**Information entities**

Any ontology of biomedical investigations needs to represent pieces of information such as collected data, investigation designs, and reports of investigation results. When we began our work, the BFO, being a realism based ontology, did not have a way of representing of such entities.

As developing a general theory of information entities is outside the scope of OBI, and will be of use to many projects, the Information Artifact Ontology(IAO) [ref: http://purl.obofoundry.org/obo/iao] was spawned as separate effort.

The IAO is intended as a middle level ontology, being based on BFO, and imported and used by OBI and other projects. It is structured by two being explicit about two types of existential dependency – that information entities are *generically dependent* (i.e. they can be borne by several entities and exist as long as there is at least one bearer for them), and that they are *about*other entities.

The relation of *aboutness* captures the fact that information doesn’t appear out of nowhere - that it has as its genesis the existence of something else.

The data item that a digital scale produces has as its genesis a substance with mass. A common problem in our field is that it is difficult to share data. We diagnose this problem to have root cause that representations of data omit clearly representing what the data is about. By maintaining this relation as in integral part of our representation of information, we hope to improve on the ability to share data.

The root class of IAO is called *information content entity,* a term intended to encompass those kinds of information entities that are intentionally created. For OBI, an important subclass of these is *directive information entity (e.g., plan specification* and *objective specification*), which represent things that pertain to processes such as instructions and protocol specifications that when *concretized* in an individual are realized as *processes.*

IAO defines general classes such as *document*, for example a journal article or a patent application, parts of documents - *textual entities* such as paragraphs of displayed text, as well as terms that more specifically apply to documents with special purposes. Particularly relevant to OBI are terms such as *report* and parts thereof, *report elements* such as *report* *table*, *figure* and *scatter plot*.

OBI then defines subclasses of *information content entity* that are specific to investigations. These include *protocol* and *study design*, which are types of *plan specifications,* representing specification of single procedure or an entire investigation, respectively. An important part of many *study designs* is the identification of *dependent* and *independent variables specifications*. Central to investigations are the types of data are generated. This includes *measurement datum*, as are obtained in assays, and a growing set of subclasses that define the structure and relations for more specific data items. For example, *scalar measurement datum* represents measurement data composed of a single numeral and a unit label (e.g. *length measurement datum, about length quality instances).* Finally, the information generated as the outcome of an investigation is represented as instances of *study interpretation*, for example as found in journal articles in a conclusion section.