Szymon Klarman *13.03.2019*

A proof-of-concept of a UN SDGs linked data infrastructure

In this note and an attached diagram, we propose a proof-of-concept for a **linked data infrastructure** (including data items and architectural components) for *describing*, *relating*, and *exposing* information about diverse resources relevant to UN SDGs: *datasets*, *documents*, *reference taxonomies* and *identifiers*. Building on the principles of linked data, we suggest employing a selection of broadly supported, machine-accessible vocabularies, identifiers, and service types for representing and sharing the meaning of the information resources over the Web.

The key **conceptual components** of the infrastructure are fine-grained *subject taxonomies* and *spatio-temporal identifier systems*, whose aim is to enable precise semantic description of datasets and documents and their adequate linking to specific sustainable development goals, targets and indicators.

On the **architectural level**, the central facility is an *RDF triple store*, which (at least during the development phase) should support complex traversals and pattern matching of linked data. Different fragments of that data are meant to be further exposed via diverse *services*, suited towards different consumers and use-cases:

- *Json-LD-enabled web APIs* for applications (REST-, GraphQL-based);
- *resolvable URIs* to enable dereferencing of owned identifiers by external consumers of the SDG data;
- *websites with embedded structured data* (e.g., Json-LD) for advertising access points to the published datasets for human consumers as well as third-party search engines such as Google Dataset Search.

As a roadmap towards a minimum viable proof-of-concept, we propose a two-stage delivery process comprising (*cf*. the diagram below):

(1) :

- providing semantic metadata for a sample of statistical datasets,
- linking these to SDG ontology via subject taxonomies,
- setting up RDF triple store for hosting and querying linked data resources,
- exposing web APIs serving fragments of collected linked data as Json-LD.

(2) :

- providing semantic metadata for a sample of documents,
- linking these to SDG ontology via subject taxonomies,
- representing a sample statistical dataset in RDF Data Cube Vocabulary,

- linking that dataset to SDG ontology via subject taxonomies and spatio-temporal identifiers,
- publishing static websites with embedded structured data advertising a sample of statistical datasets,
- exposing a simple linked data server enabling URI dereferencing for a subset of resources.