Building a Knowledge Graph on Innovation in Horizon Europe Projects

Adriana AnaMaria Davidescu[1,†], Vasile Alecsandru Strat[1,*,†], Madalina Zurini[1,*,†] and Dumitru Roman[1,2,†]

[1] *Bucharest University of Economic Studies, Bucharest, Romania*
[2] *SINTEF AS, Oslo, Norway*

Abstract

Innovation ecosystems are complex networks of entities and relationships that drive research and development. Both Horizon 2020 (H2020) and Horizon Europe (HEU) promote innovation at the EU level by financing collaborative research projects. To achieve effective ecosystem mapping and analysis, we propose the creation of a Knowledge Graph centred around innovation in H2020/HEU projects. This Knowledge Graph aims to demonstrate relationships between universities, start-ups, consortia, publications, researchers, and AI companies across the projects as an intricate web of interactions. These patterns and insights can be used for policy-making and strategic decision-making at large. Five different use cases will be developed on top of the Knowledge Graph to map seamlessly and integrate them with relationship mapping between universities involved in certain projects with startups, relationship mapping between startups and publications or researchers in the same field of study within a particular consortium under given projects, etc. Other use cases can be developed when other sources are added into the Knowledge Graph so that it provides holistic views about how different parts fit together across Europe towards achieving common goal around innovation through H2020/HEU programmes.

Keywords

Innovation knowledge graph, big data, horizon projects

1. Introduction

Innovation ecosystems are intricate systems comprised of different entities and relationships that foster the creation, development, and commercialisation of new technologies and ideas. These systems are dynamic in nature, involving universities, research institutions, start-ups, and established companies where policymakers and investors interact. Better understanding these interactions can be a baseline for more innovations with tangible economic and social benefits.

Horizon 2020, alongside its successor program Horizon Europe, have been instrumental in promoting innovation across the European Union. Through such massive funding schemes, collaborative research projects spanning various fields have been supported, which emphasise international cooperation towards tackling global complex challenges through multidisciplinary approaches. These programs do not stop at immediate results obtained from scientific investigations; they go further to influence policy-making processes, thereby impacting strategies formulated around innovations besides giving birth to new sectors within economies.

RuleML+RR'24: Companion Proceedings of the 8th International Joint Conference on Rules and Reasoning, September 16--22, 2024, Bucharest, Romania
* Corresponding author.
† These authors contributed equally.
adriana.alexandru@csie.ase.ro (A. AM. Davidescu); vasile.strat@bbs.ase.ro (V. A. Strat); madalina.zurini@csie.ase.ro (M. Zurini); dumitru.roman@sintef.no (D. Roman)
0000-0002-9847-1200 (A. AM. Davidescu); 0000-0002-5074-1419 (V. A. Strat); 0000-0001-9872-8566 (M. Zurini); 0000-0001-6397-3705 (D. Roman)

In this paper we propose a mapping of the H2020/HE innovation ecosystem through the development of a Knowledge Graph that could capture the intricate relationships in this ecosystem.

2. Knowledge Graph for Mapping Innovation Ecosystems

The proposed Knowledge Graph aims to allow detailed mapping and examination of interconnections within the H2020/HE innovation ecosystems. This Knowledge Graph will serve as a holistic instrument to visualise and understand the relationships between key players in funded projects, such as universities, startups, consortiums (groups), publications, researchers, and companies.

Another expectation from this Knowledge Graph is to show how these entities work together in many ways by sharing resources among themselves, thereby leading to an innovative life cycle. It is envisioned that this Knowledge Graph will enable detection of patterns or trends that may not be obvious through traditional analysis methods while integrating various data sources using advanced AI techniques. Therefore, we can learn more about where innovation takes place faster within such settings.

In terms of the methodological aspects of constructing the Knowledge Graph, one first needs to create an ontology that defines entities (universities, startups, researchers) and their relationships (e.g., collaborations, project participation, publications). This ontology acts as a base for the Knowledge Graph; it ensures all data points go into the correct categories and are related to each other accordingly. To accommodate different data sources, an ETL mechanism has to be devised to ingest various types of data such as structured (e.g., databases), semi-structured (e.g., JSON, XML) and unstructured (e.g., text from publications) into the Knowledge Graph. The data sources should consist of, but not limited to, databases such as the Horizon Europe projects database or academic publication repositories such as Scopus and Web of Science, among others containing scholarly records, as well as university rankings plus business registries. Therefore, data used here must be put into a uniform format before being fed into the graph. Entities in the Knowledge Graph, such as university start-up projects, are expressed as nodes, whereas edges represent their interactions/relationships between them. This means showing how those entities interact based on their roles relative to the Horizon projects in which they participate.

Knowledge Graphs with causal inference models allow an understanding of cause-effect relationships between entities. Such models can reveal how changes in funding affect project success or show that certain patterns of collaboration lead to more innovative outcomes. Large queries over big datasets should be allowed for quick execution using, e.g., Neo4j or Amazon Neptune, which are meant for such tasks and can scale up as more information becomes available. Five different use cases will validate the usefulness of the Knowledge Graph by looking at various parts of Horizon's innovation ecosystem.

The Knowledge Graph will be used to map innovation ecosystems, showing how different entities work together and contribute to innovations in an interconnected manner. It will integrate diverse data sets from various sources and apply various analytics techniques to uncover the complex processes underpinning research and development within Horizon programs.

In the following we describe the five use cases, each utilising the Knowledge Graph to address specific aspects of innovation within Horizon programs.

2.1. Adapting the Gravitational Model for EU Universities in Horizon Projects

This first use case will adapt the traditional gravity model from physics to investigate the network of universities within Horizon projects, measuring the so-called "attraction" towards

universities by projects. This gravitational model includes information related to project funding, university reputation, and the size of collaborative networks. From a methodological perspective, this use case will make use of regression and network analysis methods and shed light on the roles and influence played by universities within the broader innovation system. The regression analysis will investigate how these factors correlate with the success rate and frequency of participation in various projects. Meanwhile, a network analysis will be conducted to map and study collaborations between the universities and other entities within Horizon's boundaries, helping in identifying major players along with central nodes in the network and understanding the roles played by different universities in driving innovations forward. The Knowledge Graph will make use of various data sources, combining information related to horizon project databases, academic publication repositories such as Scopus or Web of Science, university ranking databases, and funding bodies, among others. Additionally, a temporal dimension will be added, capturing developments over time concerning university participation in Horizon projects, thereby indicating how shifts in funding/reputation influence them.

2.2. Identifying Relationships Between the Start-Up Ecosystem and Horizon Projects

This use case will analyse the relationship between start-up ecosystems and Horizon projects, specifically start-ups from universities. The Knowledge Graph will contain links showing which founding universities have participated in any of their respective projects. Logistic regression analysis and cluster analysis will be utilised to uncover important relationships and patterns of collaboration. In this use case, logistic regression will be applied to identify significant predictors of the startup's involvement in Horizon projects, such as the reputation of the founding university, the sector focus of a startup, the level of innovation or previous connection with academic institutions. Using cluster analysis, it will be possible to group startups based on their characteristics and project participation, which could indicate successful collaboration strategies among them.

The Knowledge Graph will be useful for mapping out the relationships between startups – both founders' universities and participation in Horizon projects; this relational mapping can show areas where academia-industry linkages are weak or strong. For this purpose, data from start-up registries, university spin-off databases together with records from the horizon project will be employed, while financial reports could provide additional information sources, including funding rounds or growth metrics within a startup firm over time, amongst other sources if available also should not be left out since they might give us some insights we did not know about before. Furthermore, this information is cross-referenced against their performance after joining any Horizon initiatives.

2.3. Analyzing the Relationship Between ERASMUS Mobility Patterns and Horizon Consortiums and Publications

Research mobility is one of the most important factors contributing to global partnerships and the sharing of knowledge among scholars. This use case aims to explore the link between European researchers under the Erasmus scheme and the Horizon-funded consortium and their resultant publications through various international academic journals. To prove this hypothesis, econometric tools such as panel data analysis, fixed effects models, and social network analysis will be used within the Knowledge Graph. Therefore, it is important to evaluate this approach's impact on research productivity and collaborative network formation in terms of mobility.

Social Network Analysis (SNA) will be used to visualise and quantify the collaboration networks formed by ERASMUS mobility. To understand their structure and efficiency, this method will identify central figures and key hubs within such networks. Records from the

ERASMUS+ mobility program, Horizon project participation records, and academic publication databases will all be combined with a view to giving an account of these scholars' profiles, including their history of movements, publications made so far, and projects involved.

2.4. Identifying the Relationship Between AI Company Clusters and Collaboration in Horizon Research Projects

This use case will investigate the clustering of AI companies around universities within Horizon research projects. The Knowledge Graph will map AI companies and their collaborations with universities, enabling the use of spatial econometrics and Geographic Information System (GIS) tools in identifying clusters and analysing them. The aim here is, therefore, to understand how innovation outcomes are affected by geographical proximity together with clustering around innovation hubs in AI-related fields. Analysis for identification of active projects engaged artificial intelligence firms shall involve the use of cluster analysis. This entails looking at the density of firms within a region together with the intensity of their interaction with academic institutions. Business registers, project datasets and geospatially referenced data sets will provide raw information for this analysis. Information about the location of the company, size, specialisation and history of collaboration with universities drives this analysis. Spatial relationships between clusters [1], including proximity to innovation hubs, availability of skilled labour, and access to funding, should be investigated in the present study.

2.5. Pattern Identification of European Researchers' Participation in Horizon Projects and Start-up Founding

The Knowledge Graph will enable linking researchers with their project participations as well as start-up creation from which survival analysis or latent class analysis will be utilised to identify patterns. This collaboration between the participants with Horizon projects is also research in papers such as [2], where the authors apply several clustering techniques for determining, among others, the willingness to share information under legal confidentiality. This use case will discuss different career paths for scientists, their role in innovative processes, and how entrepreneurial activities are influenced. We will analyse the rate at which researchers transit into entrepreneurship after conducting a survival analysis of the time taken from participating in such projects up to founding a start-up company. By utilising latent class analysis, we can uncover various segments among scientists based on Horizon project participation, formation of startups and other critical milestones achieved during their careers.

This use case will link information related to researcher profiles within the Knowledge Graph, including publication records, project involvement, and entrepreneurial activities. Data from academic databases, business registries, and patent filings will also be utilised.

3. Outlook

Building a Knowledge Graph based on various use cases linking various data sources can enable a better understanding of the H2020/HE innovation ecosystem at the European level. This Knowledge Graph facilitates mapping out an innovative landscape across various Horizon programs. Visualising network relationships and challenging assumptions could provide strategic insights relevant to policymaking upon which institutions act, revealing key drivers for successful innovation systems to policymakers where current provisions are not meeting needs or enabling the creation of specific interventions meant to foster collaboration required for realising an innovative outcome.

Additionally, it will enhance collaboration between universities, start-ups and established firms, thereby boosting the overall impact of research and development endeavours. With a

deeper understanding of how dynamic these innovation ecosystems are working, this project aims to provide better strategies for fostering innovations.

Acknowledgements

This research is financed under the Romanian National Recovery and Resilience Plan, by the Romanian Government, under the contract number 268/29.11.2022, Entitled “CAUSEFINDER-CAUSALITY IN THE ERA OF BIG DATA”.

References

[1] Colella, Fabrizio and Lalive, Rafael and Sakalli, Seyhun Orcan and Thoenig, Mathias, ”Inference with Arbitrary Clustering” IZA Discussion Paper No. 12584 (2019): 2-32

[2] Benissa, Takwa, and Patil, Anish, “Drivers for Clustering and Inter-Project Collaboration – A Case of Horizon Europe Projects”, Administrative Sciences, Vol. 14, Issue 5 (2024), doi: 10.3390/admsci14050104