Science Focus
Workgroup 2.2 Authenticity, Reliability and Accuracy of e-Science Records

Activity: Survey and analysis of scientific encoding languages for non-textual records.

Description. The preservation of authentic e-science data, information and records requires an understanding of the syntax and semantics of their serial encoding formats and the kinds of scientific digital objects that are so represented. Such an understanding is also needed for a conceptual analysis of notions of accuracy, reliability and authenticity of e-records created in scientific activities, another task of Workgroup 2.2.

Approach. Survey the e-science literature for file formats and encoding languages that are used for non-textual scientific data, information and records. List the file format/encoding language name, filename extension (if any), MIME type (if any), Publication name and source, viewer name and source, and references to usage in e-science. Also attempt to obtain examples of each file type and viewers for those file types.

Analyze the file formats and encoding languages to determine the data, information, and/or record structure and other properties related to the accuracy, reliability and authenticity of the object represented, e.g., metadata. Determine equivalence classes of data formats and encoding languages with respect to the kinds of objects they represent, that is to say, data formats that could represent the same semantic object. Identify conversion tools that might be used for migration. Report the results of the survey and analysis. Report will contain a glossary of technical terms regarding scientific file formats and encoding languages.

Responsibility. Bill Underwood and GRA at Georgia Tech. Advice form Reagan Moore and Fraser Taylor.

Schedule. Interim Report LA Workshop in September, 2003 Final Report Vancouver, February, 2004