A User's Perspective:
Developing a Universal Metadata Structure for Genealogical Content Providers

D. Joshua Taylor, MA, MLS | jtaylor@brightsolid.com

The Need for a Universal Standard
As genealogy continues to shift away from traditional research methods, such as microfilm and printed books, to digital images and online resources, it is essential that the field communicates and agrees upon a universal metadata standard to ensure continual access and long-term preservation of our records. The fundamental need to ensure record access is key to the success of every genealogist across all generations and is in part ensured by the creation of such a standard.

Within the past several years, organizations and individuals have begun to accept the limitations of the GEDCOM standard in handling genealogical data. In addition to its age, the GEDCOM standard does provide adequate integration with digital images, particularly when handling online records and source citations. Organizations such as BetterGEDCOM <http://bettergedcom.wikispaces.com/> have begun the process of refining the existing standard to meet current practice. However, there remains a large gap between digital images used by genealogists and the data stored and collected by genealogists. This new standard should integrate with a GEDCOM-like schema on all levels.

With the adoption of a universal standard, genealogical users will be able to easily add digital images and objects to their research from a variety of sources, increasing the demand for genealogical content providers. Newer generations of genealogists will become more engaged as genealogy's methodological elements are simplified, making the field more accessible to new hobbyists. In addition, the long-term preservation of the world's most precious resources will be greatly enhanced by the adoption of the standard.

Further, a universal standard will enhance collaboration, among key content providers within the genealogical field and potentially allow the field to further integrate itself with other communities, including archives, libraries, and other disciplines including social sciences, history, and cultural studies.
Major Roadblocks
Due to the complexities of genealogical research and methodologies, as well as the nature of the data itself, the following roadblocks must be overcome as the field moves towards a universal standard.

1. Source Citations and Templates
Source citations are a contentious area for genealogists, as the field does not follow a single standard despite the presence of a few leading guidebooks.

Potential Solution(s): The acceptance of basic components - rather than strict format - for genealogical citations within the new standard. The standard should provide advanced flexibility in line with creating detailed source citations but not rely upon their presence for successful implementation.

2. Multiple Non-Profit and Commercial Data Providers
The number of data providers within the family history field has grown tremendously over the past three years and will continue to grow as the hobby becomes more popular across the world. In addition, the vast quantity of data available to be digitized or is already digitized represents an immense amount, leaving room for new providers in the space.

Potential Solution(s): The formation of a concrete consortium of major genealogical providers from the non-profit and commercial space to collaborate with the larger genealogical community in an agreement to adopt the universally created standard. In addition, the standard must allow for easy and cost-effective migration from current data structures to the new standard. This consortium must work carefully within the non-profit space to ensure smaller organizations and state archives are able to adopt the new structure.

3. Genealogical Data is Unique and Complex
Very few genealogical documents are the same. Each contains various data and are housed and created by different repositories. It is impossible to create a universal data structure that encompasses all types of genealogical documents without allowing for some degree of customization. However, with customization the idea of a universal structure begins to fall apart.

Potential Solution(s): The recognition that a universal standard must be basic enough to encompass the essential elements of genealogical documents, while allowing for a small degree of customization to ensure unique data sets are preserved and catalogued as such. The possibility of an advanced version of the universal standard exists, if needed and agreed upon by the larger community.
Sample Models
Multiple fields have successfully adopted universal metadata standards, though not without a considerable amount of discussion and debate. The genealogical community needs to look at each model and implement successful elements as appropriate. While this method saves time and debate, it also potentially leads to a deeper level of integration with other disciplines that might have an increased interest in the genealogical data set, fostering larger collaborative efforts with other disciplines. Other models to examine include:

DDI (Data Documentation Initiative) <http://www.ddialliance.org/>
Scope: International (Archives/Social Science)

DublinCore <http://dublincore.org/>
Scope: International (Libraries)

EAD (Encoded Archival Description) <http://www.loc.gov/ead/>
Scope: International (Archives)

MARC (MAchine-Readable Cataloging) <http://www.loc.gov/marc/>
Scope: International/US (Libraries)

METS (Metadata Encoding & Transmission Standard) <http://www.loc.gov/standards/mets/>
Scope: International/US (Libraries)

MODS (Metadata Object Description Summary) <http://www.loc.gov/standards/mods/>
Scope: International/US (Libraries)

ONIX (ONline Information eXchange) <http://www.editeur.org/8/ONIX/>
Scope: International (Book Publishing/Online)

Some Essential Element Areas
While far from complete, the areas enumerated below demonstrate essential elements to ensuring the success of the universal schema among content providers. A key component to the structure will be the integration and addition of the genealogical module. The following are meant to serve as a point for beginning the discussion:

Localities: Standardizing the localities applicable to the document including town/township, city, county, state, country, and geocoding.

Titles: Incorporating a record’s official title as well as the addition of a standard referenced title to ensure consumer accessibility.

Author/Creator: The author/creator of the record set, including provisions for individuals, unknown authors, government entities, and others.
Creation Date(s): Providing for an accurate description of the date (likely years) included in the record.

Original Repository: The original, official repository of the data set, such as a government archive or civil repository.

Original Call Number: Reference to microfilmed or printed versions of the data sets including call numbers from major repositories such as the Family History Library, National Archives and Records Administration, etc.

Digital Creation Details: Including the date, method, and format of digitization for long-term preservation needs.

Data Owner/Copyright: Detailing the owner of the data, and the holder of its copyright (for all levels, as needed).

Standard Citation: Including sample citations from universally accepted formats, such as the Chicago Manual of Style and Evidence Explained.

Content Provider: Providing key details in regards to the content’s online location(s) and its provider(s).

Genealogical Module: An intricate optional module encompassing genealogical data to ensure the schema can be integrated with family history software and other uses as deemed necessary by genealogists. An expanded version of this module may become the next generation of the GEDCOM format and could include specific details such as individual facts (birthdates, places, events, etc.) as well as analytical genealogical data.

Next Steps
While the task may seem daunting, there are a few key steps that can be taken to promote a standard:

1. Engage Genealogists: Begin talking with other genealogists about their needs and desired input towards a universal structure.

2. Engage Content Providers: Lobby existing content providers to form a consortium towards creating a universal standard.

3. Preserve Ideas: Form a collaborative environment to capture and preserve input from genealogists of all levels towards the new standard.