Brussels, 17 May 2024

COST 060/24

DECISION

Subject: Memorandum of Understanding for the implementation of the COST Action "Global Network on Large-Scale, Cross-domain and Multilingual Open Knowledge Graphs" (GOBLIN) CA23147

The COST Member Countries will find attached the Memorandum of Understanding for the COST Action Global Network on Large-Scale, Cross-domain and Multilingual Open Knowledge Graphs approved by the Committee of Senior Officials through written procedure on 17 May 2024.

MEMORANDUM OF UNDERSTANDING

For the implementation of a COST Action designated as

COST Action CA23147
GLOBAL NETWORK ON LARGE-SCALE, CROSS-DOMAIN AND MULTILINGUAL OPEN KNOWLEDGE GRAPHS (GOBLIN)

The COST Members through the present Memorandum of Understanding (MoU) wish to undertake joint activities of mutual interest and declare their common intention to participate in the COST Action, referred to above and described in the Technical Annex of this MoU.

The Action will be carried out in accordance with the set of COST Implementation Rules approved by the Committee of Senior Officials (CSO), or any document amending or replacing them.

The main aim and objective of the Action is to The main aim of this Action is to promote synergies across Europe between scientific communities and industrial stakeholders, in order to enhance the open knowledge available in Europe and beyond. The aim is to provide a large-scale, high quality, cross-domain and multilingual knowledge graph technology that is free to use.. This will be achieved through the specific objectives detailed in the Technical Annex.

The present MoU enters into force on the date of the approval of the COST Action by the CSO.

OVERVIEW

Summary
WHAT: The ultimate goal of the Action is to increase and enhance the public open knowledge available in Europe and beyond. The aim is to provide a large-scale, high quality, cross-domain and multilingual knowledge graph technology that is free to use, reuse, and redistribute.

HOW: By bringing people and communities interested in knowledge graphs technologies to work together on topics related to knowledge graphs engineering, knowledge graphs management and knowledge graphs utilization.

WHY: To align and consolidate the research results on open knowledge graphs, and strengthen the links between the involved scientific communities and public and industrial stakeholders. This will result in a comprehensive, open, multilingual and structured knowledge resources. Such resources can aid the creation of innovative businesses and services that deliver social and commercial value.

FOR WHOM: For scientific communities and public and industrial stakeholders, who develop open knowledge graph solutions for domains, such as cultural heritage, food and agriculture, life sciences, and news and media.

BY WHOM: A network consisting of researchers and innovators affiliated with academic institutes, enterprises, open knowledge networks and initiatives with complementary research and/or business focus and expertise in knowledge technologies.

Areas of Expertise Relevant for the Action	**Keywords**
● Other engineering and technologies: Databases, data mining, data curation, computational modelling for food science and technology ● Electrical engineering, electronic engineering, Information engineering: Databases, data mining, data curation, computational modelling	● knowledge graphs ● open knowledge ● linked data ● multilingualism

Specific Objectives
To achieve the main objective described in this MoU, the following specific objectives shall be accomplished:

Research Coordination
● To prepare a common research agenda for the DBpedia community and the Knowledge Graph community in general. Identify challenges related to knowledge graphs technologies and in a coordinated manner address those challenges.
● To intensify the collaboration among different knowledge graph communities around DBpedia and beyond.
● To improve existing and deploy novel knowledge graphs solutions solutions and improve the quality of existing knowledge graphs solutions.

Capacity Building
● To scale-up and strengthen the DBpedia chapter network by deploying national chapters in COST Member Countries.

● To support young researchers with different research backgrounds to work on new KG related methods and foster knowledge exchange.
● To support knowledge exchange and raise the awareness of the knowledge graph technologies.
● To promote geographical, age, gender and language balance in the activities of the Action.

TECHNICAL ANNEX

1. S&T EXCELLENCE

WHAT: The ultimate goal of the Action is to increase and enhance the public open knowledge available in Europe and beyond. The aim is to provide a large-scale, high quality, cross-domain and multilingual knowledge graph technology that is free to use, reuse, and redistribute.

HOW: By bringing people and communities interested in knowledge graphs technologies to work together on topics related to knowledge graphs engineering, knowledge graphs management and knowledge graphs utilization.

WHY: To align and consolidate the research results on open knowledge graphs, and strengthen the links between the involved scientific communities and public and industrial stakeholders. This will result in a comprehensive, open, multilingual and structured knowledge resources. Such resources can aid the creation of innovative businesses and services that deliver social and commercial value.

FOR WHOM: For scientific communities and public and industrial stakeholders, who develop open knowledge graph solutions for domains, such as cultural heritage, food and agriculture, life sciences, and news and media.

BY WHOM: A network consisting of researchers and innovators affiliated with academic institutes, enterprises, open knowledge networks and initiatives with complementary research and/or business focus and expertise in knowledge graph technologies.

1.1. SOUNDNESS OF THE CHALLENGE

1.1.1. DESCRIPTION OF THE STATE OF THE ART

Open Knowledge has enjoyed a significant increase in popularity and usage in the last decade. Vast amounts of open and integrated knowledge have been published and made available for exploitation by the academy and industry. This knowledge has been primarily nurtured and generated by the **Linked Open Data** (LOD) initiative, which has created the largest publicly available **"knowledge graph of graphs"**, i.e. **LOD Cloud KG**. The LOD initiative has attracted communities from many domains, such as life sciences, government, geography and linguistics. Datasets from the LOD cloud are used as a knowledge source in many general artificial intelligence (AI) applications, including question answering systems, such as IBM Watson that beat human champions in the Jeopardy knowledge quiz in 2011 (Ferucci et al, 2012), or major search engines such as Google. Moreover, in 2018, the international consultancy firm Gartner has identified Knowledge Graphs as a key technology for Artificial Intelligence (Gartner, 2018). Knowledge graphs can also enhance the **reliability and interpretability** of **Large Language Models**, such as ChatGPT, by supplying verifiable information (Pan et al., 2023). Notably, they can both mitigate bias derived by the training data (e.g., harmful stereotypes, misinformation) and alleviate the issue of "hallucination", where LLMs generate erroneous or fabricated information.

Large commercial companies, such as global Internet search vendors, rely increasingly on proprietary knowledge graphs. While these are algorithmically distilled typically from Wikipedia and other open knowledge sources and integrated with proprietary sources, they are not freely available to competition and their transparency is very limited. It is vital that companies as well as researchers retain access to **open knowledge graphs as public good**. These need to have competitive quality and scope to those used internally by global IT hegemons. Possibly even more important is the maintenance of transparent, community-driven processes determining which information enters the knowledge graph and how it is processed. Only in this way can the public retain accountability and control of knowledge-based algorithmic decision making and **prevent algorithmic biases** and discrimination.

Since 2007, when the LOD cloud was started with only 14 datasets, the number of published datasets has increased exponentially. From 45 datasets in 2008, 295 in 2011, and 570 in 2014, to 1,314 published datasets as of September 2023 (LOD stats, 2023). Within the LOD cloud, the "links" are the key enabler for retrieval of related information from different datasets. They are also one of the key prerequisites for a dataset to be included as part of the LOD network of datasets. Over the years, **DBpedia** has become one of the **most widely used knowledge graphs and one of the central interlinking hubs in the**

LOD cloud. The ultimate goal of the DBpedia project is to build a large-scale, multilingual knowledge graph by providing structured information extracted from Wikipedia, and to integrate and complement this knowledge with knowledge from other sources. Over the last ten years, DBpedia has grown into a large community with **22 language chapters** established in several European and overseas countries. Over the years, also the user community around DBpedia has significantly grown. According to the bibliographic database Google Scholar, there are 278,000 articles citing DBpedia; using DBpedia or developing technology for DBpedia. In parallel to DBpedia, there are also several complementary knowledge graph initiatives, such as Wikidata, YAGO, BabelNet and DBkWik. YAGO is a semantic knowledge graph, derived from Wikipedia and complemented with information from WordNet and GeoNames. YAGO (Tanon et al, 2020) provides a comprehensive entity classification schema which has been developed based on the Wikipedia article category information. Wikidata (Ismayilov et al, 2018) is a collaboratively created knowledge graph supported and hosted by the Wikimedia Foundation. Wikidata is highly focused on "centralization" of the information for different Wikimedia projects, i.e. i) centralization of interlanguage links in Wikipedia and ii) central access to data in a similar fashion as Wikimedia Commons for multimedia content. BabelNet (Navigli et al, 2021) is another knowledge development initiative which creates and maintains a multilingual knowledge resource containing concepts and named entities lexicalized in different languages. BabelNet is primarily based on information derived from Wikipedia and Wordnet, but also Wiktionary, GeoNames and Wikidata. DBkWik (Hertling, 2020) is a recently created knowledge graph, which consolidates knowledge extracted from thousands of Wikis. DBkWik extracts knowledge from Wikis available in Wikia, which is one of the most popular Wiki farms. In addition to these open or semi-open initiatives, there are proprietary knowledge graphs (Ji et al, 2022), such as Google's Knowledge Vault, Microsoft's Satori or Facebook's Entities graph.

There are several aspects which make the presented COST Action proposal focus primarily on DBpedia rather than on any of the previously mentioned knowledge graphs.

i) DBpedia is a **decentralized, community-driven effort**. Individual languages in DBpedia are maintained by national DBpedia language chapters. These in many cases unite several institutions. In comparison, the efforts behind most other knowledge graphs are centralized and usually driven by a single institution.

ii) Unique to DBpedia are its communities centralized around a particular **language** or **domain**. These have different and complementary research focuses, such as linguistics, knowledge extraction, knowledge engineering and ontologies, Web technologies, scalable systems, data mining, information retrieval and data quality. The entire **DBpedia ecosystem**, and the DBpedia knowledge graph in particular, can **benefit from the diverse expertise of the members** behind each DBpedia language chapter. Also, different language chapters have different domains in their focus, making DBpedia as a whole improve in its horizontal perspectives (research), but also in vertical perspectives (domains).

iii) DBpedia is **highly integrated** (i.e. interlinked) **with other knowledge graphs** in the LOD cloud, such as YAGO, Wikidata, BabelNet, DBkWiki and GeoNames. These provide rich and comprehensive information, which is linked and integrated with the DBpedia knowledge graph. DBpedia, as a consequence, acts as a **data access proxy** to these and many other knowledge sources. It enables efficient and effortless retrieval and integration of knowledge from multiple sources. In short, DBpedia provides **Global and Unified Access to Open Knowledge** and can thus enable future applications aimed at, for example, **verifying LLM output**.

1.1.2. DESCRIPTION OF THE CHALLENGE (MAIN AIM)

Freely available knowledge, also known as Open Knowledge or Open Data, is a very important factor and has a direct impact on the ongoing technological revolution and advancements and direct societal benefits. Open Knowledge creates a wider range of opportunities for small and medium enterprises, big corporate organizations, government institutions and non-profit organizations to address various societal problems (OpenDataPortal report, 2020).

DBpedia aims to provide **multilingual, open and structured knowledge** for various domains. DBpedia is language driven and in the past decade, the DBpedia network has grown to **22 DBpedia language chapters**. Currently, the DBpedia network includes the following European language chapters: **English (core), German, Dutch, Spanish, Catalan, Basque, Portuguese, French, Italian, Czech, Polish, Swedish, Irish and Greek**. The network also spans beyond the boundaries of the European Union and it currently also includes DBpedia chapters for Japanese, Indonesian, Korean, Russian, Ukrainian and Arabic, but also for the international auxiliary language Esperanto. These DBpedia language chapters work on localization and internationalization of DBpedia and support the extraction of non-English Wikipedia editions as well as build a data community around a certain language, region or special

interest. The language chapters are primarily focused on improving the extraction of data from language-specific Wikipedia versions. They are also part of the DBpedia executives board and contribute to the infrastructure of DBpedia and improvement of the DBpedia knowledge graph.

In spite of the increased popularity and contributions of the DBpedia chapters, **DBpedia has run into noticeable barriers to growth**. Possibly the biggest challenge is related to the diversity of its communities: individual DBpedia language versions have different levels of data quality and different levels of coverage for particular domains. Methods developed and applied in one language version are rarely adopted and applied to other languages. Individual chapters have different dataset release plans; release intervals are not synchronized. While considerable resources are focused on widely spoken languages such as English, for minor and under-resourced languages there is a lack of established DBpedia chapters. This undermines the growth of industries that depend on the availability of comprehensive and accurate structured open knowledge. Most severely affected are small and medium enterprises (SMEs), which cannot afford to build their own knowledge graphs. In particularly disadvantaged positions are the SMEs in European countries, where minor or under-resourced languages are spoken.

The key problems that hinder the development and growth of DBpedia, as a source of machine- readable open knowledge, are as follows:

i) Low accuracy and coverage across different languages and domains. Currently, the quality of the extracted and provisioned knowledge varies significantly across languages. For some major languages (e.g. English, German and Dutch) a number of methods have been developed for efficient extraction and integration of information. However, minor languages, for instance Slavic languages (Czech, Slovak, Slovene, Macedonian, Bulgarian, etc.) and Semitic languages (such as Maltese) still lack resources. The scope and accuracy of extracted knowledge could increase substantially when even basic computational linguistics methods are applied (e.g. stemming, lemmatization, part-of-speech tagging, named entity recognition, etc.). Furthermore, individual domains have different degrees of development. For example, while general domains, such as sports, music and geography, are well represented, more specific domains, such as finance, biology and medicine, are poorly covered. Thus, there is a requirement to increase the accuracy and the coverage of knowledge extracted especially for **minor languages and less popular domains.**

ii) Limited interaction with users. Wikipedia, the main source of information for DBpedia, relies almost exclusively on contributions of content from users. These contributions do not constitute only newly added material, but also refinements and error corrections. While DBpedia has a large and highly skilled user base, individual users rarely contribute to refinement of the content they use and benefit from. With adequate tools and incentivization, the “wisdom of the crowd” of DBpedia users could be activated to improve the quality and scope of DBpedia.

iii) Lack of a cross-language transfer of knowledge. Many advanced methods have been proposed in the scientific literature that can be used to increase the scope and coverage of available open knowledge. So far, these techniques have been used to the benefit of English as well as a few other languages. This can be attributed to two major factors: i) limited availability of Natural Language Processing (NLP) tools for minor languages and ii) lack of coordination and cross-language support. Transfer of knowledge, methods and technology across different language chapters is essential for the growth in quality and coverage of DBpedia across different language chapters. This requires **scientific coordination of the DBpedia research and development activities.**

iv) Lack of a systematic development strategy. The research and development activities behind DBpedia are primarily driven by the requirements coming from individual language chapters. These requirements are as diverse as the languages the chapters represent. In addition, there are requirements from the end-users (organizations, academy, industry), which are usually captured and processed in an ad-hoc manner. A particular manifestation is duplicated efforts when a particular language-specific solution is developed. This can be attributed to the problem that the required technology has not been integrated and made available via the respective technology stack. Furthermore, the data management in DBpedia requires improvement, and in particular, rich metadata should be made available, the integration with other datasets should be regularly validated, and finally, data preservation plans and archiving plans should be clearly stated and executed. In summary, there is a requirement for a coordinated R&D.

The Action addresses the above-mentioned problems and articulates them into the following set of research questions (RQ):

RQ1: How to assess and improve the quality of DBpedia in terms of accuracy and coverage for different languages and domains? The Action will advance the state of the art in knowledge extraction and develop novel methods. The Action will explore novel methods for domain and language-specific

extraction of knowledge. Existing methods will be extended (where possible) and applied for different languages and domains. Novel data fusion techniques will be utilized in order to improve the quality and the confidence in the provisioned knowledge. The domain specific improvement will be achieved in collaboration with domain experts and users.

RQ2: How to increase the amount of knowledge for under-resourced and minor languages? The Action includes representatives from countries with under-resourced and minor languages, such as Slovene, Macedonian, Czech, Slovak, Bulgarian, Maltese etc. The representatives, being native speakers for the targeted languages, will help adapt existing methods and develop models for efficient knowledge extraction for these under-resourced languages. The Action will provide all necessary means (tooling, guidelines and best practices) for deployment of a localized DBpedia knowledge graph for over 120 world-spoken languages, including many minor and under-resourced languages. This will reduce the costs of extending the scope of open knowledge for any (major or minor) language.

RQ3: How to enable seamless and effortless transfer of knowledge, methods and technologies across different DBpedia language chapters? The Action will investigate the current state of the methods and technologies developed within the context of the individual DBpedia language chapters. Also, the robustness of existing language-specific methods and their portability to other languages will be evaluated. Finally, a strategic plan for scientific coordination of the DBpedia research and development activities will be developed.

RQ4: How to exploit the "wisdom of the crowd" and enhance the efficiency of knowledge extraction and provisioning? The Action will enhance the DBpedia web portal with tools and methods allowing users to be directly involved in the process of: i) validation of existing knowledge, ii) collection of training data for knowledge extraction methods, iii) collection of general feedback on knowledge requirements (data quality and domain specific requirements). The portal will adopt the characteristics of a social Web platform. It is recognized that customized communication channels need to be created for individual types of users, such as developers, knowledge providers and consumers.

Since findability, accessibility, interoperability and reusability of the knowledge is in the main focus of the Action, all the knowledge generated by the Action will be published along the F.A.I.R. (Findable, Accessible, Interoperable and Reusable) principles. The Action will implement the F.A.I.R. principles in the context of Action and provide guidelines for publishing FAIR Linked Data.

1.2. PROGRESS BEYOND THE STATE OF THE ART

1.2.1. APPROACH TO THE CHALLENGE AND PROGRESS BEYOND THE STATE OF THE ART

The Action's approach to the challenge is to **bring people and communities together** that work on technologies which support creation, management and exploitation of knowledge graph solutions. The Action is organized into two technological Working Groups (WGs) concerned with the **development and management** of knowledge graphs.

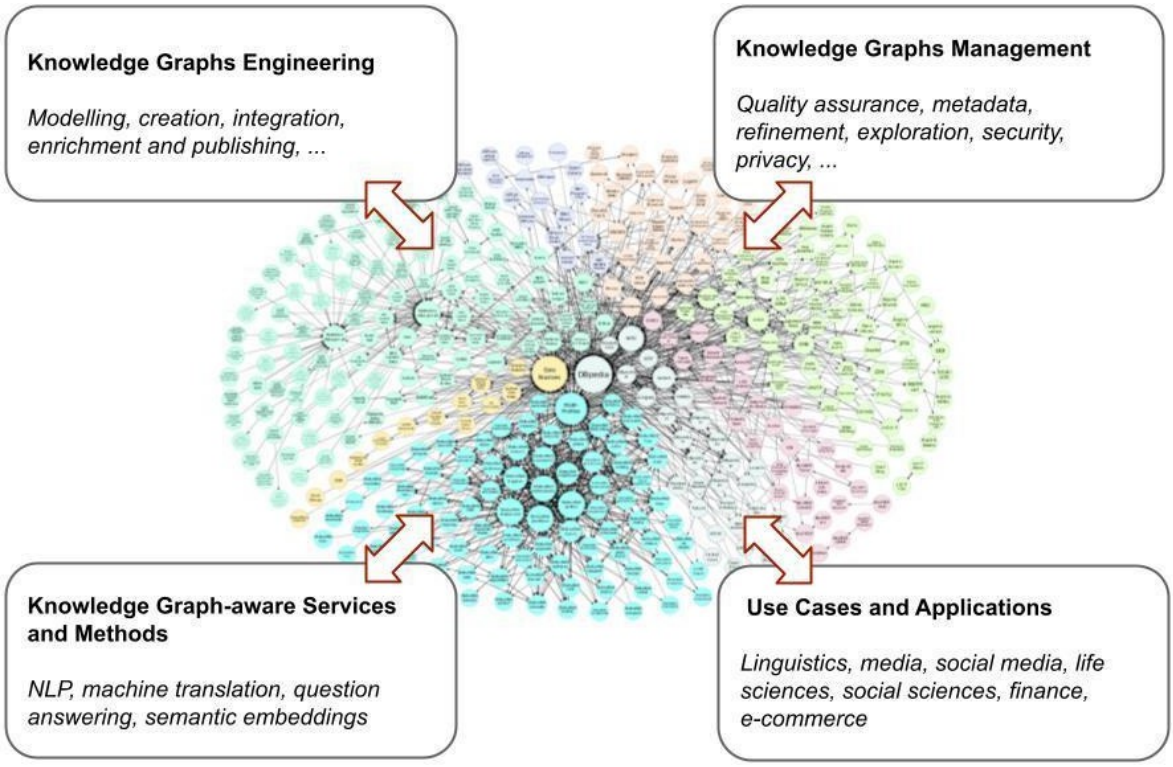
- The WG on "Knowledge Graphs Engineering" is concerned with **engineering** aspects of the knowledge graphs such as modelling, creation, integration, enrichment, linking, and publishing.
- The WG on "Knowledge Graphs Management" is concerned with **management** aspects of the knowledge graphs such as quality assurance, metadata, refinement, exploration, security and privacy.

The Action also covers the **exploitation** aspect with two exploratory WGs concerned with the use and benefits of knowledge graphs for particular methods, services, applications and use cases.

- The WG on "Knowledge Graph-aware Services and Methods" is concerned with **utilization** of open knowledge graphs in order to enhance the performance of AI methods and services.
- The WG on "Use Cases and Applications" is concerned with the **consumer** aspects and possible use cases and applications of the knowledge graph technologies.

DBpedia will facilitate and support the work within the Action since DBpedia provides all the means required, i.e. strong multilingual community, community with strong expertise in linked data and semantic web technologies, knowledge engineering and data science, large pallet of tools and services exploiting DBpedia, number of methods developed, benchmarked or integrated with the DBpedia KG, number of applications and use cases developed with help of DBpedia, and strong engagement of the academic and industrial players in the development and exploitation of DBpedia. However, although the main motivation for the Action is strengthening the DBpedia community and the advancing DBpedia knowledge assets, **the focus of the Action is not only on DBpedia but in its focus are also other knowledge graphs** such as Wikidata, GeoNames, WordNet, YAGO, BabelNet, and other knowledge

graphs from the LOD cloud, and knowledge graphs that will emerge during the course of the Action. Selected sets of external knowledge graphs will be linked, integrated and exploited in combination with the DBpedia core knowledge graph.



The Action aims at **advancing the state-of-the-art in knowledge graphs technologies** in the following aspects.

- **Deploy novel national knowledge graphs.** Currently, the DBpedia community is organized into chapters where the common denominator is the language. While this has been proved as successful approach, this Action will further advance the DBpedia community and expand the DBpedia network with a novel concept, i.e. **National Knowledge Graphs**. The focus will be on provisioning of knowledge of high accuracy and coverage for a domain of particular interest of the stakeholders. As a proof of concept, the DBpedia team has deployed the **Dutch National Knowledge Graph**, which "through the eyes of DBpedia", integrates authoritative datasets relevant for the Dutch nation. The Action will further advance these achievements and develop novel methods which will support the creation of national knowledge graphs. The Action will provide novel methods for knowledge integration, fusion and refinement. During the course of the Action our estimation is to deploy over 10 novel national knowledge graphs in different counties participating in the Action.
- **Increase the quality of the DBpedia KG.** In the current state-of-the-art, a lot of methods concerned with Knowledge Graph (KG) engineering and KG management have been developed and applied just on a particular DBpedia dataset or language. For example, current approaches for knowledge extraction, knowledge refinement and quality assurance have been applied only to a particular DBpedia dataset. Thus, this Action will investigate new approaches for knowledge integration, extraction, enrichment, refinement and quality assurance (within WG 1, 2 and 3) which will advance the **syntactic and semantic quality** of the DBpedia knowledge graph.
- **Improve the coverage for minor languages and domains.** In the current state-of-the-art, the majority of the works are focused on popular, major languages such as English, German, French or Italian and popular, wide-spread domains such as linguistics or geography. The Action will study the adaptability of the existing methods for other languages and domains. This work will be supported by WG4 in providing domain specific analytical information for the work in the KG engineering (WG1), KG management (WG2) and KG services and methods (WG3) working group. The current list of minor languages includes Slovene, Macedonian, Serbian, Hungarian, Czech, Albanian, Bulgarian, Slovak, Polish, Maltese, Greek and Turkish. Support for other languages will be also considered during the course of the Action. The domains will be selected based on the interest of the network members, and the stakeholders coming from different

domains.

- Enable augmenting **large language models (LLMs)** by using high-quality data. The crowd-sourced creation, validation, and improvement of DBpedia language chapters, thanks to the broad network of the Action, would directly increase openly specific language that can serve as a valuable resource for training and fine-tuning specific-language models. This facilitates the development of models that can understand and generate content in a wide array of languages, improving language support for linguistic diversity. Feedback from users, including language-specific requirements and insights into under-resourced languages, can guide the development of large language models to cater to a broader user base, ensuring more equitable and open access to AI technologies.

In addition to the advancements elaborated above, the Action will also define guidelines and best practices for engineering, management and exploitation of knowledge graphs. The Action will consider the entire data pipeline in knowledge graph creation and management and for each stage include experts in the field of engineering, management, services and users applications.

Other foreseen advances in the state-of-the-art concern knowledge-driven services and methods (cf. WG3) which will be able to increase their coverage and performance thanks to the availability of the improved and better integrated knowledge served through DBpedia. A key benefit of the knowledge graph-aware services and methods is their capability to be adapted to new languages and domains.

1.2.2. OBJECTIVES

1.2.2.1. Research Coordination Objectives

The aim of the Action is to further advance the Knowledge Graph technology landscape by connecting scientists and stakeholders from the industry. The main research coordination objectives of the Action are as follow:

- To prepare a common research agenda for the DBpedia community and the Knowledge Graph community in general. Identify challenges related to knowledge graphs technologies and in a coordinated manner address those challenges.
- To intensify the collaboration among different knowledge graph communities around DBpedia and beyond.
- To improve existing and deploy novel knowledge graphs solutions solutions (e.g. applications, tools and services using KGs) and improve the quality (quality measured according to the state-of-the-art Linked Data quality assessment measures) of existing knowledge graphs solutions.

1.2.2.2. Capacity-building Objectives

The aim of the Action with respect to capacity building is to strengthen and widen the knowledge graph community around DBpedia and beyond. The main capacity-building objectives of the Action are as follows:

- To scale-up and strengthen the DBpedia chapter network by deploying national chapters in COST Member Countries.
- To support young researchers with different research backgrounds to work on new KG related methods and foster knowledge exchange.
- To support knowledge exchange and raise the awareness of the knowledge graph technologies.
- To promote geographical, age, gender and language balance in the activities of the Action.

2. NETWORKING EXCELLENCE

2.1. ADDED VALUE OF NETWORKING IN S&T EXCELLENCE

2.1.1. ADDED VALUE IN RELATION TO EXISTING EFFORTS AT EUROPEAN AND/OR INTERNATIONAL LEVEL

In relation to other COST Actions. According to the COST.eu database (http://cost.eu) there are few COST actions which are related and complementary to this Action. The ongoing "Distributed Knowledge Graphs" (CA19134) COST Action is a very close and complementary initiative that aims at providing and disseminating good practices and lessons learnt in the field of distributed knowledge graphs. The main concerns of the related Action are the scalability and the degree of distribution of knowledge graphs. These aspects are not in the main focus of the proposed Action but are highly relevant and complement

the goals of this Action. The outcomes of the related Action will constitute a valuable input for our own activities, in particular in WG 2 “Knowledge Graphs Management”. There is one more COST Action that is focused on research that could directly benefit from improved availability of open knowledge: European network for argumentation and public policy analysis (CA17132).

In relation to the Open Data initiatives. The Open Knowledge Foundation (OKF) is a related network which is primarily focused on promoting the idea of open knowledge, raising awareness for the importance of open knowledge and supporting the use of open data. DBpedia and OKF are complementary in that DBpedia acts as a multilingual knowledge provider, while OKF spreads the word on the importance of open knowledge. The EU Open Data Portal is an initiative primarily focused on enabling open access to data published by EU institutions. The actual knowledge extraction and integration is out of the scope of the EU Open Data Portal, as well as related open data initiatives, such as Data.gov, launched by the U.S. government, and the Open Government in Canada.

In relation to the Knowledge Extraction efforts. In the past decade, a number of projects, initiatives and networks centered around extraction and provisioning of knowledge. The most relevant results from these efforts include the creation of YAGO (a comprehensive entity classification schema), Wikidata (a collaboratively edited knowledge base), BabelNet (a multilingual lexical knowledge base) and DBkWik (a multi-Wiki consolidated knowledge graph). There are several aspects which make DBpedia stand out in relation to the existing knowledge sources. First and foremost, DBpedia is a decentralized community-driven effort, where the quality of the knowledge in a particular language highly depends on the effort spent by the dedicated DBpedia language chapter, which is represented by one or more institutions. In comparison, most of the related efforts are centralized and led by a single institution. DBpedia, in comparison to the other knowledge sources, is highly integrated with many other open datasets. This makes DBpedia a data proxy, or in other words, a **knowledge integration point**. DBpedia is also unique with its very strong community concentrated around individual language chapters. These language-specific sub-communities, via participation in European and other national projects, have developed a number of methods for efficient knowledge extraction, enrichment, integration, quality assurance and retrieval. However, these methods have been developed for a particular language and/or domain. Moreover, usually these methods are not consolidated with the core DBpedia extraction framework, left unmaintained and no longer applied. The Action aims at **harmonization, alignment and integration of these research efforts**, where networking tools are highly required to maximize the results.

Added value of networking. In the last decade, a number of methods for and using DBpedia have been developed. According to the Google Scholar bibliographic database, there are 278,000 scientific articles mentioning DBpedia, many of them using DBpedia or developing a technology for DBpedia in a way or another. This includes methods for identifying or fixing data quality issues in DBpedia, methods for enrichment of DBpedia with additional information, methods for integration of knowledge sources, methods exclusively developed for particular languages, crowdsourcing based methods for DBpedia and other methods and techniques for more efficient knowledge extraction. These methods and technologies have been developed as part of many European, national and international projects, such as LOD2, Big Data Europe, LinkedTV and LIDER. Although potentially highly valuable for increasing the quality and scope of open knowledge available through DBpedia, most of these results have not yet been integrated and consolidated within the DBpedia framework. The Action will leverage and consolidate the results from these efforts. The networking aspect of the Action will help harmonize, align and integrate these research efforts, where networking tools are highly required to maximize the results and the impact. As a consequence, this will 1) improve the quality of DBpedia in terms of improved accuracy and coverage for different languages and domains, and 2) increase the amount of knowledge for under- resourced and minor languages. The ultimate result from the networking behind the Action is to enable creation of large-scale, high quality, cross-domain and multilingual open knowledge graphs.

2.2. ADDED VALUE OF NETWORKING IN IMPACT

2.2.1. SECURING THE CRITICAL MASS, EXPERTISE AND GEOGRAPHICAL BALANCE WITHIN THE COST MEMBERS AND BEYOND

Improving the data quality and coverage of knowledge for different DBpedia languages and domains requires experts with strong expertise in knowledge extraction, enrichment, integration, data quality assurance, retrieval, data mining, natural language processing and scalable software engineering. This requires involvement of experts from different domains and representatives of countries with a variety of spoken languages. The Action's network of proposers consists of top researchers in the field of Linked Data and Knowledge Engineering where many proposers are familiar with the DBpedia technology, have used it or contributed to its development. In order to provide stronger support for under-resourced languages, the network includes representatives from COST inclusiveness target countries (ITC).

Further, to provide support and better coverage of knowledge for different domains (cf. WG 4), the network also includes representatives with expertise in linguistics, news and media, life sciences, libraries and cultural heritage.

In order to further strengthen the network and reach the critical mass of expertise in the field of knowledge graphs technologies the Action will organize **tutorials, workshops and hackathons**. The Action will also maintain **web presence** through the website and various social media channels. The DBpedia association will also support the promotion of the Action and encourage people to get involved in the Action via the existing **DBpedia communication channels** such as the DBpedia blog, DBpedia forum, DBpedia Slack workspace and the regular DBpedia newsletter.

2.2.2. INVOLVEMENT OF STAKEHOLDERS

The Action is initiated by **mix of stakeholders** with different background motivations: (i) Semantic Web researchers exploiting DBpedia in their everyday research work, (ii) Open Data specialists, (iii) knowledge providers, (iv) DBpedia language chapter leads, (v) academics and practitioners interested in deploying national knowledge graphs, (vi) linguists interested in exploiting DBpedia in linguistic science and (vii) specialists from the domains of the life-sciences, cultural heritage, media, language technologies and others, eager to explore knowledge graph technologies for new use cases and applications. Moreover, the network of proposers, although primarily composed of academics, has strong ties with industry stakeholders from various national and international projects which provide commercial knowledge graph solutions for their customers. The Action will include a **dedicated task on stakeholder selection and engagement** (see Task 5.3). This task will also make sure that the effort is equally distributed to the different stakeholders' fields so that all the stakeholders and areas are equally addressed and treated. The stakeholders will be involved from the very early stage of the Action and gradually the list will be extended with specialists from different fields. The Action will enable stakeholders to learn, establish connections with other relevant scientists and exchange knowledge and experiences.

The **DBpedia community** will be directly involved and members will be invited to join the Action. The network of proposers already includes DBpedia community members which have developed and/or used DBpedia technology in the past. DBpedia community members will contribute methods, data, tools, guidelines, but also take an active part in the dissemination activities.

The initial **network of proposers** consists of many experienced researchers with strong collaborations with relevant entities, which will help further expand the network of stakeholders. The Action will organise tutorials, hackathons, workshops and training schools at which a special focus will be put on direct involvement of relevant stakeholders coming from both academia and industry. Stakeholders will be trained to efficiently use the developed knowledge graph technologies. Feedback on the data and tools will be collected and used for further improvement. Relevant stakeholders will also be invited to present current challenges in their work and discuss how the knowledge graph technology can address those challenges.

A particular focus will be put on **engagement with the relevant industry partners**. This will be primarily realized via WG4 Use Cases and Applications and Task 5.4, where in close collaboration with the industry partners new use cases will be developed and ideas for new novel applications discussed. The industry partners will be also supported in the use of the developed knowledge graph technologies and based on the received feedback the technologies will be further improved.

3. IMPACT

3.1. IMPACT TO SCIENCE, SOCIETY AND COMPETITIVENESS, AND POTENTIAL FOR INNOVATION/BREAKTHROUGHS

3.1.1. SCIENTIFIC, TECHNOLOGICAL, AND/OR SOCIOECONOMIC IMPACTS (INCLUDING POTENTIAL INNOVATIONS AND/OR BREAKTHROUGHS)

Scientific and technological impact. Open knowledge is a key enabler of many research projects across OECD research categories. For instance, in computational NLP, the entity linking task is directly dependent on knowledge graphs. Outside of computer science, linked data have found many uses in education, biomedicine, law, linguistics, etc. But these applications are not without their challenges. For instance Pereira et al, 2017 identified data integration and metadata management as two factors undermining future growth of linked data in education. Precisely, the presented Action has improvement in these two aspects among its core objectives. In the short term, the objective is to support at least five new educational applications using DBpedia in minor or under-resourced languages. In the realm of natural language processing, the Action aims for measurable and statistically significant improvement in

the performance of applicable NLP techniques, such as machine translation for the target languages. In the long term, it is expected that a number of new applications in research and technology will use DBpedia across different OECD categories. This objective is measurable via cited reference search and meta-analysis collapsed over the primary science category of citing journals.

Regarding NLP techniques, we are witnessing an incredible development and adoption of techniques based on deep learning, in particular on Transformer-based pre-trained language models. The combination of the implicit semantics encoded in language models with the explicit semantics of KGs has a lot of potential for research and practical applications (to be analysed in WG3). For instance to inject common sense knowledge on language models, to enhance the interpretability of their results, or to complement such results with fact checking techniques. A possible application is **"hallucination detection" in** LLMs (Pan et al., 2023). Further, a number of Action's proposers participate actively in standardisation bodies (e.g., W3C, ISO). It is expected that the Action will have an impact on the standardisation of KGs technologies through these proposers.

Socio-economic impact. According to a recent OECD report, 2021, only 8% of SMEs in Europe sell via e-commerce in other EU countries and an even lower share (4%) sell outside of the European Union borders. Language barriers are one of the causes of such low numbers. SMEs that sell their products and services internationally can leverage the benefits and opportunities generated by engaging in global markets.This Action will contribute in that direction by increasing the accessibility of SMEs to multilingual KGs that can be freely exploited, thus facilitating the internationalisation of the SMEs product catalogues and digital services as well as the free flow of their data and knowledge across borders. In particular, the Action aims to increase the number of SMEs using DBpedia in their products by 100% over the duration of the Action. Progress on this objective will be tracked by reporting new DBpedia "certified" users (i.e. users which explicitly confirmed use of DBpedia). Also, the public sector will take advantage of the Action's outcomes. Governmental data can be published as part of an open KG on the Web and enriched with knowledge from other sources such as DBpedia and/or other LOD datasets, to allow for improved and more transparent data-driven decision making.

Further, many technological solutions currently used by SMEs in Europe largely rely on global IT companies, usually from the USA, and on their proprietary knowledge graphs. As recognised by the EU, it is vital to reduce technological dependency from outside Europe, while at the same time allowing companies as well as researchers retain access to open knowledge as public good. This Action will promote precisely such principles. An example in that direction is the use of open knowledge in systems that perform fact checking or detection of bias. Some current solutions such as FactCheck (IDAFix) (Kalchgruber, 2018) partly relies on the proprietary Google Knowledge Graph, which limits transparency and accountability of the results. An alternative open-knowledge based system for truth discovery using DBpedia demonstrated 18% improvement compared to the base model (Beretta et al, 2018). The increasing use of LLMs makes the availability of an EU-based open general knowledge KG even more important, as it can be used to verify LLM output.

Potential innovations and breakthroughs. This Action aims for the creation of a critical mass of multilingual, cross-domain, linked data on the Web, using DBpedia as a cornerstone. On top of it, novel knowledge graph-aware services and methods will be developed. This envisaged scenario will enable new ways of scientific cooperation and access to information. Pan et al., 2023 describe a host of applications of general knowledge KGs, including mitigating LLM errors in digital healthcare, such as patient diagnosis, or refining LLM output when dealing with common sense questions, as common sense knowledge can be highly-culture specific and also biased. DBpedia, with its multiple independent yet interconnected language versions and content sourced mainly from curated sources, can help address this. While this Action does not explicitly aim to address significant challenges in digital healthcare, it aims to improve and maintain the underlying KG, which is a precursor and fabric for potential innovations and breakthroughs.

Overall, **the target state of the Knowledge Graphs technology addressed in this Action will enable further developments, less costs, better KG management and consumption, reduced and better quality checks.**

3.2. MEASURES TO MAXIMISE IMPACT

3.2.1. KNOWLEDGE CREATION, TRANSFER OF KNOWLEDGE AND CAREER DEVELOPMENT

Knowledge Creation and Transfer of Knowledge. The main contribution of the Action is to create high quality knowledge, encoded as a **knowledge graph**, for various domains and languages. The Action will focus on knowledge that has not been previously available in a machine-readable format. Various related

documents, such as specifications, user manuals and software, will be also created. The Action will also contribute to the creation of scientific knowledge, in terms of novel knowledge engineering and management methods which will be documented and published as scientific papers. One of the main goals of the Action is to foster **knowledge exchange** and **harmonize**, **align** and **integrate research efforts** across different language chapters. Relevant methods for knowledge extraction, data quality assurance, data governance and NLP, will be transferred across different language chapters in order to increase the overall quality (i.e. accuracy and coverage) of the available knowledge. The Action also aims at **transfer of the created scientific knowledge and technology** from academics to practitioners from various industrial domains, such as agriculture, libraries, news and media, life sciences, and cultural heritage, and to consulting stakeholders on it. The Action will support knowledge creation and knowledge transfer by organizing joint meetings among different WGs as well as meetings and training schools exclusively organized for the new DBpedia chapters.

Career Development. The Action will **support young researchers to extend and strengthen their research, development and innovation skills** in the knowledge graphs realm. The Action will **extend the profiles of existing jobs and create the foundation for new job profiles** in the context of open knowledge. This will enable young researchers to become open knowledge graphs engineers, knowledge quality assurance experts or knowledge integration experts. To aid career development, the Action will support the creation of learning resources and supporting tools for knowledge retrieval (SPARQL, Virtuoso), data integration (SILK, LIMES), quality assurance (RDFUnit), knowledge engineering (DBpedia Extraction Framework). For example, currently data journalists spend much of their time looking for sources of data and extracting these data from semi-structured sources, like web sites or proprietarily formatted governmental databases. The availability of structured, consistently formatted knowledge in DBpedia will not only increase efficiency of data journalists' work, but will also transform it. This will allow for a qualitative leap in the outputs of data journalism. On the other hand, this will require new skills on the side of "data journalists 2.0", such as knowledge of the SPARQL querying language.

3.2.2. PLAN FOR DISSEMINATION AND/OR EXPLOITATION AND DIALOGUE WITH THE GENERAL PUBLIC OR POLICY

Dissemination strategy. The goal of the Action's dissemination efforts is to establish the network as a worldwide focal point for academic and industrial parties interested in provisioning open knowledge graphs. In general, dissemination activities will consist of:

- **Informing user and developer communities** of the state-of-the-art developments taking place in the field via established communication channels (i.e. website, blogs, microblogs and news feeds) and establish groups in social and professional networks.
- Disseminating the results at **scientific and industry conferences and workshops** in the area of Semantic Web and Knowledge Engineering (ISWC, ESWC, EKAW, Web Conference, SEMANTiCS), NLP and Language Technologies (LREC, ACL, EACl, LDK, COLING), Data Mining and Machine Learning (SIGKDD, WKDD, ICML, ECML PKDD), and scientific journals (Semantic Web Journal, Journal of Web Semantics, Language Resources and Evaluation). Many of the considered venues provide opportunities for **collaboration with SMEs and big corporations**. In close interaction with relevant industry stakeholders, the Action will disseminate the results.
- Organizing **hackathons, workshops and conference** to disseminate and promote the results from the Action. The Action will **organize dedicated sessions and tracks for dissemination of the results with the participants coming from the industry**. The aim is to organize four workshops with co-located hackathon events and at least one Action dedicated conference.
- Participating in specific events such as **trade shows** or the like in order to reach a large number of companies.
- Organizing **tutorials and training schools** to educate target user groups in engineering, management and utilization of knowledge graphs. The Action will organize three training schools and tutorials.
- **Standardizing** the efforts wherever possible (W3C, ISO).

Exploitation of results. Over the last decade, the DBpedia knowledge graph has been extensively exploited by **academic and industrial users** in projects from different domains and languages. The Action is primarily focused on enriching the multilingual information of DBpedia and improving its data quality. Specifically, in terms of accuracy and coverage, which is a main prerequisite for its further exploitation and use in research and industrial contexts. The Action's network includes members from

domains such as food and agriculture, news and media, libraries, linguistics, life sciences and cultural heritage, which will exploit the Action's results in order to derive new knowledge, integrate internal (private) and public knowledge, publish knowledge or perform content enrichment. Some specific actions may include but are not limited to: the education sector, as researchers, educators, and student from diverse linguistic backgrounds will gain access to high-quality, multilingual resources; the life sciences sector can use DBpedia's enriched content to derive new insights and support their work in areas like genomics and drug discovery; similarly, cultural heritage organizations can leverage DBpedia to enrich their collections and enhance the visitor experience. The DBpedia knowledge graph is also widely exploited by **data scientists** in order to **discover novel information**, where the Action will introduce new challenges by expanding language support. The DBpedia knowledge graph has already been used by many scientists to **validate and evaluate data analysis methods**, and we expect that this Action will further strengthen this use of DBpedia.

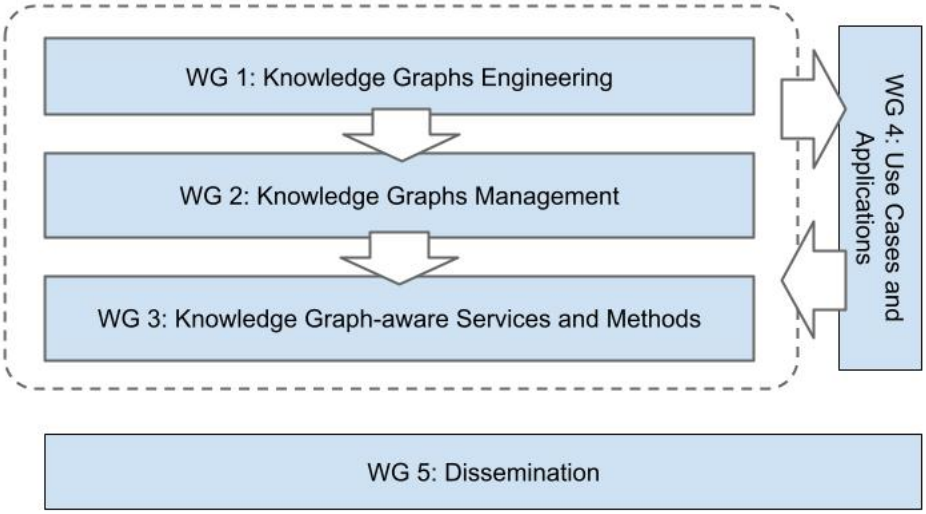
Communicating the results with the general public. To bring the results closer to the general public, the Action will be promoted in newspapers at universities, online news portals targeting specific users (farmers, librarians, musicians, sportsmen, citizens), but also at various events organized by related initiatives such as the Open Knowledge Foundation or the Wikimedia foundation. Also, non- technical users will be invited to dedicated DBpedia community meetings to get familiar with the available knowledge and supporting services.

4. IMPLEMENTATION

4.1. COHERENCE AND EFFECTIVENESS OF THE WORK PLAN

4.1.1. DESCRIPTION OF WORKING GROUPS, TASKS AND ACTIVITIES

The Action will be driven by five Working Groups (WGs), each focused on a particular aspect.



Working Group 1: Knowledge Graphs Engineering
Objectives. WG1 will lay foundations and develop best practices for creation, modelling, integration, enrichment and publishing of knowledge graphs.
Tasks Description **Task 1.1 KG modelling and ontology engineering.** This task deals with modelling and ontology engineering aspects for knowledge graphs representation. The task will study the evolution, coverage and quality of existing models, ontologies and vocabularies w.r.t. specific domains and languages. **Task 1.2 Knowledge extraction for KGs.** This task is concerned with knowledge extraction methods and their domain and language specific adaptation. It will also investigate the *use of LLMs* to extract structured information from text. **Task 1.3 Integration and interlinking of KGs.** This task will study the methods and tools for integration and interlinking of KGs at the schema and data level. **Task 1.4 Enrichment of KGs**. This task will study automated, semi-automated and manual (i.e. human-in-the-loop) techniques by which knowledge graphs are enriched from diverse sources. **Task 1.5 Publishing and release management of KGs.** This task will study principles, approaches,

protocols and best practices for efficient publishing of KGs.

Working Group 2: Knowledge Graphs Management

Objectives. WG2 will lay the foundations and develop best practices in data quality assurance, metadata, provenance, cleansing, exploration, mining and security aspects of knowledge graphs.

Tasks Description

Task 2.1 Data quality assurance of KGs. This task will deal with methods and approaches for data quality assurance of knowledge graphs.

Task 2.2 Metadata and provenance in KGs. This task will study how metadata and provenance are implemented in the context of knowledge, identify challenges and propose solutions.

Task 2.3 KG refinement. This task will focus on (semi-)automated data cleansing mechanisms for completion and correction of knowledge graphs.

Task 2.4 Knowledge exploration in KGs. This task will study existing knowledge exploration mechanisms for access and exploration of knowledge, which aid users to discover accurately while at the same time serendipitous and diverse information in knowledge graphs.

Task 2.5 Security and privacy in KGs. This task will study security and privacy for knowledge graphs and investigate the challenges and opportunities in the context of licensing, usage control, user authentication and authorization.

Task 2.6 Mining KGs. This task will examine novel techniques for mining open knowledge graphs. Graph mining, social network analysis methods and association rules mining methods will be exploited to discover novel knowledge from existing knowledge graphs.

Working Group 3: Knowledge Graph-aware Services and Methods

Objectives. This WG focuses on utilization of open knowledge graphs in order to enhance the performance of various methods and services and benefit from the growing amount of the open knowledge. The list of services and methods is not closed, and others might be added at the request of the Action members or interested communities.

Tasks Description

Task 3.1 KGs and Deep Learning. The aim of this task is twofold: (1) investigate how KGs can enhance deep learning methods by providing structured, semantically rich representations of concepts and relationships, and (2) explore how deep learning methods can enrich KGs by techniques for automatic extraction and linking of entities, concepts, and relationships from unstructured text data to populate and expand KG.

Task 3.2 KGs and LLMs. This task will investigate how knowledge graphs can support LLMs to boost the domain-specific reasoning, enhance interpretability, and mitigate biases and hallucinations.

Task 3.3 KGs and machine translation. This task will be concerned with the exploitation of knowledge graphs in machine translation systems.

Task 3.4 KGs and question answering. This task will examine how knowledge graphs can be integrated and exploited in question answering systems.

Task 3.5 KGs and information retrieval. This task will examine how retrieval augmented generation (RAG) approaches can be additionally injected with a knowledge graph resources to improve information retrieval, e.g. Web (semantic) search capabilities.

Working Group 4: Use Cases and Applications

Objectives. This WG will study possible use cases and applications of the knowledge graph technologies. The WG identifies a potential list of domains, however, the list will be dynamically adjusted according to the interests of the Action members.

Tasks Description

Task 4.1 Use cases in linguistics. By collaborating with linguists, lexicographers, and language researchers, this task aims to investigate and study specific applications of knowledge graphs such as for language preservation, dialect analysis, cross-lingual information retrieval, etc

Task 4.2 Use cases in news, media and social media. This task will be concerned with the role of knowledge graphs in the context of news, media, publishing and social media sectors, and in particular, how knowledge graphs can support the development of novel services and solutions. It will investigate how KG can enhance content recommendation systems, improve fact-checking or personalised content, etc.

Task 4.3 Use cases in finance and e-commerce. This task will study the use of KGs in the Fintec domain for data analytics, sentiment analysis or risks assessment, and the role of KGs in various business scenarios concerned with selling consumer goods and services. Moreover, it will investigate the role of KG in product recommendation and enhancing customer experience.

Task 4.4 Use cases in life sciences. Great majority of life science branches are collecting and making use of a vast amount of data assets. This task will study how knowledge graphs can deeply influence studies in life sciences allowing to efficiently integrate multiple sources, model knowledge spaces or support research across different life science branches. This task will investigate for example drug discovery through the analysis of the new enriched information in DBpedia.

Task 4.5 Use cases in social sciences. Knowledge graphs can significantly support the social sciences branches such as anthropology, education or geography. This task will investigate how knowledge graphs can support social sciences for a more in-depth understanding of the societies, by investigating cultural trends and geographic patterns and the overall behaviour of the system.

Working Group 5: Dissemination

Objectives. This WG will coordinate the dissemination activities as well as the training, stakeholder engagement and the alignment with other projects and communities.

Tasks Description

Task 5.1 Training and capacity building. This task is dedicated to coordination of the STSMs, tutorials, workshops, hackathons and training schools.

Task 5.2 Dissemination, community building and cross-fertilization. Dissemination of the project results, supported by community building and cross-fertilization activities, such as creation of a website, blog, online community groups and press releases, to ensure scientific communication of the results. This task will also include the coordination and monitoring of the Action's publications.

Task 5.3 Stakeholder selection and engagement. This task is dedicated to engagement and selection of relevant stakeholders. The task will also ensure that the effort is equally distributed to the different stakeholders' fields. For instance, this task aims to initiate focused dialogues with academic researchers, governmental bodies, and industrial partners to pinpoint their specific needs and expectations. By establishing channels for open communication, conducting targeted surveys, and facilitating their participation in relevant events and conferences, we aim to maintain a high level of engagement. Stakeholder engagement will be a continuous process, which may involve collaborative workshops, periodic consultations, and feedback mechanisms.

Task 5.4 Orchestration with other projects and industry partners. To maintain these synergies, this task aims to engage in reciprocal activities such as co-organized workshops, joint conferences, and collaborative research efforts. Moreover, the task targets to explore partnerships with complementary projects, creating opportunities for mutual support. These partnerships may involve data sharing, co-development of tools and methodologies, and shared knowledge transfer.

4.1.2. DESCRIPTION OF DELIVERABLES AND TIMEFRAME

WG 1: Knowledge Graphs Engineering • Guidelines and best practices on engineering KGs (M24, M48) • Intermediate and final activity reports (M24, M48) • Scientific papers on engineering KGs (M48)
WG 2: Knowledge Graphs Management • Guidelines and best practices on KGs management (M24, M48) • Intermediate and final activity reports (M24, M48) • Scientific papers on KGs management (M48)
WG 3: Knowledge Graph-aware Services and Methods • Guidelines and best practices in use of KGs in services and methods (M24, M48) • Intermediate and final activity reports (M24, M48) • Scientific papers on KG driven services and methods (M48)
WG 4: Use Cases and Applications • Requirements specifications and use case/application landscape (M24) • Intermediate and final activity reports (M24, M48) • Scientific papers on use cases and applications of KGs (M48)
WG 5: Dissemination • Action Vision (M6) • Report on stakeholder selection and engagement (M36) • Report on dissemination activities and dissemination materials (M48) • Materials (slides, manuals, code) generated during the training schools (M24, M36, M48)

Milestone	Month	Means of verification
M1	6	First MC meeting organized, working groups established, website deployed, 2nd MC meeting organized, first set of WG meetings, Action vision document.
M2	12	First workshop organized including a hackathon session, first set of STSMs.
M3	18	First summer training school on open knowledge graphs organized, 3rd MC meeting organized, WG meetings.
M4	24	Second workshop organized including a hackathon session, second set of STSMs, WG meetings, intermediate reports by WGs, KG guides and best practices. 4th MC meeting organized.
M5	30	Second summer training school on open knowledge graphs organized.
M6	36	KG in Action conference organized, third workshop organized including a hackathon session, third set of STSMs, WG meetings, report on stakeholder selection and engagement, 5th MC meeting.
M7	42	Third summer training school on open knowledge graphs organized.
M8	48	Final Action workshop organized including a hackathon session, final set of STSMs, WG meetings, final reports by WGs, dissemination report, updated KG guides and best practices. Final MC meeting.

4.1.3. RISK ANALYSIS AND CONTINGENCY PLANS

Risk	Mitigation measure

Loss of a member Probability: Low Impact: Low	All the Action proposers have been involved in the proposal preparation and have shown their commitment to working together. Moreover, individual groups of proposers have been/are working together in other projects and a foundation of trust and common understanding has been established. These collaborations are longer lived than Action will be.
Late delivery of Action outputs Probability: Medium Impact: High	The Action has a strong management model with emphasis on clarity and efficiency. All outputs will be tracked from six weeks before their due dates, and any additional resources needed to meet deadlines will be deployed by the Action management team in a timely manner.
The Action does not succeed to reach a critical mass of relevant stakeholders Probability: Low Impact: High	The size and the diversity of the network reduces the risks. The stakeholders are well known and the interest in knowledge graphs is high.
DBpedia does not accept tools and methods proposed by the Action Probability: Low Impact: High	The network of proposers consists of researchers directly involved in the development of DBpedia and the Action has received strong support from DBpedia Association. This assures that the technology developed during the course of the Action will be accepted by DBpedia. Of course, extensive testing will be executed before deployment of the technology in production environments.
Lack of knowledge for specific languages. Probability: Medium Impact: High	The Action's network spans over many COST associated countries and languages which reduces the risk.
Complexity of language and domain specific adaptation is too great Probability: Low Impact: Medium	The network of proposers is highly experienced in development of methods and services for various languages and domains. The network provides expertise in different languages and domains which will ease the adaptation process.
Too many use cases ("Jack of all trades master of none") Probability: Low Impact: Medium	The current list consists of five use cases which have been selected based on the current state of the community in the field and based feedback from the network of proposers. The use cases, although different, require similar tools and methods. Thus, the effort put in the Action will be beneficial for all the use cases.

4.1.4. GANTT DIAGRAM

Activities	**Year 1**				**Year 2**				**Year 3**				**Year 4**			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Milestones		M1		M2		M3		M4		M5		M6		M7		M8
Meetings																
MC meeting	X	X			X			X				X				X
WG meeting		X			X			X				X				X
Events																
Training schools						X				X				X		
Workshops				X				X				X			X	
KG in Action conf												X				
STSMs																
All WGs																

REFERENCES

[Ferrucci et al, 2012] "Introduction to “This is Watson”." IBM Journal of Research and Development 56.3.4 (2012): 1-1.

[Gartner, 2018] Gartner Identifies Five Emerging Technology Trends That Will Blur the Lines Between Human and Machine, URL: https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine

[Pan et al., 2023] "Large Language Models and Knowledge Graphs: Opportunities and Challenges." arXiv preprint arXiv:2308.06374 (2023).

[LOD Stats, 2023]. The Linked Open Data Cloud, URL: https://lod-cloud.net

[Tanon et al, 2020] Pellissier Tanon, T., Weikum, G., & Suchanek, F. (2020, May). Yago 4: A reason-able knowledge base. In European Semantic Web Conference (pp. 583-596). Springer, Cham. URL: https://link.springer.com/chapter/10.1007/978-3-030-49461-2_34

[Ismayilov, 2018] "Wikidata through the Eyes of DBpedia." Semantic Web 9.4 (2018): 493-503.

[Navigli et al, 2021] Navigli, R., Bevilacqua, M., Conia, S., Montagnini, D., & Cecconi, F. (2021). Ten Years of BabelNet: A Survey. In IJCAI (pp. 4559-4567). URL: https://www.ijcai.org/proceedings/2021/0620.pdf

[Hertling et al, 2020] Hertling, S., Paulheim, H. DBkWik: extracting and integrating knowledge from thousands of Wikis. Knowl Inf Syst 62, 2169–2190 (2020) https://doi.org/10.1007/s10115-019-01415-5 URL: https://link.springer.com/article/10.1007/s10115-019-01415-5

[Ji et al, 2022] Ji, S., Pan, S., Cambria, E., Marttinen, P., & Philip, S. Y. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. IEEE Transactions on Neural Networks and Learning Systems, 33(2), 494-514. URL: https://ieeexplore.ieee.org/abstract/document/9416312

[OpenDataPortal report, 2020] European Union Open Data Portal, URL: https://www.europeandataportal.eu/en/using-data/benefits-of-open-data

[Zaveri et al, 2016] "Quality assessment for linked data: A survey." Semantic Web 7.1 (2016): 63-93.
[OKF] Open Knowledge Foundation, main website, URL: https://okfn.org

[Pereira et al, 2017] "Linked data in Education: a survey and a synthesis of actual research and future challenges." IEEE Transactions on Learning Technologies11.3 (2017): 400-412.

[Kalchgruber et al, 2018] "Factcheck-identify and fix conflicting data on the web." International Conference on Web Engineering. Springer, Cham, 2018.

[Beretta et al, 2018] Combining Truth Discovery and RDF Knowledge Bases to Their Mutual Advantage. In: Vrandečić D. et al. (eds) The Semantic Web – ISWC 2018. ISWC 2018. Lecture Notes in Computer Science, vol 11136.

[OECD report, 2021] The Digital Transformation of SMEs, URL: https://www.oecd-ilibrary.org/industry-and-services/the-digital-transformation-of-smes_1386638a-en