTOLOGY
- DILIGENT
- On-To-Knowledge
- Gruninger & Fox: “The Role of Competency Questions in Enterprise Engineering”
- Other:

Q3.2. Choose from the tasks/steps below that align with the steps you followed:

**Ontology Requirement Specification**

I specified use cases.

I identified the purpose and scope.

I captured ontology requirements through competency questions.

I defined ontology requirements in a different way.

Other:

**Ontology Implementation**

I checked for ontology reuse.

I took the time to draw/sketch a conceptualisation.

I encoded the ontology in an ontology implementation language.

I evaluated the ontology.

Other:

**Ontology Publication**

I wrote a documentation for the ontology.

I published the ontology.

Other:

**Ontology Maintenance**

I, or someone else, found bugs in the ontology.

I checked for new requirements.

Other:

Q3.3. Are there any other steps you performed that you were unable to mark? Please specify into which of the groups you would put the missing step or task.

**Knowledge Graph Engineering**

.

## Knowledge Graph Engineering

Here we want to know about the process of Knowledge Graph (KG) construction and ask about the details of the process. With KG, we focus now on the **Abox**, meaning the **facts** or **statements**.

General Remarks/Glossary:
"**Automated**" refers to any form of software, script, tool that was involved in the process or step. For example, using a tool to automatically derrive SHACL shapes from the ontology or a python script which transforms a CSV file into RDF.
"**LLM**" refers to the usage of a large language model through prompting.
If there would be a different way you executed a step that would not fall under "Automated" or "LLM", please specify this in the comment box after the task.

Q4.1. If you followed any Knowledge Graph Engineering methodology please provide a reference or description here:

Q4.2. **KG Implementation**
Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

With regards to the actors, we anticipate that often you will select more than one, e.g. for a semi-automated task you can select both 'Automated' and 'Onto/KG Engineer'. If a task was not part of your process, please select 'Not applicable'.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Data preparation: I prepared/ preprocessed the data. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Mapping generation: I wrote or generated mappings (e.g. RML) or a script to be able to transform the data into RDF. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data transformation: I executed the script or mappings. The output of this step was the RDF data. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Constraints generation: I wrote or generated constraints for data validation purposes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data validation: I validated my RDF data (and fixed mistakes that might have been found) | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q4.2.1. **KG Implementation**

Are there any other steps you performed that you were unable to mark or add above? Please describe:

Q4.3. **KG Publication**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Documentation: I wrote or generated documentation for the KG. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data publication: I published the KG (RDF, dump or other) and its documentation. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q4.3.1. **KG Publication**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

Q4.4. **KG Maintenance**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Bug detection: There are contingencies in place to detect and fix bugs in the KG. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q4.4.1. **KG Maintenance**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

**Change Activities**

.

# Change and Update Activities

When it comes to maintaining your ontology and KG, we want to know which steps you perform or which steps are automated and you are ready to perform if the need arises.

We refer to a change as e.g. the triple that is actually changed vs. an update being the task of changing the ontology, schema or KG.

Q5.1. What are the types of evolution and changes you foresee for your graph:

**Change in ontology/schema**

- Change of ontology requirements (e.g. business/task requirements)
- Change of domain knowledge (which will require a change in modelling)
- Data-driven ontology change, e.g. available information changes and requires a change in the ontology

**Change in data** - Mapping dependent

- Change of data modelling in already integrated data sources (change of column label in a database (DB schema) or change of class label/ID in a KG)
- New data to be added or integrated (new data source or new column)
- Deletion/Removal of data source, e.g no longer available or privacy issues, removal of column.

**Change in data** - Mapping independent

- New data items in already integrated data source (more rows)
- Deletion/Removal of data items (fewer rows)

Not relevant or not applicable

Other:

Q5.1.1. Please explain why you do not care about changes or why the question is not applicable:

Q5.2. **Change Analysis**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

With regards to the actors, we anticipate that often you will select more than one, e.g. for a semi-automated task you can select both 'Automated' and 'Onto/KG Engineer'. If a task was not part of your process, please select 'Not

applicable'.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Change conceptualisation: I noted down the changes to be applied and conceptualized them, e.g. in a visual or formal way. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Change evaluation: I evaluated the changes to ensure ontology consistency or other artifacts like shapes, mappings or downstream tasks that might be affected. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q5.2.1. **Change Analysis**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

Q5.3. **Ontology Update**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Change encoding: I implemented the change in the ontology, producing a new version. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Ontology evaluation: I executed an ontology evaluation. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Incremental ontology evaluation: I executed an ontology evluation incrementally. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q5.3.1. **Ontology Update**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

Q5.4. **Change Impact Assessment**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Ontology delta detection: If a list of changes is not available, I run an ontology diff program or extract the changes in some way. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Analysis of change impact: I analyse the impact these changes will have on the KG to determine the best course of action and, hence, refine it to changes that need further processing and filter out those that do not have an effect. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q5.4.1. **Change Impact Analysis**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

Q5.5. **KG Update in case of schema changes**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Mapping update: I updated the mapping rules/scripts based on the ontology changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data update: I run the mapping to regenerate the RDF data. This step is executed incrementally. Meaning only updating the RDF triples which need to be changed or adding new ones and deleting ones that are no longer true and hence not reconstructing the entire RDF graph again. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data update: I run the mapping to regenerate the RDF data. This step reconstructs all the data, does not detect which ones remain the same | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Constraints update: I update the mappings based on the ontology changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data validation: I run the data validation of the new RDF data using the new constraints. This step is executed incrementally only on the data which was updated. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data validation: I run the data validation of the new RDF data using the new constraints. This step is executed over all the RDF data, no matter the changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q5.5.1. **KG Update in case of schema changes**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

Q5.6. **KG Update in case of data changes**

Choose from the tasks below that align with those you followed and select the actor that performed this step. You can comment on the actor in the first text box and in the second text box please specify any tool, library or similar you used for the task.

| | Who was involved in this step: | | | | | If you performed this step, with what tools or libraries? | Comment about this task: |
|---|---|---|---|---|---|---|---|
| | Onto/KG Engineer | Domain Expert | Automated | LLM | Not applicable | Tool, Library, etc. | Comment |
| Mapping update: I updated the mapping rules/scripts based on the ontology changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data update: I run the mapping to regenerate the RDF data. This step is executed incrementally. Meaning only updating the RDF triples which need to be changed or adding new ones and deleting ones that are no longer true and hence not reconstructing the entire RDF graph again. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data update: I run the mapping to regenerate the RDF data. This step reconstructs all the data, does not detect which ones remain the same | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Constraints update: I update the mappings based on the ontology changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data validation: I run the data validation of the new RDF data using the new constraints. This step is executed incrementally only on the data which was updated. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Data validation: I run the data validation of the new RDF data using the new constraints. This step is executed over all the RDF data, no matter the changes. | ☐ | ☐ | ☐ | ☐ | ☐ | | |
| Other Task | ☐ | ☐ | ☐ | ☐ | ☐ | | |

Q5.6.1. **KG Update in case of data changes**

Are there any other steps you performed that you were unable to mark or add above? Please describe below:

**Closing**

.

# Closing remarks

Q6.1. Do you have any additional comments you would like to share with us?

Q6.2. Besides this survey, we are also carrying out in-depth interviews to deepen our understanding of how KG are constructed and maintained, especially in industry settings. If you are interested in providing more information and have a discussion with us, answer yes below. We will use your email contact from the beginning of the survey to contact you to arrange the interview:

- Yes
- No

Q6.3. Thank you for completing the survey and helping us validate the LOT4KG Methodology. Feel free to reach out to us with any additional questions or comments that surpass what you are able to write below. Leave us your name or other form of reference if you would like to be included in the acknowledgements: