

Do you need a knowledge graph?

Helping organizations determine whether a knowledge graph is needed for their problems with the Smals KG Checklist

Knowledge graphs (KGs) are graphs representing entities and their relationships that adhere to three conditions: **they integrate data from heterogeneous sources; they have a schema (or ontology) describing their UoD; and they are used to derive knowledge, insights, etc.**

Organisations have a good understanding of the problems they face, but **do not know whether or how these problems can be solved using KGs**. The problem is that KGs are a difficult concept for organisations to grasp. There are no clear indications when a KG solves a problem and organisations need to rely on the extrapolation of existing case studies.

We present the **Smals KG Checklist**, which has been developed to address this gap. The checklist is to be used in a collaborative setting in which a facilitator fills in the forms. Starting from a concrete problem, the form will help an organisation decide whether a KG provides a viable and elegant solution.

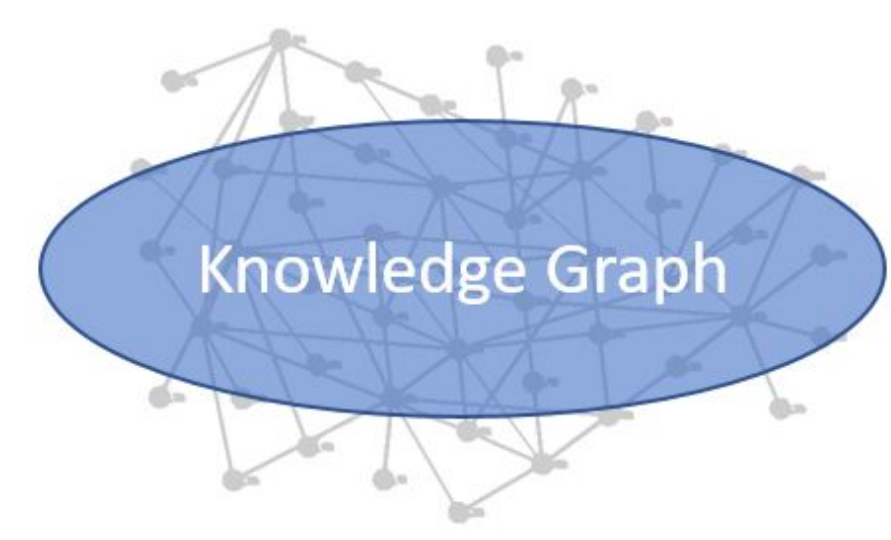
In Part I, we first identify the problem, stakeholders, and core concepts. Then we aim to answer three questions by filling in Part II. These three questions are related to the conditions of a KG. The final question on Part I is used to identify future opportunities for the KG, or its applicability in the longer term.

The [Smals KG Checklist](https://www.smalsresearch.be/) has been made available with a CC BY-NC-SA 4.0. Scan the QR code to obtain the document.

Smals KG Checklist – Part I			
1	Problem identification	Identify a problem and use this document to determine whether a KG may provide or contribute to a solution.	Inspectors need to consult three distinct applications of Belgium's National Social Security for their investigations. They have access to all these applications and information can be conceptually integrated. They lose time doing so manually.
2	Who are the stakeholders? What are the different types of stakeholder?	The presence of multiple (types of) stakeholders is an indication that information needs to be accessible in a meaningful way.	The inspectors conducting the investigations.
3	What are the business concepts that are important in solving the problem?	A list of (high-)level concepts will allow you to identify the information sources.	Employer, Employee, Work Permit, Contract declaration,...
4 Answer these questions by filling in the table on the next page. Then follow the directions to circle "YES" or "NO".			
	Do we need to integrate information and data from multiple (types of) sources?	<ul style="list-style-type: none">• Domain knowledge• Unstructured data• Structured data• ...	Count the number of boxes that have been checked in sections I, II, III, and IV. If the number of checked boxes is greater than 1, then the answer is yes. YES - NO
	Do we need to represent, share, use, reuse, ... information in a meaningful way?	<ul style="list-style-type: none">• Type hierarchies• Role hierarchies• Definitions and glossaries• Business rules• ...	Count the number of boxes that have been checked in section I. If the number of checked boxes is greater than or equal to 1, then the answer is yes. YES - NO
	Do we need to distill (implicit) knowledge and insights from the KG?	<ul style="list-style-type: none">• Via schema reasoning• Via AI and ML techniques• Via applications• ...	Count the number of boxes that have been checked in sections V, VI, and VII. If the number of checked boxes is greater than 1, then the answer is yes. YES - NO
5 Count the number of "YES" you have circled: When that number equals 3, then the use of a knowledge graph seems appropriate. If not, then answer the following questions to see whether the use of a knowledge graph may become appropriate.			
6	If a KG were to be constructed: <ul style="list-style-type: none">• Are there other problems that the KG could help solve?• Would the integration of additional sources help solve other problems?• Are there other stakeholders that may find the KG useful?	Answering these questions may allow you to identify other (potential) use cases for KGs.	<ul style="list-style-type: none">• Integrated information may be useful to facilitate administrative tasks.• Integration of external sources such as the Belgian Crossroads Bank for Enterprises• Integration of internal documentation such as prior investigation reports may facilitate research on an employer's history

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



Smals KG Checklist – Part II		
I. Top-down <ul style="list-style-type: none"><input checked="" type="checkbox"/> Experts (domain, users,...)<ul style="list-style-type: none">> Inspectors<input checked="" type="checkbox"/> Schemas (conceptual, database,...)<ul style="list-style-type: none">> checkinetwork, LIMOSA, and Déclaration Unique de Chantier<input type="checkbox"/> Ontologies, thesauri, standards<input checked="" type="checkbox"/> Rules, laws, regulations<ul style="list-style-type: none">> Belgian Social Security	V. Reasoning and inferencing <ul style="list-style-type: none"><input checked="" type="checkbox"/> Reasoning over concepts<ul style="list-style-type: none">> Queries using class hierarchies<input checked="" type="checkbox"/> Reasoning over relationships<ul style="list-style-type: none">> Queries using role hierarchies<input checked="" type="checkbox"/> Business and domain rules<ul style="list-style-type: none">> Use rules to find problems or violation (e.g., inconsistencies in dates)	VI. AI and Machine Learning <ul style="list-style-type: none"><input type="checkbox"/> Classification<input type="checkbox"/> Prediction (labels, types, relations)<input type="checkbox"/> Clustering<input type="checkbox"/> Recommendations
II. Bottom-up: structured data <ul style="list-style-type: none"><input checked="" type="checkbox"/> Systems, databases,...<ul style="list-style-type: none">> checkinetwork> LIMOSA> Déclaration Unique de Chantier<input type="checkbox"/> Spreadsheets<input type="checkbox"/> Datasets		VII. Interaction and applications <ul style="list-style-type: none"><input checked="" type="checkbox"/> Search<ul style="list-style-type: none">> On employer, employee, form...<input checked="" type="checkbox"/> Look up<ul style="list-style-type: none">> Overlap employees of employers> History of employees/employers<input type="checkbox"/> Explore
III. Metadata <ul style="list-style-type: none"><input type="checkbox"/> Knowledge enrichment<input type="checkbox"/> Provenance (who, what, how,...)<input type="checkbox"/> Governance (roles, responsibilities,...)	IV. Bottom-up: unstructured data <ul style="list-style-type: none"><input type="checkbox"/> Electronic documents and forms<ul style="list-style-type: none">> Reports (currently out of scope)<input type="checkbox"/> Physical and scanned documents<input type="checkbox"/> Social media	Notes <ul style="list-style-type: none">• POC to use databases and systems within the organization's control.• Integration of (or links with) documents stored on file servers not within scope for the first iteration.• Integration of external databases of the public administration not within scope of the project.

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).



In the images above, we use one of Smals Research's KG projects with the Belgian Social Security to illustrate the checklist by filling in the forms. This project aims to integrate the data of three databases to which inspectors have access but are unable to query as a whole. Integrating the data into a KG requires an ontology and would greatly facilitate their work by allowing them to gain insights in, for instance, one's employment histories.

Christophe Debruyne, Katy Fokou, and Paul Stijfhals

Smals Research, Smals, Avenue Fonsny 20, 1060 Brussels, Belgium

research@smals.be

<https://www.smalsresearch.be/>

