Open Data Kit Goes Semantic –
A Contribution to the Interpretability and
Interoperability of Citizen Science Data

Markus D. Steinberg[1], Sirko Schindler[2], Friederike Klan[3]

[1]*Friedrich-Schiller-Universität, Jena*
[2]*Data Management Technologies Group, DLR Institute of Data Science, Jena*
[3]*Citizen Science Group, DLR Institute of Data Science, Jena*

Mobile data collection is playing an increasingly important role in Citizen Science projects.

In Order to Be (Re-)Usable and analyzable, Citizen Science Data Need to Be …
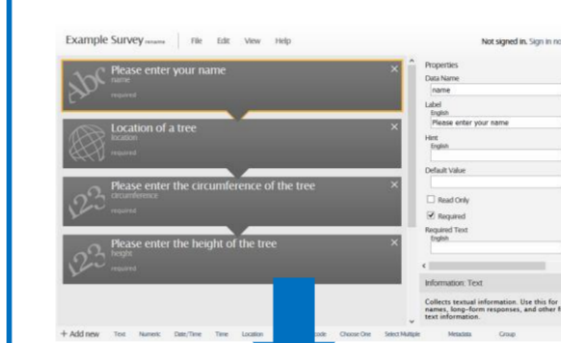
… (1) easily and automatically processable on a computer, (2) need to be interpretable by humans and (3) interoperable with other data.
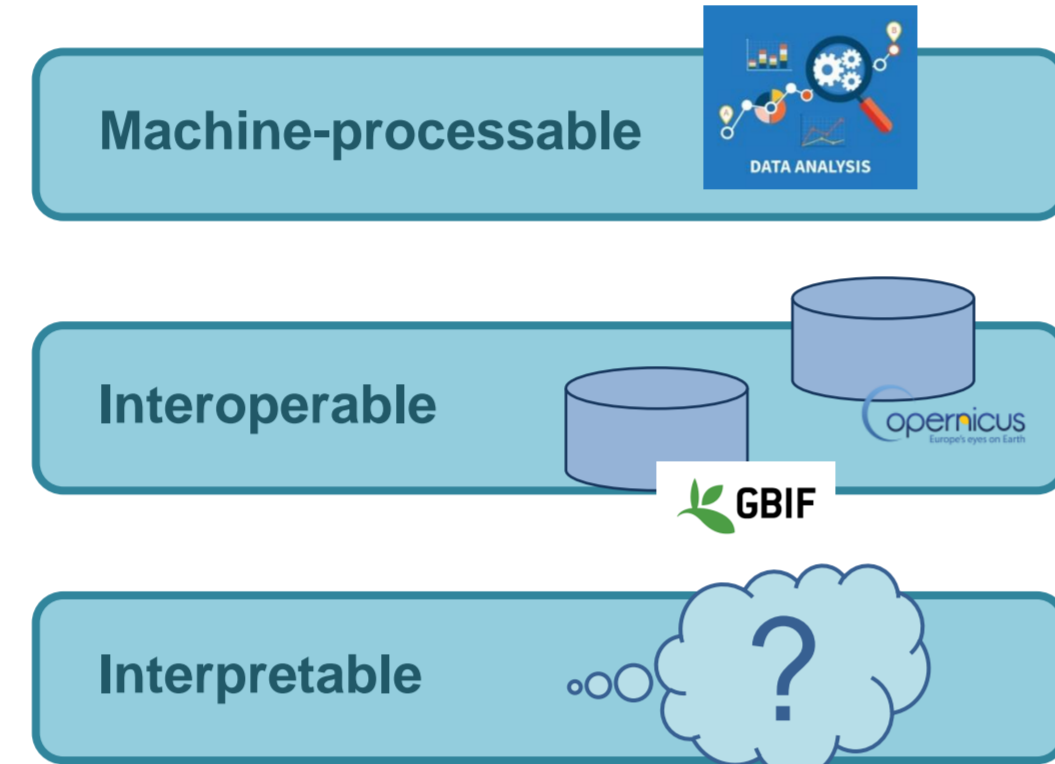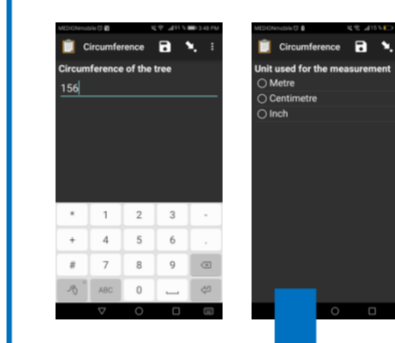

Mobile Data Collection Frameworks

… such as the Open Data Kit (ODK) facilitate the collection of Citizen Science data by enabling drag-and-drop creation of online surveys and data collection via smartphone applications.
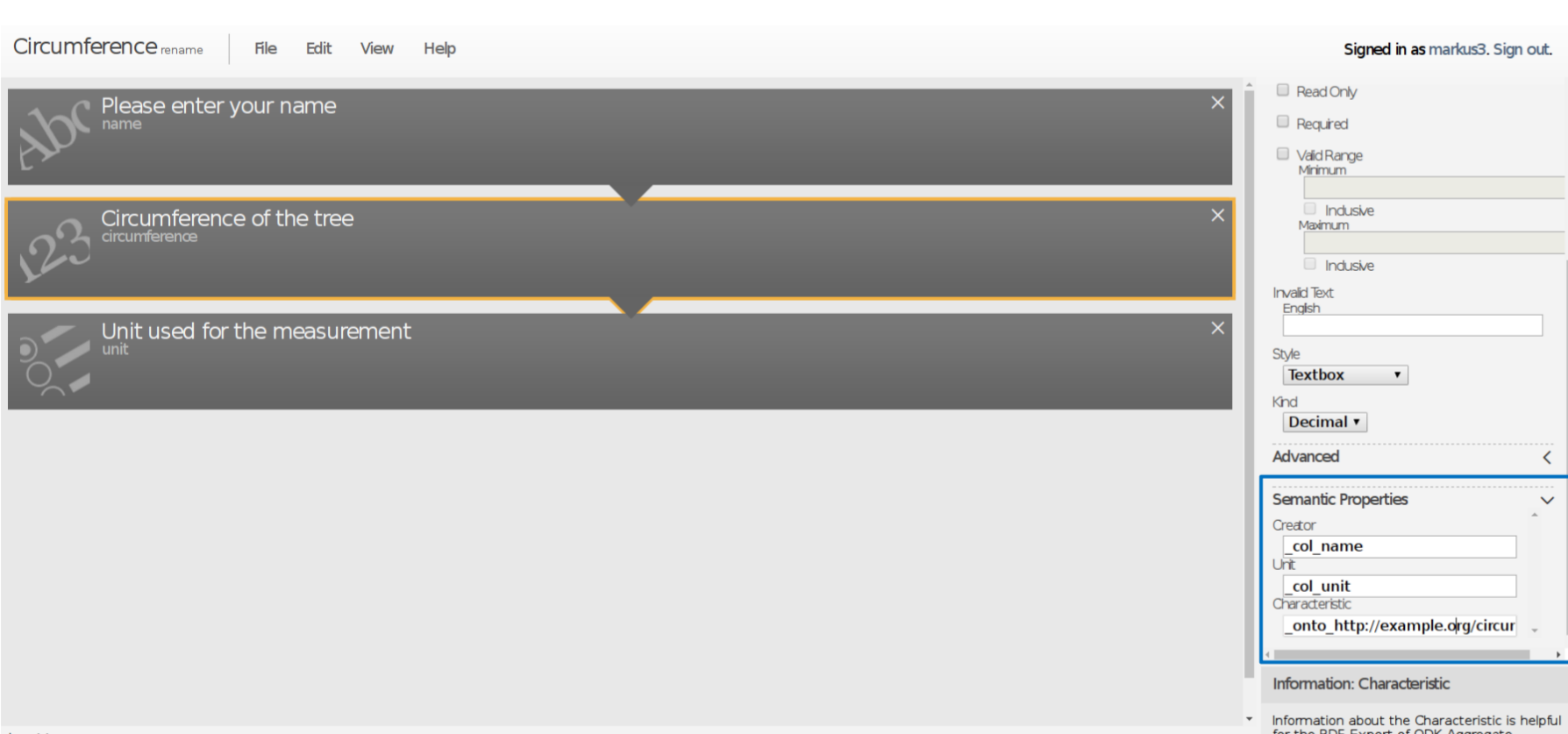
SURVEY DESIGN

MOBILE DATA COLLECTION

Limitations
- data export limited to standard data formats and models (e.g. CSV or Excel)
- metadata which describe the semantics of the data and the circumstances under which they have been collected (what was measured / observed?) typically not provided

DATA EXPORT
JSON, KML, CSV

Limited (re-)use of Citizen Science data, as interpretability and integration with other data is compromised

We've developed a methodology and implemented it as a software extension

to the mobile data collection framework Open Data Kit 1 (ODK1)[1]. Our solution contributes to the interoperability and interpretability of Citizen Science data by …

(1) Enabling Interpretability by Enriching CS Data with Metadata at Design Time

Designing a mobile survey for the collection of tree parameters (e.g. tree circumference) using ODK's form designer ODK Build using the highlighted fields metadata can be linked with form elements.

Metadata fields can be linked with form elements, e.g. the metadata information unit of measure for the input element tree circumference is linked with the unit selected via the single choice field unit.
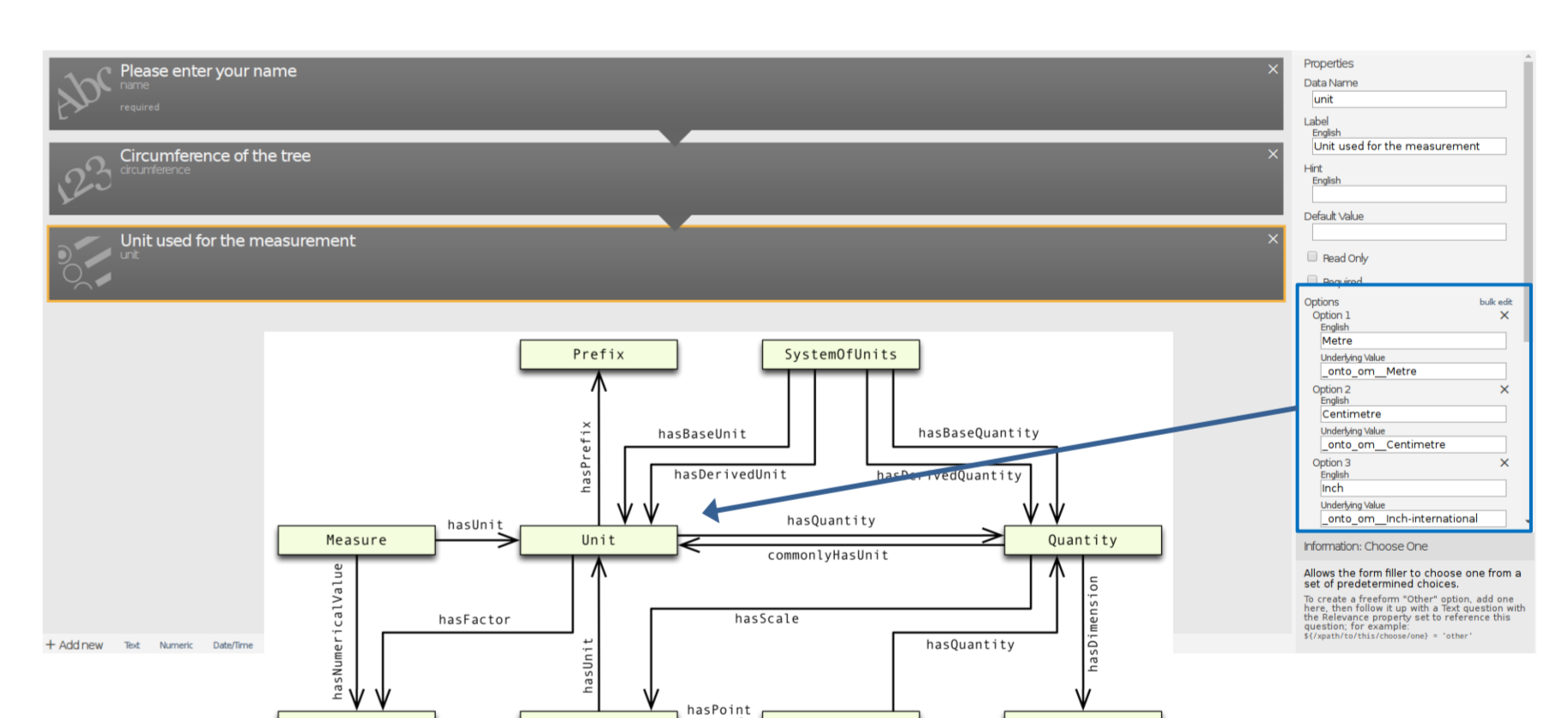
Metadata fields (here the field unit) can be linked with formal knowledge (here the Ontology of units of Measure (OM)[2] providing information about units, measurable quantities, and the conversion of units).

(2) Enabling Interoperability by Offering Flexible Template-Based Data Export …

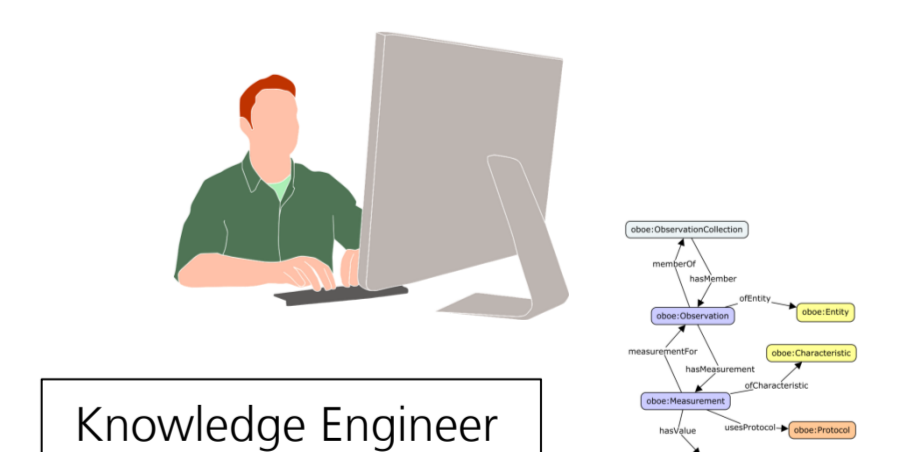… to customized data formats and models including XML-, CSV- and RDF-based (Linked Open Data) formats.
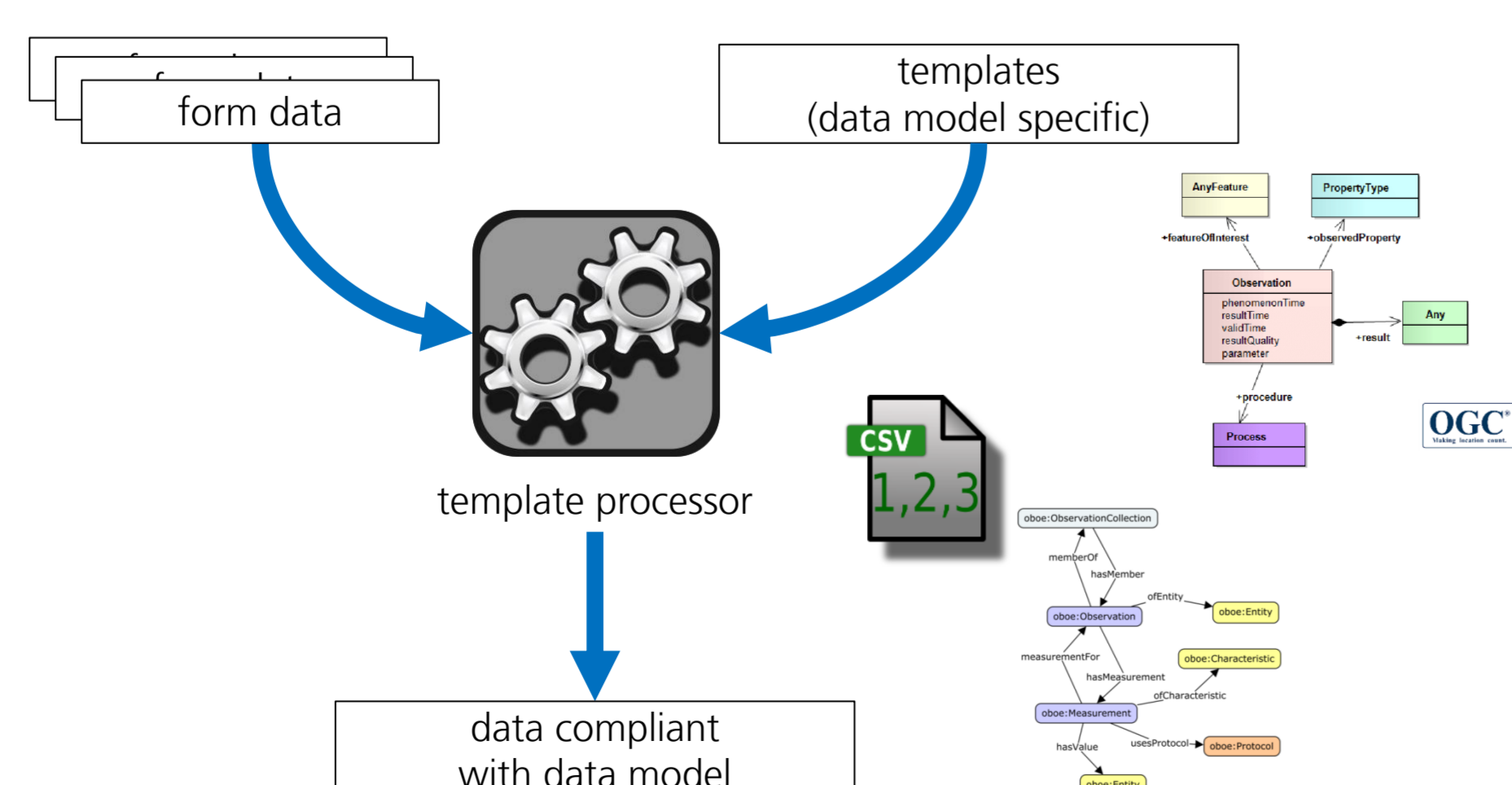
Separation of Concerns & Transparency …

… ensures that data managers are only bothered with form design and the collected data,

Data Manager

knowledge engineers are only concerned with the data model and format, and

Knowledge Engineer

citizens data collectors just interact with the mobile application.

Data Collector

Deutsches Zentrum
DLR für Luft- und Raumfahrt

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

Contact friederike.klan@dlr.de, sirko.schindler@dlr.de
Software https://gitlab.com/dlr-dw/semantic-data-collection

[1]https://github.com/opendatakit/, [2]H. Rijgersberg, M.F.J. van Assem, J.L. Top, "Ontology of Units of Measure and Related Concepts." *Semantic Web*, **4**, 1, 2013, pp. 3-13.