Deliverable from the COST Action CA19134 "Distributed Knowledge Graphs"

First Working Group Activity Report

Due date: 2022-08-30
Reporting period: 2020-09-23 to 2022-08-30
Editor: Tobias Käfer (DE)
Authors: Remzi Çelebi (NL), Andreas Harth (DE), Antoine Zimmermann (FR), Olaf Hartig (SE)

Abstract:

This report summarises working-group specific activities as conducted by the working groups of COST Action CA19134 "Distributed Knowledge Graphs" in the first half of the Action's funding period. As such, the report does not cover an in-depth treatment of activities that the Action ran as a whole such as plenary meetings, talks, and short-term scientific missions (STSMs).

WG1

As of October 15, 2022, the working group consists of 61 members (as registered in the e-Cost platform), and 93 people are subscribed to the mailing list of the working group.

The first major activity of the working group was its kick-off workshop, which took place online on January 27, 2021. A total of 23 participants attended this workshop. Besides getting to know one another, we discussed the topics of the Working Group. Various topics were identified and some of them discussed in more detail. Notes taken during these discussions are available to the participants in shared documents.

All working groups worked together on May 20, 2021, to catalog existing knowledge graphs regarding Covid-19 data as a common use case for distributed knowledge graphs. The role of WG1 was to provide a metadata that can catalog other KGs in order to aid in their discovery and utilization. It was observed that we need a clear set of metadata elements and tools for metadata generation.

The WG1 and WG3 leads organized a three-day workshop on "Workshop on Metadata 4 DKG (Metadata Authoring for Distributed Knowledge Graphs)", which took place in Lyon from May 18 to May 20, 2022. The workshop had 32 on-site participants and another 13 participants online. We had two subgroups focusing; i) a core set of metadata properties for KGs and ii) usability of tools for metadata generation. WG1 lead and participants worked on metadata specification to provide concrete guidance on which metadata should be included for a description of a KG. We reviewed existing schemas for relevant elements and brainstormed KG specified metadata elements. After discussing each candidate metadata element, we have identified relevant schema.org and RDF vocabularies and defined ranges. We voted on their inclusion and their cardinality. Finally, the group identified a minimal set of metadata elements for the tool session and re-examined cardinality constraints and added few more candidates. We also added Wikidata metadata as an example.

We also approached the W3C Knowledge Graph Construction Community Group on possible collaboration on the development of the KG Metadata specification. We decided to hold regular meetings with the community group to jointly decide what metadata elements should be included when producing knowledge graphs.

Finally, WG1 contributed to the UC & requirements deliverable with 7 UCs.

WG2

As of October 15, 2022, the working group 2 Consumers of Distributed Knowledge Graphs (WG2) consists of 36 members (as registered in the e-Cost platform), and 84 people are subscribed to the mailing list of the working group.

The first major activity of the working group was its kick-off meeting, which took place online on January 27, 2021. A total of 18 participants attended this meeting. Besides getting to know one another, we discussed the topics of the working group. Various topics were identified and some of them discussed in more detail. Notes taken during these discussions are available to the participants in shared documents.

All working groups met on May 20, 2021, to catalogue existing knowledge graphs regarding Covid-19 data as a common use case for distributed knowledge graphs. The activity on Covid-19 was related to the Health theme as the first focus area we chose for the project. The role of WG2 was to collect existing systems that make use of Covid-19 knowledge graphs. The description of the identified systems have been added to the COST DKG wiki.

At the International Semantic Web Conference (ISWC) 2021, which was held virtually in October 2021, WG2 members organised a challenge ("All the Agents Challenge" - ATAC), which attracted submissions, a substantial virtual audience, and proceedings published with CEUR-WS.

WG2 members contributed to the Use Cases and Requirements deliverable with ten use cases. The identification of use cases was the topic of an online meeting on January 19, 2022.

For the second focus area Digital, Industry and Space, WG2 members Kimberly Garcia, University of St. Gallen, Switzerland, Andreas Harth, FAU Erlangen-Nuremberg and Fraunhofer IIS, Germany and Andras Micsik, SZTAKI, Hungary organised an online workshop on Consumers of Distributed Knowledge Graphs in Digital, Industry and Space. We accepted seven submissions for presentation during an online meeting on February 6, 2022 via Zoom with over 35 participants from all over Europe. The workshop was followed by a social event on gather.town with some long running conversations.

During the plenary meeting in Budapest, Hungary on October 14, 2022, a subgroup of WG2 members formed to discuss requirements and architectures of Linked Data Portal systems. Given that more and more knowledge graphs are available, there is a need for users to interact with, update and visualise knowledge graphs, which the WG2 subgroup aims to address in the future.

WG3

As of November 15, 2022, working group 3 on prosumers of distributed knowledge graphs has 29 members registered in the e-Cost platform, and 59 people are subscribed to the mailing list of the working group.

The working group started its activities with a kick-off meeting on January 27th, 2021 that took place online due to travel restrictions following the Covid-19 pandemic. The workshop was attended by 23 participants. After a short introduction of the working group and the participants, attendees were invited to join a virtual breakout room to discuss a specific topic. Participants could propose their own topics to be discussed in these rooms. Three topics were discussed separately during the first breakout session (45 minutes): specificities of prosumers, KG refinements, and provenance. The second breakout session (45 minutes too) had the following three topics: replication, multi-agent systems, and nanopublications. Notes taken during these discussions are available to the participants in shared documents.

In May 20th, 2021 all working groups collaborated over a common use case for distributed knowledge graphs, namely cataloguing existing knowledge graphs about Covid-19 data. WG3's rôle was to build a knowledge graph describing as many existing Covid-19-related KGs, that is, provide a metadata KG cataloguing other KGs that should help discovering and exploiting them. By extension, the group also discussed how to catalogue arbitrary kinds of Covid-19 datasets, assuming that data not existing in a KG can be translated to a KG by way of transformation tools. It was concluded that we need more systematic ways of describing metadata and more tooling to partially automate the task.

In September 9th, 2021, at the plenary meeting, the work of WG3 on cataloguing continued with a better specification of the metadata templates, edited with an RDF editor, and a simplified web form using Semantic MediaWiki for the most common metadata properties. The goal was to assess whether the metadata of the templates were sufficient and suggest more if necessary, and also find areas of improvement in terms of the process. In particular, one goal was to identify what could be automatised.

In May 2022, WG1 and WG3 jointly organised a 3-days workshop in Lyon, France. The workshop took place from the 18th to the 20th of May, 2022 and 32 people participated. The workshop had an initial and final hybrid session that could be joined remotely, but the main activities were conducted in-person only. The goal was two-fold: 1) identify a core set of metadata properties that are key to the description of knowledge graphs; and 2) to identify existing tools for editing metadata, test them on the core set and suggest improvements. Since the tools presented were ones from the participants themselves, some improvements could be implemented during the workshop. At the beginning of the workshop, Prof. Paul Groth, from the University of Amsterdam gave a keynote on metadata in relation to knowledge graph creation.

Finally, WG3 contributed to the use-cases & requirements deliverable with 8 UCs. The identification of use cases was the topic of an online meeting on the 19th of January 2022.

WG4

As of October 15, 2022, the working group consists of 33 members (as registered in the e-Cost platform), and 83 people are subscribed to the mailing list of the working group.

The first major activity of the working group was its kick-off workshop, which took place online on February 3, 2021. A total of 13 participants attended this workshop. Besides getting to know one another, the main objectives of this workshop were i) to identify cross-cutting concerns relevant for both knowledge graph producers and knowledge graph consumers, as well as for prosumers, and ii) to identify challenges related to these concerns. Various concerns were identified and four of them discussed in more detail (architectures, interoperability, discoverability, and availability). Notes taken during these discussions are available to the participants in shared documents.

In the context of the kick-off workshop, a subgroup was formed with participants interested in privacy and security concerns in ecosystems of knowledge graph producers and consumers. The initial aim of this subgroup has been to identify potential privacy and security issues in decentralised architectures of personal knowledge graphs, with the Solid ecosystem as the major subject of study. The initial findings of this work were presented at the plenary meeting of the Action on September 9, 2021, in Amsterdam. Thereafter, the findings were documented in the form of a position paper that was submitted for peer review to the SEMANTiCS 2022 conference. The submission was rejected with a number of helpful reviewer comments. A revised version of the paper was then submitted to the 21st Workshop on Privacy in the Electronic Society (WPES 2022), which was also rejected. Currently, another revision of the paper is under review for the 2023 ACM/SIGAPP Symposium on Applied Computing (SAC).

The privacy & security subgroup also organised a three-day workshop on "Privacy Issues in Distributed Social Knowledge Graphs" (PIDSKG), which took place in Luxembourg from June 13 to June 15, 2022. The workshop had 45 participants (32 on-site + 13 online). The main objective of the workshop was to explore the legal and cyber security aspects of privacy in the context of distributed social knowledge graphs, with Solid as an example. The first day focused on Solid as a system and on personal identity, with a keynote by Bart Jacobs from Radboud University, NL (title of the keynote: "IRMA as precursor EU-wallet-ID") and an invited talk by Ruben Verborgh, Ghent University, imec, BE (title: "The Solid API for Distributed Knowledge Graphs"). Afterwards, different participants demonstrated their Solid-related systems and the day ended with a round table discussion on the question "What are the Privacy Issues with Solid?" The focus of the second day was on Solid in relation to privacy law & government. The program of that day consisted of presentations (e.g., Bart Buelens from VITO, BE, on "Solid and data governance" and Harshvardhan Pandit from Trinity College Dublin, IE, on the "W3C Data Privacy Vocabularies and Controls Community Group"), round table discussions (topics: gap between privacy law and Solid privacy, privacy required by law), and breakout sessions on policy & governance. The focus of the second day was on Solid privacy as a cyber security problem, with an opening talk by Sabrina Kirrane from the Vienna University of Economics and Business, AT, and related round table discussions and breakout sessions.

Another activity of the working group as a whole was to contribute descriptions of use cases and requirements for the corresponding deliverable of the Action. To kick off this activity, the working group held an online meeting on January 19, 2022. A follow-up meeting for

consolidating the use case descriptions was held online on January 28, 2022. As a result, the members of the working group contributed descriptions of 11 use cases, covering topics as diverse as smart farming, smart cities, health intervention, healthcare risk governance, vaccination roll-out, personalised oncology, etc.