Data Validation 10 Years On: The Use of Interactive Ontologies, Sydney R Hall[a] and Nick Spadaccini[b], [a]*School of Biomedical & Chemical Sciences;* [b] *School of Computer Science & Software Engineering, University of Western Australia, Nedlands 6009, Australia.* E-mail: syd@crystal.uwa.edu.au

Keywords: Data-validation; Data-ontology; Methodology

Data dictionaries are widely used by databases and journals to support the validation of submitted data. These dictionaries contain precise computer-readable information about data items used in particular disciplines. In crystallography, CIF data dictionaries exist for data used for structure analysis, macromolecular structures, powder diffraction, symmetry, incommensurate structures, and precision density studies. These are described in detail in the soon-to-be-published *International Tables Volume G* [1]. The primary function of these dictionaries is the precise identification and characterisation of frequently-used data items. This characterisation, which includes the definition of attributes that specify the dependencies between data items; whether they are numbers or text; and their allowed enumeration, underpins many data validation processes used currently by journals and databases when accepting deposited data.

Data dictionaries, or ontologies, as they are more generally referred to, can also provide detailed relational knowledge about data. This is usually in the form of 'methods' that record the functional relationship of *derivative* data items to *primitive* (i.e. measured or postulated) and other derivative data. In the main, methods are algorithms that allow non-primitive data to be evaluated from other data, and may be applied to classes of data as well as to individual items.

The next generation of ontologies will be capable of direct application to a particular data instantiation i.e. they are interactive and executable. Moreover ontologies will provide method scripts for the dynamic *redefinition* of the attributes (e.g. the enumeration of an item can be changed according to the value of another), *conformance* (important in DDL dictionaries where the instantiated data are data ontologies) and *validation* (i.e. methods for both consistency and quality checks).

This paper will describe an interactive ontology approach based on the *StarDDL* and *dREL* languages [2], and demonstrate how these are applied to particular data instantiations.

[1] *International Tables for Crystallography Volume G* (2004) Eds: S R Hall and B McMahon. Kluwer Academic Press: London.

[2] Spadaccini, N., Hall, S.R. and Castleden, I.R. (2000) *J. Chem. Inform. & Computer Sci.* **40**, 1289-1301.