



Conference Proceedings | DOI: <https://doi.org/10.18174/FAIRdata2018.16272>

Towards a semantic approach for environmental timeseries data re-usability

Samourkasidis, A.¹, I.N. Athanasiadis¹

¹ Wageningen University & Research, Information Technology Group, Hollandseweg 1, Wageningen 6706 KN, the Netherlands

Corresponding author's e-mail: argyrios.samourkasidis@wur.nl

The data deluge following the rise of Internet of Things contributes towards the creation of non-reusable data silos. Especially in the environmental data science domain, syntactic and semantic heterogeneity hinders data re-usability as most of the times manual labour and domain expertise is required. To this end contributes, not only the different syntaxes under which environmental time series are formatted, but also the implicit semantics which are used to describe them. Usually, the real meaning of data is obscured in a combination of short data labels, titles and various value codes, that require domain or institutional knowledge to decipher. In this work, we present EDAM (Environmental Data Acquisition Module), a template framework which deals with syntactic and semantic heterogeneity. Data curators draft templates to describe in an abstract fashion, the syntax of the time series datasets they want to acquire and/or disseminate. They complement each template with a metadata file, which is used to annotate observables and their properties (i.e. units of measurement, etc.) with terms from an ontology. EDAM employs a reasoner to infer compatibility among syntactically and semantically heterogeneous datasets, and enables on-the-fly time series, format and units of measurement transformation. Our approach utilises a local ontology to store metadata about datasets, which enables EDAM to acquire and transform datasets which are originally stored under different semantics and syntaxes. We demonstrate EDAM in a case study where we transform meteorological input files of four agricultural models.