

Semantic Technologies and Semantic Web: Structuring Data for STI Policy Analysis

Agenda of a joint EC-OECD workshop in support of the H2020 [REITER project](#)

19 June 2017, 8 Square Frère Orban, Brussels

- 10:00 Welcome – Román Arjona (EC) and Dominique Guellec (OECD)
- 10:15 Overview of the REITER project – Michael Keenan (OECD) and Ana Nieto (EC)
- 10:45 Session 1 -- **Reshaping policy monitoring and analysis within the STI policy field**
 Chair: Marnix Surgeon (EC)
 Presenters: Philippe Larédo (ENPC, Paris), Kincsö Iszak (Technopolis, Brussels), Cameron Neylon (Curtin University, Western Australia & FORCE11)
 Discussants: Andrea Bonaccorsi (Univ. Pisa), Edward Ziarko (BELSPO, Brussels)

A key challenge in the science, technology and innovation (STI) policy domain is to cope with ever increasing amounts of textual data describing strategies, policy initiatives and evaluation practices. Large amounts of textual data (e.g. official documents describing policy initiatives, reports, evaluations and websites) that provide policy-relevant information are published on the web in different formats. While readily available for consultation, analysts struggle to navigate, compare and synthesise such information. This session will explore how different projects are facing this challenge by structuring data within the STI policy field, and heralding a paradigmatic shift in the way policy-relevant data is collected, curated and analysed. For example, in the context of the REITER project, the EC and OECD are in the process of shifting their policy monitoring and analysis arrangements into a taxonomy management tool that harmonises descriptions of STI policies, generates surveys for collecting country data, and drives a platform for interacting with and visualising STI policy data that supports policy analysis.

- 12:30 Lunch
- 13:30 Session 2 – **Using semantic technologies and semantic web to structure the STI policy field**
 Chair: Alina Deniau (OECD)
 Presenters: Thierry Vebr (OECD), Diana Maynard (Univ. Sheffield), Cinzia Daraio (Univ. Roma Sapienza)
 Discussants: Frédérique Sachwald (OST, France), Abdullah Gök (Univ. Manchester)

This session discusses how semantic technologies, e.g. Natural Language Processing (NLP), the semantic web approach, e.g. linked data, and sophisticated web scraping can be mobilised for STI policy analysis. Ontologies and taxonomies built by human operators can be used to structure and enrich data on policy initiatives. Another approach is based on NLP algorithms that extract the main themes in the STI policy field from a large corpus of documents. The session will provide an opportunity to demonstrate and discuss experiences and to identify opportunities and lessons arising from these methods. It will highlight key issues in building interoperable data models: identifiers, linked data, and modelling of human knowledge for machine use.

- 15:00 Coffee break
- 15:30 Session 3 – **Supporting STI policy data analysis through embedded analytical and visualisation tools**
 Chair: Andrés Barreneche (OECD)
 Presenters: Juan Mateos-Garcia (Nesta, UK), Frédéric Olland (Ministère de la Recherche, France), Francesco Osborne (Open University, UK)
 Discussant: Pascale Dengis (FRIS, Belgium)

It has so far been difficult to develop data visualisation interfaces to support analysis of unstructured data. But the knowledge structures discussed in the previous sessions now allow for more advanced interfaces, which can help analysts to find patterns that could inform STI policy processes. This session will discuss current practices and examples of visualisations, dashboards and semantic filtering/linking tools that allow analysts to readily query and navigate policy data. For example, in the context of the REITER project, the EC and OECD are developing a platform that presents data in interactive dashboards that are built on analysts' queries. In this way, the platform's visualisation capabilities support policy analysis, providing an overview of major points of interest and assisting policy analysts in report preparation.

- 16:45 Wrap-up and follow-up steps – Michael Keenan (OECD) and Ana Nieto (EC)
- 17:00 Workshop close