

An XML Implementation of the Genealogical Data Model

Hans Fugal (hans@fugal.net)

November 27, 2001

1 Background

All commercial and freely available genealogical software today supports a conclusional paradigm: Users do research and enter their conclusions in the software. Genealogical research, when done well, is a process of first collecting evidence and then making assertions that lead to conclusions. Documenting the sources used to draw these conclusions is vital. The only support software in the past has provided for recording such documentation was a “notes” field; but now the modern genealogical community is more aware of the need for proper documentation, so most modern genealogical software provides more advanced methods for doing so. However, many beginning genealogists still document only as an afterthought or as an aside. The conclusional paradigm fails to handle these and other documentation issues. In 1998, the Lexicon Working Group, created by GENTECH and the FGS, released an RFC (Request For Comment) describing the Genealogical Data Model (GDM)[1], which addresses these issues, and others, and provides a new paradigm for understanding genealogical data and the genealogical research process. The GDM is a valuable and sound model of genealogical information and processes, with great potential for changing the way we store genealogical information. While the GDM does not strive to change the way a professional genealogist does research, the influence of software based on the GDM may indeed change—for the better—the way an amateur genealogist does research. The GDM is not a database schema or a data structure; it is simply a logical data model. It was created as such to avoid being influenced by the details of implementation or by the limitations of technology. The Lexicon Working Group wished to leave implementation of the model up to other researchers and developers.

2 Proposal

My research project will implement the GDM using XML (Extensible Markup Language.) Reaching a practical software implementation of the GDM requires the following: (1) an internal data representation that fits the GDM and (2) an effective and standard scheme for communication between users and between programs. My XML proposal addresses the communication requirements. It follows the lead of GENTECH’s LexXML project. LexXML was a project to develop an XML implementation of the GDM. I was unable to find anything more than a mention of the LexXML project on the Web, so I contacted Beau Sharbrough, GENTECH president, who told me that the project has seen little to no activity recently and encouraged me to pursue the idea.

XML is ideal for an implementation of the GDM for several reasons: It is human-readable, extensible, flexible, hierarchial, and a quickly growing standard in the industry. Another significant reason is that the Church of Jesus Christ of Latter-day Saints has announced their intentions to migrate to XML for data storage and communication.[2] It is my desire to lay some groundwork for that effort.

3 Methodology

I will develop an XML DTD (Document Type Definition) that will describe the GDM. The GDM is well-developed and concisely stated so creating an XML DTD should be straightforward. There will be issues of implementation to consider, and there may be some questions that will take serious deliberation. When I come across these questions I will solicit the volunteer help of GENTECH volunteers and the Lexicon Working Group, as well as other email lists such as GEDCOM-L that

may show interest. If the nature of the question is appropriate, I will draw upon resources of the Family History department faculty here at BYU.

The GDM leaves some areas for future research, namely the expert systems for person names, place names, and dates. I will provide a basic implementation of the data of these systems that will be easily extended or replaced by future research in these areas, and should be able to handle common cases for most basic (amateur-level) genealogical research. I do not intend for these systems to be in final form—they will each require a great deal of research and effort. I do hope to provide a basic and working implementation that can show the potential for an XML representation of the GDM and which may prove to be a useful foundation for these expert systems.

4 Anticipated Results

This XML implementation will be the first practical implementation of the GDM. It may bring to light some problems or questions about the GDM which can then be ironed out by the Lexicon Working Group. It will allow for practical and standardized sharing of data represented by the GDM and should prove to be valuable in continued research in genealogical data and systems. This implementation will hopefully be a springboard to more uses of the GDM including relational-database-driven software in the long term. Such software will give the genealogist a great deal of power in searching, organizing, and understanding data. It will more closely fit the needs of professional researchers and if designed correctly will still be easy for amateur researchers while instilling proper research practices as they learn.

References

- [1] Lexicon Working Group: *GENTECH Genealogical Data Model*, <http://www.gentech.org/gdm/> (1998)
- [2] Randy Bryson, The Church of Jesus Christ of Latter-day Saints: At the *GENTECH 2001 Conference*, Dallas, TX (2-3 Feb 2001)