

THREE DIMENSIONS OF INTEROPERABILITY

Timothy B. Patrick

College of Health Sciences, University of Wisconsin-Milwaukee

A goal of this workshop is to consider approaches to developing an information environment that makes it possible to fully and efficiently utilize a number of diverse biological databases and data analysis tools on the Internet. In this talk, I report on the programmatic efforts towards this end by the Laboratory for Integrative Informatics, previously at the University of Missouri-Columbia, and since 2005 at the University of Wisconsin-Milwaukee. We have focused our efforts on three dimensions of the problem of interoperability of biological information resources. These three dimensions are *intensional data-type semantics*, *scientific/editorial culture*, and *principles of representational choice*.

Intensional Data-type Semantics In order to achieve the kind of information environment that is the goal of this workshop, it will be necessary to develop ways to measure the quality of information processing workflows. A typical representation of an information processing workflow is a directed graph where the nodes are information resources and the arcs are data links between adjacent resources. A data link indicates that data output by one resource is input to another resource. The data outputs and inputs of information resources are characterized by semantic data-types such as *dna sequence length*, *dna sequence descriptor*, *keyword*, *controlled vocabulary term*, and *gene ontology term*. Ostensibly, if two adjacent resources share the same output and input data-type, a corresponding data link between those resources will be legitimate. If, however, the output and input uses of the data-type are different enough, the data link may be illegitimate. For example, a data link between two resources, one of which outputs data of type *gene ontology term*, and the other of which inputs data of type *gene ontology term*, may be illegitimate if the output and input uses of that data-type are restricted to different sensu or species specific applications. We model such differences in the application of a data-type using a possible worlds framework for ontology conceptualization of the sort proposed by (Guarino, 1998). We assume that workflows are defined over a given collection of information resources. We assume that resources in the collection are related according to some point of view or set of purposes and that the collection is characterized overall by a set of semantic data-types. The data-types are inter-related at the output and input interfaces of information resources in the collection, and they may be inter-related differently at the interfaces of different resources. We treat each interface as corresponding to a particular data-type ontology. The problem of determining the quality of a workflow is reduced in part to the problem of determining the ontological legitimacy of its data links. The problem of determining the ontological legitimacy of a data link is the problem of determining whether there is sufficient ontological similarity between the data link's output and input data-types.

Scientific/Editorial Culture Biological information resources may be composed into workflows of varying quality, leading to results of varying quality. Thus, achieving the kind of information environment that is the goal of this workshop requires a culture of documentation of the information processing methods used to generate scientific results. To this end we have investigated aspects of scientific/editorial culture with regard to information processing methodology and how it is reported in biological science papers. Our preliminary results suggest that information processing methods are reported in much less detail than wet bench methods.

Principles of Representational Choice Web services and the service oriented architecture are particularly well suited to the biomedical domain since biomedical research is carried out by independent labs and investigators who nevertheless have a need to share data, tools and other resources. Web services provide a common framework for facilitating the sharing of resources. Presumably, a scientist who makes a resource available as a web service will have, in at least some cases, an interest in interoperating that resource with the resources made available by other researchers. Yet the particular ways the scientist chooses to represent the informational content of his or her resource (perhaps with a particular controlled terminology or ontology) may impede such interoperability and this in part due to the representational choices of other resource providers. Accordingly, we have undertaken an investigation of the principles of rational choice of representational scheme, by the provider of a resource, in a context where (1) interoperability of resources is highly valued and (2) that interoperability may be impeded by the possibly conflicting and simultaneous representational choices of other providers of otherwise potentially interoperable resources.

Guarino, N. (1998). *Formal Ontology and Information Systems*. In N. Guarino (ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. Amsterdam, IOS Press, pp. 3-15.