Interoperability Briefing Paper

By Daisy Abbott, University of Glasgow

1. Introduction

Interoperability is the transfer and use of information in a uniform and efficient manner across multiple organisations and IT systems. Its purpose is to create a shared understanding of data.

Data exchange requires the data to be semantically matched (i.e. ensuring that the data describe the same thing) and for any differences in representation within the data models to be eliminated or meaningfully handled.[1] Data integration is the process which takes heterogeneous data and their structural information and produces a unified description and mapping information to allow seamless access to all existing data. Interpretation of these data must be unambiguous. More generally, interoperability goes beyond data compatibility as we also need interoperable hardware, software, and communication protocols to allow data to be interpreted correctly and unambiguously across system or organisational boundaries.

Interoperability can be divided into five different conceptual levels:

1. **No Data Exchange**
2. **Unstructured Data Exchange:** exchange of human-interpretable, unstructured data (e.g. free text)
3. **Structured Data Exchange:** exchange of human-interpretable structured data intended for manual and/or automated handling, but requires manual compilation, receipt, and/or message dispatch
4. **Seamless Sharing of Data:** automated data sharing within systems based on a common exchange model
5. **Seamless Sharing of Information:** universal interpretation of information through co-operative data processing[2]

2. Short-term Benefits and Long-term Value

Using multiple, independent data sets risks incompleteness, inconsistency, duplicated effort, and conflicts of management responsibilities. Interoperability can solve many of these issues and may lead to:

- Improvement of the consistency of information and ensuring it can be used across different technological and organisational boundaries — this increases the number of people benefiting from access to or use of the data
- Wider propagation of standards, giving stakeholders a common language
- Acceleration of the adoption of and compliance with relevant data standards
- Quicker transfer of data and more efficient working
- Simplified flow of data results in reduced administrative costs
- Reduction in duplication of effort

- Unexpected data use which could potentially lead to innovative research and analysis

3. Examples of Interoperability in Practice

Some domains have been quicker to embrace interoperability than others, for example geographical information systems (GIS) and healthcare. For example, the use and distribution of various spatial data formats can be achieved through proprietary applications such as ArcGIS[3] and high levels of standardisation and data accreditation, as are often found in medical information systems, makes interoperability easier to achieve.[4]

Significant steps have already been taken towards interoperability in structured data such as database and XML schema.[5] However, the issues become more complex when data are held in discrete, varied formats (e.g. spreadsheets, free text, external websites) rather than structured databases or XML schema and may require sophisticated knowledge discovery techniques. The overall system may also need to evolve with changes to the availability and structure of diverse source data.

4. HE/FE Perspective

"*Digital data are increasingly both the products of research and the starting point for new research and education activities. The ability to re-purpose data — to use it in innovative ways and combinations not envisioned by those who created the data — requires that it be possible to find and understand data of many types and from many sources. Interoperability (the ability of two or more systems or components to exchange information and to use the information that has been exchanged) is fundamental to meeting this requirement.*"

— U.S. Office of Cyberinfrastructure[6]

5. e-Science Perspective

"*Interoperability is key to all aspects of scale that characterize e-Science, such as scale of data, computation, and collaboration…We need interoperable information in order to query across the multiple, diverse data sets, and an interoperable infrastructure to make use of existing services for doing this.*"

— Hendler, J. & De Roure, D. (2004). "E-Science: the Grid and the Semantic Web", *IEEE Intelligent Systems*, 19 (1). p. 65.

6. Issues to be Considered

- Data interoperability is only one aspect of the overall interoperability problem. Data could be requested that is not actually held, the hardware or connectors, operating systems, metadata, protocols, or applications could be incompatible. These are interoperability issues that go beyond data integration.[7]
- Design issues need to address both human-centred aspects (e.g. authority, co-operation, negotiation) and data-centred aspects (data integration, schema evolution). Operational issues need to address system interoperability such as new transaction types, query processing algorithms, and security.
- Other human issues include the fear of losing control over interoperable data, the training and process change required to technically and administratively managed complex systems over a dispersed community.
- Consistency or mapping between terminological choices. This will affect use and re-use of the data. Use of thesauri and relevant ontologies could help improve consistency.

- Interoperability goes beyond technical implementation, and into the area of conceptual modelling. Different data will likely have different structures and specialisation criteria. How can data be integrated into a single semantic representation without over-generalising this structural information?8
- The four types of basic conflict that need to be resolved are:
    1. **Semantic Conflicts** (different schemata do not match conceptually and therefore must be aggregated)
    2. **Descriptive Conflicts** (e.g. terminological choices, naming difference between conceptually identical data, different measurement values that need to be resolved or mapped)
    3. **Heterogeneous Conflicts** (the methodologies being used to describe the concepts differ substantially)
    4. **Structural Conflicts** (concepts are structured differently, e.g. one schema uses an attribute whereas another uses a reference)9
- Data use over a distributed architecture can have disadvantages as well as advantages and which need to be taken into account.

7. Additional Resources

- Hendler, J. & De Roure, D. (2004). "E-Science: the Grid and the Semantic Web" , *IEEE Intelligent Systems*, 19 (1). pp. 65-71.
- Open Geospatial Consortium Interoperability Institute (OGCii)
- Renner, S. (2001). "A 'Community of Interest' Approach to Data Interoperability" in *Federal Database Colloquium*.
- Simulation Interoperability Standards Organization Sun, Y. & Guoqing, L. (2008). "Interoperability Research of Heterogeneous GIS Based on Spatial Information Grid" in *2008 International Conference on Computer Science and Software Engineering*, vol.3, no., pp.41-44.
- Tolk, A. & Muguira, J. (2003). "The Levels of Conceptual Interoperability Model" in *Fall Simulation Interoperability Workshop*.
- Wu, P. et al (2008). "The Architecture of a Virtual Grid GIS Server" in *Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments*. *Proceedings of the SPIE*, vol. 7143, pp. 71432A-71432A-9.

1 For more details see Renner (2001): A "Community of Interest" Approach to Data Interoperability

2 NATO C3 Technical Architecture Reference Model for Interoperability

3 See ArcGIS Data Interoperability
4 For example, see the US National Incident Management System and the DCC Briefing Paper on Data Quality and Accreditation

5 For example, Parent, C. "Database Integration: The Key to Data Interoperability" in Papazoglou, M. *et al* (ed.s) Advances in Object-Oriented Data Modeling (MIT Press, 2000) pp. 222-224 and Open XML
6 See Community-based Data Interoperability Networks (INTEROP)

7 For more details see Renner (2001): A "Community of Interest" Approach to Data Interoperability
8 Tolk, & Muguira (2003)
9 Tolk, & Muguira (2003)

The DCC is funded by

Joint Information Systems Committee