The role of registries within a spatial data infrastructure

S. J. D. Cox (1), Rob Atkinson (2)

(1) CSIRO Exploration and Mining, Bentley, WA, Australia, (2) CSIRO Land and Water, Lucas Heights, NSW, Australia (Simon.Cox@csiro.au / Phone: +61-8-6436-8639)

Spatial data infrastructures, including systems to support environmental science applications, are expected to be implemented using a service-oriented architecture. However, the only significant element of SDI's that has seen wide scale deployment are directories of datasets, primarily based on metadata records using one of the well-known standards (FGDC, ANZLIC, ISO 19115), and primarily accessed through human user interfaces. Examples of these have been available for close to a decade, but there has been a frustrating lack of progress in moving from the latter (dataset directories) to the former (spatial data infrastructures). This has led to a degree of skepticism in the community concerning the SOA approach and the plausibility of SDI's, despite general acknowledgement of their key role in mitigation of environmental challenges at all scales.

What is the nature of the blockage? It is essentially the traditional data-assimilation problem. Interaction with a *dataset directory* typically ends with the user getting, at best, a URL to download a pre-packaged dataset. Then what they retrieve is in a proprietary format, structured according to a schema defined by the custodian, and labelled using tokens derived directly from the persistence layer. The user is responsible for interpreting the structure, mapping it to the data model required for their processing environment, and performing the necessary conversions. Even when the structure is well understood, the content must be reconciled. In an SDI the service provider faces a one-off burden of describing the data structure, content and service behaviour. This has proved difficult, with few scalable precedents.

What has caused the blockage? Reducing the burden on the data provider, and the data consumer trying to understand the service offering, requires standardization of a lot more metadata elements, at a fine level granularity. Many of the elements in a conventional metadata record are entered separately every time as text (e.g. "keywords", type-names), or structured data (e.g. contact information). However, interoperability requires that all parties subscribe to the same set of definitions for concepts that they share, rather than using different words or structures to describe the same things. Hence, descriptors within metadata records should be pointers to items provided from standard sources, managed on behalf of the interoperating community. This applies to keywords, data types, service types, vocabulary bindings, etc. In particular, automation of service-chain composition requires subscription to standards for all elements of service classification, except for those that are unique to an individual service instance.

How should standard classifiers be provided? Through registries, that host registers for every "classifier" required in the description of every other resource type. The registers are managed on behalf of the interoperability community, and the register items are normatively denoted by identifiers assigned by the registry.

The role of registries is central to an SOA. Interoperable distributed systems require all shared elements to be managed or published in registries, at the same fine level of detail required to create interoperable services. An SDI, at its core, comprises framework data sets and reusable metadata about these that enable additional data and applications to be deployed using this metadata as the basis of interoperability agreements.