Use of Semantic Technologies to Inform Progress Toward Zero-Carbon Economy*
Extended Abstract

Stefano Germano[1] [0000-0001-6993-0618], Carla Saunders[2] [0000-0001-6699-2848], Ian Horrocks[1] [0000-0002-2685-7462], and Rick Lupton[2] [0000-0001-8622-3085]

[1] Department of Computer Science, University of Oxford
{stefano.germano,ian.horrocks}@cs.ox.ac.uk
[2] Department of Mechanical Engineering, University of Bath
{cs2537,R.C.Lupton}@bath.ac.uk

Abstract To investigate the effect of possible changes to decarbonise the economy, a detailed picture of the current production system is needed. Material/energy flow analysis (MEFA) allows for building such a model. There are, however, prohibitive barriers to the integration and use of the diverse datasets necessary for a system-wide yet technically-detailed MEFA study. Here we present a demonstration of the "Physical Resources Observatory" (PRObs) system, which exploits Semantic Web technologies to integrate and reason on top of this diverse production system data. We demonstrate how this system supports domain experts to load datasets, map them to the ontology, and query the extended data to meet several modelling use cases.

Keywords: Semantic Technology · Resource efficiency · Rule-based approach · Data integration · Material Flow Analysis · Ontology · Decision Support System

This is an extended abstract related to our full paper ([1]) accepted in the In-Use Track of ISWC 2021. It complements our paper with a more specific description of some details of our prototype and the material we are going to show in our demo. The type of this submission is *demo*.

1 Introduction

A whole-systems understanding of production systems is essential to navigating the necessary rapid transition to a zero-carbon economy. Understanding how

* This work was supported by the EPSRC project UK FIRES (EP/S019111/1), the SIRIUS Centre for Scalable Data Access (Research Council of Norway), and Samsung Research UK. We would like to acknowledge the support of José Azevedo and Christopher Cleaver, whose work on the UK production system case study has provided essential context for the development of this work, and the Oxford Semantic Technologies team for their support.

we produce and consume physical resources is fundamental to understanding the impacts human activity has, and opportunities to increase efficiency. This understanding relies on access to data about the production and consumption of physical resources (materials, energy, etc.) and their associated environmental impacts. Due to the economy-wide yet detailed nature of these questions, they must be based on many pieces of data from diverse sources. Moreover, this data is incomplete, and defined using inconsistent categorisations of the types of resource and activities. In addition, the lack of well-defined data models for this type of data is limiting to data reuse and holding back academic research [7,3].

We propose and develop a solution using a domain ontology and the RDFox triple store to efficiently implement Datalog rules integrating diverse data points into a consistent structure. This forms part of the "Physical Resources Observatory" (PRObs) system, being developed within the *UK FIRES* research programme[3], where it supports a wider research agenda on resource efficiency and decarbonisation in UK industrial strategy.

2 Modelling and reasoning in PRObs

To allow quantified data points on resource use to be expressed in RDF, we build on the data model proposed by Pauliuk et al. [6]. More details about our conceptual data model can be found in [1]. Different data points have been defined with respect to different classification systems, and must be easily and transparently retrieved for reuse in new analyses. More broadly, new information needs to be inferred from the raw data using rules. Generally, this might involve converting data between different definitions of time, location, activity, and object type. We use the Datalog language with stratified negation and aggregates to perform these computations. This allows the complex behaviours required to be expressed in simple rules, while benefiting from the efficient solvers available for evaluating Datalog programs.

The ontology and the RDFox scripts implementing the rules and algorithm are available online at [2].

3 Implementation in PRObs

The PRObs system is intended to be used to support modelling by domain experts who are not familiar with semantic web technologies. As such, they should be supported to enter information (e.g. about materials of interest and their hierarchical structure) and retrieve results without becoming experts in RDF and complex SPARQL queries. The system should be usable (as far as possible) on typical researchers' computers, and integrate with typical modelling workflows involving e.g. Python scripts and notebooks. Further, because defining the system is subjective, the implementation should support users in clearly documenting this input information.

[3] `https://ukfires.org`

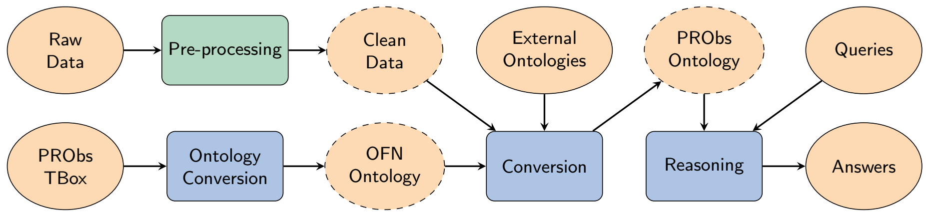

Figure 1. Backend pipeline. The rectangles represent the steps of our pipeline (green for Python scripts and blue for RDFox scripts). The ellipses represent inputs and outputs (dashed for internal results).

Specifically, our system is composed of a frontend interface for defining and documenting the system definitions as input RDF data, and the backend implementation based on RDFox [5]. Our backend pipeline is shown in Figure 1. More details on the full pipeline of the PRObs system are available in [1].

To make the use of the PRObs system accessible to domain experts we adopt a literate programming approach to produce code (RDF) and documentation (HTML) from a single source.

This allows for full documentation-writing features within Python executable notebooks. They are easy-to-use, interactive, and widely used by scientists in various fields; therefore, they are the ideal platform for a system that aims to support the needs of domain experts. To streamline use in a MEFA analysis, we have developed Python wrappers that manage the process of setting up the RDFox scripting language commands to load the relevant datasets, and running RDFox as part of a wider workflow to answer queries accessing the relevant information which form the input to subsequent modelling and analysis steps. The PRObs system runs RDFox scripts to load the input data and answer queries, supported by Python utilities to embed this within a testing or analysis workflow. The input data consists of system definitions in RDF, with external datasets provided in the form of tabular data files and mapping scripts which are read during processing by RDFox.

4 Case study and demo

To illustrate the use of the system, we describe a case study of mapping flows through the UK production system. This case study will form the main part of our demo. In particular, we will demonstrate several use cases within this case study displaying particular intricacies of the PRObs system. A working example of the PRObs system containing this case study and the material that will be used for our demo is available online at [4].

4.1 UK production system

This case study forms part of the ongoing research within the *UK FIRES* programme, motivated by seeking opportunities for innovation in manufacturing processes. *The goal is to obtain a detailed understanding of how supply chains are dependent on different manufacturing processes, and where scrap is currently arising within the system, in order to quantify the benefits of innovation in different areas.* To this end, a MEFA model is used to define the structure of supply chains and estimate the pattern of flows through the system which best matches the available measurements. The role of the PRObs ontology described here is to provide access to data from a diverse set of external datasets in a coherent structure aligned to the required inputs of the optimisation model.

Since there is no standard system definition of UK manufacturing supply chains at the level of technical detail required for this analysis, a major element of the project is to describe a suitable set of *Processes* and *Objects* to which the available data can be mapped, and which describe entities of relevance to the study's research questions. These are defined and documented using the system described in [1]. At the moment, our whole project includes 701 processes and 617 object types, but in our demo we are going to show only a subset of them. The datasets used include Prodcom[4], Comtrade[5], and BGS[6].

Example use cases will be demonstrated through the Jupyter Book[7], with the opportunity to demonstrate live queries. During our demo we are going to show the full pipeline of the PRObs system, from loading the source data to the reasoning and querying phases. The demo will encompass sample datasets loaded and mapped into the ontology, as well as use cases which include: retrieving data points with their system context, retrieving inferred observations at aggregated object level, and identifying provenance of observations. Concrete details about the challenges of our scenario and the lessons learned during this work will be provided as well.

Specifically, we will introduce the data sources mentioned above, and we will describe loading and retrieval details for some objects in the BGS dataset using the PRObs system. Then, we will show why retrieving the original data points in not enough in this scenario and how we can derive new information about the observations thanks to the modelling and reasoning capabilities of our system. We will focus on (i) the aggregation of observations, useful when we want to derive all possible measurements of an object that can be classified into further components; (ii) the propagation of observations, useful when we want to analyse equivalent objects; and (iii) the provenance of observations, useful for explainability and quality of information. Finally, we will present more complex real-world examples and explain why this inferred knowledge is useful for domain experts and for further analyses.

[4] Statistics on the production of goods and materials within the EU (Eurostat). `https://ec.europa.eu/eurostat/web/prodcom`

[5] Statistics on international trade (United Nations). `https://comtrade.un.org`

[6] *British Geological Survey Minerals Yearbook*: data tables for production of various minerals. `https://www2.bgs.ac.uk/mineralsuk/download/ukmy/UKMY2015.pdf`

[7] `https://jupyterbook.org`

References

1. Germano, S., Saunders, C., Horrocks, I., Lupton, R.: Use of semantic technologies to inform progress toward zero-carbon economy. In: International Semantic Web Conference (ISWC) (2021), [forthcoming]
2. Germano, S., Saunders, C., Lupton, R.: ukfires/probs-ontology: probs-ontology v1.5.2 (Jul 2021). https://doi.org/10.5281/zenodo.5052739
3. Hertwich, E., Heeren, N., Kuczenski, B., Majeau-Bettez, G., Myers, R.J., Pauliuk, S., Stadler, K., Lifset, R.: Nullius in Verba: Advancing Data Transparency in Industrial Ecology. Journal of Industrial Ecology **22**(1) (2018). https://doi.org/10.1111/jiec.12738
4. Lupton, R., Germano, S., Saunders, C.: ukfires/probs-ISWC2021-example: Initial release for ISWC2021 paper (Apr 2021). https://doi.org/10.5281/zenodo.5052758
5. Nenov, Y., Piro, R., Motik, B., Horrocks, I., Wu, Z., Banerjee, J.: RDFox: A highly-scalable RDF store. In: The Semantic Web - ISWC 2015 - 14th International Semantic Web Conference, Bethlehem, PA, USA, October 11-15, 2015, Proceedings, Part II. Lecture Notes in Computer Science, vol. 9367. Springer (2015). https://doi.org/10.1007/978-3-319-25010-6_1
6. Pauliuk, S., Heeren, N., Hasan, M.M., Müller, D.B.: A general data model for socioeconomic metabolism and its implementation in an industrial ecology data commons prototype. Journal of Industrial Ecology **23**(5) (2019). https://doi.org/10.1111/jiec.12890
7. Pauliuk, S., Majeau-Bettez, G., Mutel, C.L., Steubing, B., Stadler, K.: Lifting Industrial Ecology Modeling to a New Level of Quality and Transparency: A Call for More Transparent Publications and a Collaborative Open Source Software Framework. Journal of Industrial Ecology **19**(6), 937–949 (Dec 2015). https://doi.org/10.1111/jiec.12316