Geophysical Research Abstracts
Vol. 21, EGU2019-19123, 2019
EGU General Assembly 2019

Bringing Semantics to Citizen Data Collection - A Semantic Extension of Open Data Kit 1

Markus Daniel Steinberg (1), Sirko Schindler (2), and Friederike Klan (2)
(1) Friedrich-Schiller-University Jena, (2) German Aerospace Center

Citizen Science contributions gain more and more momentum in many areas of scientific research. Applications range from the classifying galaxies from telescope-images and the transcriptions of scanned texts to data collection tasks like taking photos of celestial phenomena or taking stock of the local flora and fauna. While oftentimes domain-specific tools are developed, for data collection tasks past years have seen the evolution of frameworks that allow researches without in-depth technical knowledge to create their own surveys. These frameworks also support a wide range of devices allowing citizens to use, e.g., their mobile phones to collect and submit data. While these frameworks support the creation and execution of data collection surveys, the data export functionalities are oftentimes restricted to standard tabular formats like CSV or Excel. On the other hand, we see an increased usage of the Linked Data Cloud using formats like RDF and a multitude of vocabularies that describe observations and measurements taken by citizens. Here, semantic data models associate datasets with machine-readable meaning allowing automated processing. In order to bridge this gap, we extended one popular data collection framework, Open Data Kit 1 (ODK1), with capabilities to augment collected data with semantic concepts. In the design phase, researchers define the fields used within the survey that later users are asked to fill. We extended this process with the option to provide semantic descriptions for different aspects. Our initial prototype allows adding a basic set of additional information: a concept defining the meaning for a field and a unit of measure to qualify the values collected. Here, concepts are chosen from an ontology pulled from an arbitrary SPARQL-endpoint and offered to the researcher using an autocomplete-feature inside the form designer. This lowers the burden for the researcher designing a survey, but yet provides the flexibility to adapt the range of options to the specific needs of the domain and ensure a certain degree of consistency throughout all surveys on a given installation. Furthermore, we added a new export pipeline based on sets of templates. A template set allows to define the structure of the data export in compliance with a certain data model. In our case, this structure is based on ontologies used to describe measurements. In particular, we used the OBOE Extensible Observation Ontology as an example to embed the results in. To validate the flexibility of our approach we also asked data managers with a background in semantics to define template sets for other ontologies like W3C's Data Cube vocabulary and even more simple formats like CSV or XML. In summary, we will present our additions to ODK1 that enable researches to easily define semantically enriched surveys for data collection campaigns and export the results in formats compatible with their local infrastructure and the Linked Data Cloud. We hope that this will foster the acceptance and use of FAIR data principles in the Citizen Science community.