**Common Action 2021 Quantum Hackathon Challenge**

**KGs for ESG**

**V1.1**

**Introduction**

Pressing environmental, social and governance (ESG) challenges present new, data-intensive requirements to investors, companies, and innovators alike.

This burgeoning market is large and rapidly growing, with a third of global Assets Under Management earmarked for ESG investments by 2025, and sweeping environmental regulation swiftly being passed, affecting firms across Europe and in Asia. This growth presents many new opportunities, as well as an increased demand for reliable data about companies’ performance. Though many are jumping into the market with data innovations, few are based on robust, transparent foundations. This results in a mix of conflicting metrics that reduce speed, efficacy, and clarity for market participants.

In addition to this chaos, ESG metrics present flat data: that is, numerical data and true/false criteria. The simplicity of this data presentation hides opportunities for innovation. Fortunately, new advances in computer technologies make it possible to transcend the limitations of traditional, labor-intensive scoring methodologies. In particular, the “knowledge graph” technology can be used to provide a rich, contextual view, by connecting data from disparate sources and providing information about the semantics of the elements and their relationships. As a result, this technology allows users to gain richer context about the topics in question.

The Common Action (CA) Impact Demo knowledge graph connects information about over 1900 companies--including their industries, profits’, environmental impact on various parameters, and country of headquarters--and combines that with information about impact, through country development data points. Information from multiple data sets are connected to the United Nations Sustainable Development Goals, which are a collection of 17 interlinked global goals set up in 2015 by the United Nations General Assembly and are intended to be achieved by the year 2030. Countries are the sites where SDG-related phenomena occur, and companies’ actions have tremendous bearing on both positive and negative outcomes.

The remainder of this document discusses the concepts behind knowledge graphs, how the Common Action graph was created and how it can be used.

**Knowledge Graph Technology**

Before diving into the specifics of knowledge graphs, it is valuable to overview the concepts of “ontologies” and “graphs”. The following sub-sections discuss these terms and then knowledge graphs specifically.

**Ontologies**

Much time and effort have gone into defining “what is an ontology?”[[1]](#footnote-1) But these definitions are much too complicated. … An ontology can be defined as:

A partial, simplified description of the categories of things in a domain, along with their attributes and relationships, specified in a formal way and created by a community of users for an explicit purpose

Unpacking the concepts in the definition, it is important to highlight that:

* One of the most important goals of an ontology is to communicate the concepts and knowledge (and increase the understanding) of the domain within the “community of users”, which also enables sharing and reuse of the knowledge encoded using the ontology
* “Partial, simplified description” requires understanding and describing the domain as *simply* as possible, while still addressing the “explicit purpose” of the ontology’s “community”
  + As mentioned by Katariina Kari in the “Meet the Mentors” session, an ontology’s “purpose” is made explicit by a set of *competency questions* that will be used for querying the ontology’s concepts and data
* Defining an ontology in a “formal way” does NOT mean that it uses a complex logical form, a programming or standard language (such as RDF or JSON), or specific tooling
  + What is required are a set of design criteria and rules that apply to all concepts in the domain (such as having agreed definitions and examples for the concepts, or always defining the types of entities that can participate in a relationship)
  + Defining an ontology using a machine-processible language is necessary IF the ontology is to be consumed by both humans and machines

**Graphs**

A graph is a set of “nodes” (vertices) that are connected by labeled or unlabeled “edges” (relationships or associations). In this discussion, the nodes are entities (instances of the categories of things in our ontology) or values (such as string or integer values), and the edges are the attributes and relationships that connect them.

**Knowledge Graphs**

A knowledge graph is a collection of nodes interconnected by named edges (identifying properties and relationships). The structure of the graph is defined by an ontology, which formally specifies and defines the semantics of the possible nodes and edges. The “knowledge” that is captured comes from the specific data that is rendered using the ontology and comes from multiple sources, in multiple formats, with multiple levels of specificity.

Knowledge graphs enable analysis of the data as an integrated whole. When implemented using graph database tooling, they combine features of standard, relational databases (such as structured query), with the flexibility and extensibility of graphs. In addition, graph analysis (such as centrality, clustering and associativity analyses) can be performed on the integrated data. Backing the graph by an ontology provides formal definition and structure, which provides for increased precision and accuracy when fusing and integrating data from various sources.

The figure below shows an example of a knowledge graph for the company, MAPFRE SA, from the Impact Demo data. (Additional labels have been added since the text is quite small.) Note that this is just a small amount of the available data since there are also 10 instances of Measurements for the 13 other types of impacts, as well as all the available information on the country of Spain, where MAPFRE is based.

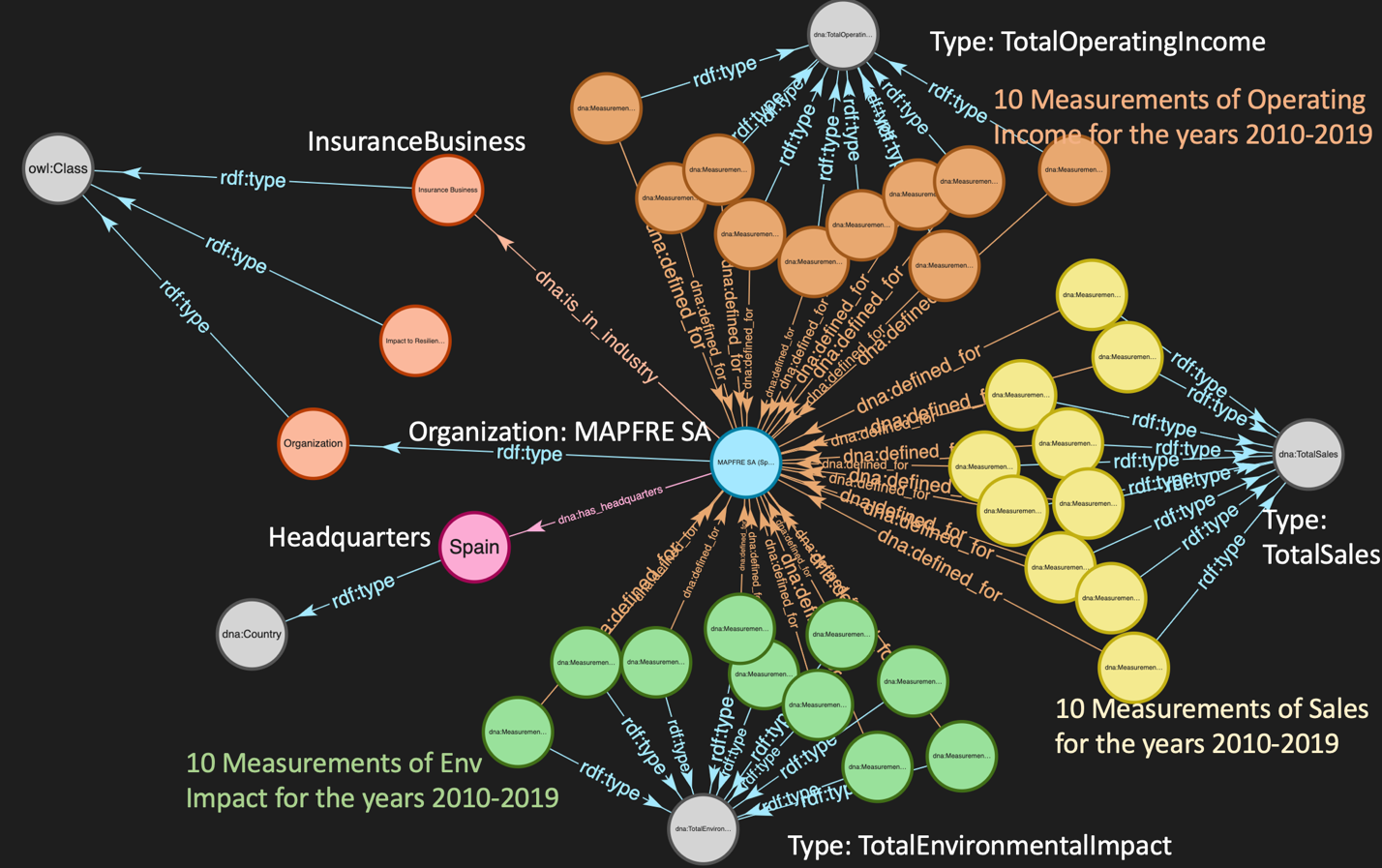


Figure 1. Portion of the CA Impact KG

## **Connecting company and country data to UN SDGs:**

The table below presents a layout of the connection points within the CA knowledge graph, describing descriptors for corresponding company and country data, and the related UN SDG goal. For more information about the SDG goal, please refer to the UN website dedicated to each (linked below).

|  |  |  |
| --- | --- | --- |
| **Company Data** | **Country Data** | **UN SDG** |
| Impact to Habitat and Biodiversity | Total Area; Total Land Area; Environmental Issues (text); Amount of Forest | [Life on Land](https://sdgs.un.org/goals/goal15)  [Life below Water](https://sdgs.un.org/goals/goal14) |
| Impact to Resilience to Events | Total Electrical Production and Consumption; Electricity from Fossil/Nuclear Fuels and Renewable Sources; Crude Oil Production and Reserves; Natural Gas Production, Consumption and Reserves | [Climate Action](https://sdgs.un.org/goals/goal13) |
| Impact of Adverse Compounds (e.g., greenhouse gases, VOCs, NOx, SOx, particulate matter, miscellaneous compounds) | Adverse Compound Emissions (CO2, methane and particulate emissions) | [Life on Land](https://sdgs.un.org/goals/goal15)  [Life below Water](https://sdgs.un.org/goals/goal14)  [Climate Action](https://sdgs.un.org/goals/goal13)  [Responsible Consumption and Production](https://sdgs.un.org/goals/goal12) |
| Impact to Reduction of Disease and Mortality | Health Care Assessment | [Good Health and Well-Being](https://sdgs.un.org/goals/goal3) |
| Impact to Access to Food | Assessment of Food Insecurity | [Zero Hunger](https://sdgs.un.org/goals/goal2) |
| Impact to Agricultural Productivity | Amount of Arable Land, Permanent Crops and Permanent Pasture Lands | [Zero Hunger](https://sdgs.un.org/goals/goal2) |
| Impact to Potable Water and Sanitation | Amount of Urban/Rural Population Lacking Potable Water; Amount of Urban/Rural Population Lacking Sanitation | [Clean Water and Sanitation](https://sdgs.un.org/goals/goal6) |
| Water Use Impact | Amount of renewable water resources | [Clean Water and Sanitation](https://sdgs.un.org/goals/goal6) |
|  | Municipal Waste Generated and Recycled | [Responsible Consumption and Production](https://sdgs.un.org/goals/goal12) |

The Measurements in the first and second columns are contained within the knowledge graph, as well as data on several additional topics including economics. An outline of all topics, and explanations and visualizations of the backing ontology are included in Appendix 1.

**Hackathon Challenge**

In this challenge, you will explore the reality of company environmental impact with the help of a knowledge graph. Using Common Action’s Venture Impact Canvas[[2]](#footnote-2), included in Appendix 2, as a research and design guide, you are challenged to create or re-design an impact business model for one of the following:

* Redesign an existing company’s business model to better direct impact investment or reduce impact
* Design a business model for a new startup, service, or tool that assists companies to track or manage or improve performance in an impact area
* Produces a better option for the industry/market at large
* Supports a country’s needs for development.

Creative ideas that do not follow one of the above prompts are also welcome.

## **Starting Questions**

A subset of possible starting questions and KG queries are included below.

You are also welcome to think of your own analytical approaches through research in the supplied knowledge graph, [at the UN SDG website](https://sdgs.un.org/goals), and within supplementary data sources such as financial databases and the Web.

Starting questions:

* What are the main SDG challenges that a country is facing? How can the market help to address them?
* How can needs for electrification and climate action be balanced?
* What natural resources are present: arable land, clean water, habitats and biodiversity, etc.? What is the status of these resources? How can the market address environmental challenges and augment solutions?
* What UN SDG goals are particularly relevant for an industry? How can a company in that industry increase its contribution to meeting those?
* How do the economic and environmental conditions of a company’s headquartered country affect meeting the UN SDG goals? Can differences in environmental impact be seen for companies in the same industry across countries?
* How can a company identify and reduce their impact in a country?
* How can a company and/or startup support the country to meet its development goals?

## **Appendix 1: About the Data**

The Company Data included in the knowledge graph provides calculated dollar amounts (in US Dollars) for each unit of environmental impact[[3]](#footnote-3). This data is extracted from a spreadsheet holding a portion of the data from the footnoted study.

The Country Data, on the other hand, provides specific area, volume, power, percentage and other statistics for the region. This data is assembled by web scraping the pages of the CIA World Factbook[[4]](#footnote-4).

All the statistics that may be extracted for an Organization or Country are specified in the file, Measurements.xlsx. Several examples are shown below.

|  |  |
| --- | --- |
| Type of Measurement | Definition |
| AmountOfArableLand | The percentage of a Country's land area that is used for the cultivation of crops that are replanted after each harvest (such as corn or wheat). The Country is specified using the :defined\_for property, and the year for which the percentage is calculated is indicated by the :reported\_year property. The Measurement's value can be compared to a Country's :land\_area\_sq\_kms property value to understand the actual amount of arable land. Note that the Measurement values are given in percents, % (as defined by the :has\_unit property). |
| AvgSalesByIndustry | The average sales for the industry indicated by the :has\_industry\_type property, for the Country specified by the :defined\_for property, as reported for the year indicated by the :reported\_year property. The Measurement's :reported\_value can be compared to an Organization's TotalSales value to understand the relative size of the company in the specified industry and country. Note that Sales values are given in US Dollars, USD (as defined by the :has\_unit property). |
| ElectricityFromFossilFuels | The percentage of the electrical generating capacity (given by the :reported\_value of the :ElectricityGeneratingCapacity Measurement) associated with the burning of fossil fuels, in the Country specified by the :defined\_for property, as reported for the year indicated by the :reported\_year property. Note that the Measurement values are given in percents, % (as defined by the :has\_unit property). |
| FoodInsecurityAssessment | An assessment (from 'Negligible or None' to 'Very High') of the food insecurity of the Country specified by the :defined\_for property, as reported for the year indicated by the :reported\_year property. The string defining the assessment is specified using the :assessment property. If food insecurity is severe, but only in certain areas of the country, then the :localized property will be set to True. |
| ImpactToAccessToFood | The calculated impact of the Organization's operations (where the Organization is indicated by the :defined\_for property) on the availability of sufficient, nutritious food, as reported for the year indicated by the :reported\_year property. Note that values are given in US Dollars, USD (as defined by the :has\_unit property). Also, note that this Measurement directly relates to the 2nd UN Sustainable Development Goal (SDG2, 'Zero Hunger') and its sub-goals, 2.1 ('By 2030, end hunger and ensure access by all people, in particular the poor and people in vulnerable situations, including infants, to safe, nutritious and sufficient food all year round.') and 2.2 ('By 2030, end all forms of malnutrition, including achieving, by 2025, the internationally agreed targets on stunting and wasting in children under 5 years of age, and address the nutritional needs of adolescent girls, pregnant and lactating women and older persons.'). Information on SDG2 is found at https://www.un.org/sustainabledevelopment/hunger/. |
| TotalSales | The yearly sales for the Organization indicated by the :defined\_for property, as reported for the year indicated by the :reported\_year property. The Measurement's reported\_value allows taking profit margins into account, when TotalSales are compared with TotalOperatingIncome. Note that the values are given in US Dollars, USD (as defined by the :has\_unit property). |

Table 1. Country and Company Data

The sub-sections below summarize the data contained in the knowledge graph, related to both companies and countries. Also included are visualizations of the backing ontology for the knowledge graph.

Note that although the two data sources used in this demo are quite different, their information is semantically aligned, and identically encoded. All the extraction and graph generation code that was used is available as an open-sourced Jupyter notebook[[5]](#footnote-5).

## 

## **Company Information:**

## For companies, the knowledge graph contains data specifying:

* The industry of the company
* The country where the company is headquartered
* Total operating income and sales in a given year
* Total environmental impact, which is a composite score based on all contributing environmental and social issues that are related to that company's individual activities
* Individual impact measurements

## 

Figure 2. Visualization of the Company/Organization Knowledge

## **Country Information:**

For countries, the knowledge graph shows a variety of economic data (GDP, unemployment rate, inflation rate, …), land use data (total land area and the amount of arable land, forest, …), information on waste generated and recycled, information on electrical production and consumption as well as the production and consumption of natural resources such as crude oil and natural gas, data on CO2, methane and particulate emissions, data on the percentages of the population in poverty and without potable water or sanitation facilities, and much more.

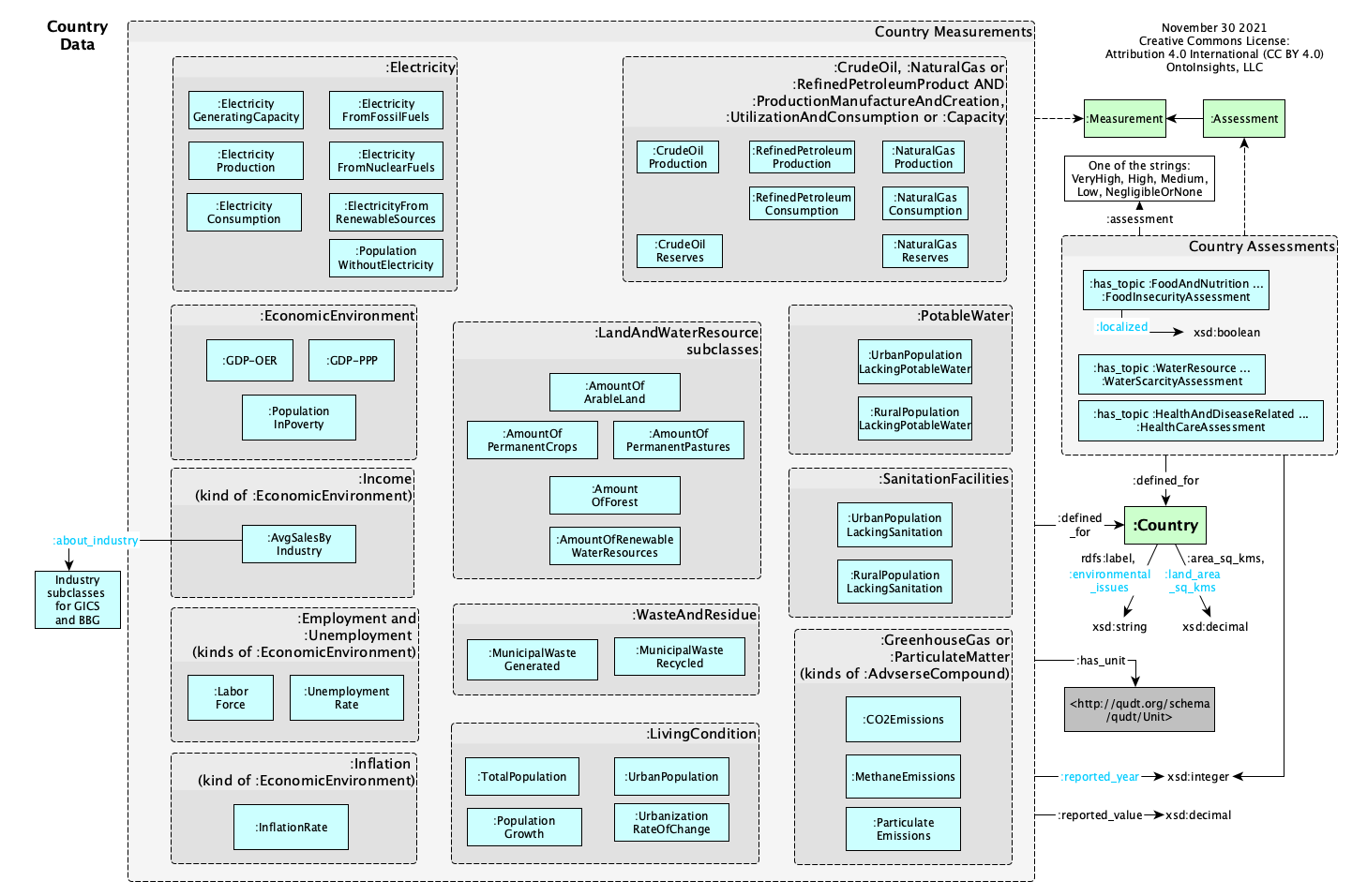
****

Figure 3. Visualization of the Country Knowledge

**Appendix 2: The Venture Impact Canvas**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Key Partners**  *Who are the venture’s suppliers?*  *Which key resources is the venture acquiring from our partners?*  *How do the venture’s partners source/produce those resources?* | **Key Activities**  *What resource-intensive activities does the venture perform (e.g., manufacturing, distribution), and at what quantity?*  *What is the venture’s waste strategy?*  *What is the overall impact of the venture’s activities?* | **Value Propositions**  **Company***:*  *What does the venture offer the market and communities they serve? What problem(s) do they solve?*  **ESG/SDG:**  *How can the venture provide value to support SDG and ESG outcomes?* | **Relationships**  *How can the venture act as an agent for impact with all of its stakeholders? Prospects, customers, investors, regulators, partners, suppliers, workers, communities, etc.* | **Customer Segments**  *What SDG/ESG metrics are relevant and important to the venture’s customers?*  *How can the venture encourage or support customers to create positive impact?* |
| **Key Resources**  *Which of the following resources does the venture use, and at what quantity?*  *How sustainable are these resources?* | **Channels**  *What is the footprint of the venture’s distribution methods? How can that impact be offset or transformed?* |
| **Cost Structure**  *How does the venture’s cost structure facilitate or restrict support to be provided for SDGs?*  *What measures and key indicators are useful (e.g., ESG, financial) for connecting impact with financial planning?* | | **Revenue Streams**  *For what kind of SDG, ESG outcomes are customers willing to pay?* | |

Adapted from the Business Model Canvas provided by [Strategyzer.com](https://www.strategyzer.com/)

1. There is a 371 page book by Effingham, a 23 page paper by Gruber, a 17 page paper by Guarino, a 10 page rebuttal of Guarino’s paper by Neuhaus, a 17 page paper by Neuhaus), the developing ontologies 101 web site from the creators of Protégé and much more! [↑](#footnote-ref-1)
2. The Venture Impact Canvas is a variation of [Strategyzer.com](https://www.strategyzer.com/)’s Business Model Canvas. [↑](#footnote-ref-2)
3. Produced by “Freiberg, David and Park, DG and Serafeim, George and Zochowski, Rob. 2020. Corporate Environmental Impact: Measurement, Data and Information. Harvard Business School, Impact-Weighted Accounts Project report.” [↑](#footnote-ref-3)
4. <https://www.cia.gov/the-world-factbook/> [↑](#footnote-ref-4)
5. <https://github.com/ontoinsights/deep_narrative_analysis/blob/master/notebooks/Hackathon-Data_Load.ipynb> [↑](#footnote-ref-5)