Semantic Specification of Data Type Information in the Deep Carbon Observatory Data Portal

Xiaogang (Marshall) Ma (max7@rpi.edu), John Erickson, Patrick West, Stephan Zednik, Peter Fox
Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY, USA

Background
• Data types are often treated only as syntax of variables, such as integer, float, boolean, character, and string, etc. Such declaration does not offer any domain specific meaning to the data types.
• Our intention is to let a data type include more meanings, such as who create the data type, the source standard that the data type derives from, the operations that can be done on datasets of that data type, and typical scientific domains, software programs and/or instruments that use the data type.
• Initial results have already been achieved in the Deep Carbon Observatory (DCO) Data Portal (http://info.deepcarbon.net).

Our Aim
Any humans or machines facing a data type can quickly understand or be in a situation to at least process details within the dataset without even downloading it.

Nature of Efforts
• A registered DCO dataset is asserted as an instance of a BASIC DATA TYPE, such as Dataset, Image, Video, and Audio, etc.
• It is possible to further annotate a registered dataset with the SPECIFIC DATA TYPES defined by the DCO community members.

Initial Results
Updates to the DCO Ontology:
• A new class dco:DataType. Each specific data type is an instance of it.
• An object property dco:hasDataType linking a dataset and a data type
• A collection of other classes and properties associated with dco:DataType

An Example in The DCO Data Portal

 WHY Should You Care?
• Data types make aspects of data more visible
• Data types group data sets with similar characteristics
• Data types will help you find data sets matching your needs
• Data types enable machines to find tools and algorithms for specific datasets
• More features in an ‘inter-linked world’...

Future Works
• More use case analyses relevant to data types in the DCO community
• Refine the schema for the annotation and provenance of specific data types
• Interoperability between DCO specific data types and data types registered in other communities
• A separate ontology for data type?

(Image credit: Ainsley Seago, PLoS Biology)

(Images credit: deepcarbon.net and X. Ma)