

CFPS 91

(Call for Papers Submission number 91)

A substantive vision and proposal to validate and ratify the GEDCOM X specification set

Submitted by: Heaton, Ryan

Created: 2013-03-29

URL: Most recent version: <http://fhiso.org/files/cfp/cfps91.pdf>
This version: http://fhiso.org/files/cfp/cfps91_v1-0.pdf

Description: A substantive vision for FHISO, and a proposal that FHISO's first step in fulfillment of this vision be to validate and ratify the GEDCOM X specification set.

Keywords: GEDCOM X, vision

March 29, 2013

To FHISO and its Founding Members:

I write to submit a substantive vision for FHISO, and to propose that FHISO's first step in fulfillment of this vision be to validate and ratify the GEDCOM X specification set.

I'm excited about the commendable work that FHISO has done to gather broad community representation in support of genealogical standardization efforts. I'd like to submit for your consideration my hopes for some of the processes to be standardized:

1. The **genealogical research process**, including gathering information, selecting evidence, making conclusions, justifying those conclusions with a proof statement, and citing the supporting sources for each step along the way.
2. **Constrained vocabularies** for genealogical data exchange, including the definition of vocabulary elements for fact types, record types, name types, relationship types, citation fields, etc. Such vocabularies will enable application developers to exchange such data across geographic and cultural boundaries. Such vocabularies will also enable innovative development of expert systems that can augment the capabilities of researchers.
3. Standard **Web service APIs** that can be implemented by online genealogical data providers so that record managers and other client applications can write programs that work the same way with different providers. For example, standardizing the parameters and result format of a search operation would allow a user to use a single query to search thousands of genealogical data providers.
4. The **record indexing process**, whereby providers that support the ability to index field-based records (such as as census records) can exchange data and make such indexes more widely available.
5. **Citation templates** whereby elements of a citation can be exchanged between systems in a style-independent way such that applications can provide research citations to their users according to their preferences.
6. **Research microdata** that can be embedded in a research document (such as a proof statement) in such a way that an automated system can provide automated tools to evaluate the document.
7. **Research plans and logs** to enable greater collaboration and to reduce duplicated effort among researchers.
8. **Research metadata for images** so that a digital image can carry its own citation, notes, etc. so that information doesn't get lost when images are shared.

The above list is obviously incomplete, but it suggests how vast and deep are the possibilities for standardization. We have so much to do.

The GEDCOM X project is an open project offered to the community as a way to meet the above-mentioned needs. The current focus is limited to item #1 (the genealogical research process) and we expect to wrap up a milestone release to that end over the next few months. The project has been designed to accommodate extensions that will address the other items. We are actively developing provisions for #2 (constrained vocabularies), #3 (Web service APIs), and #4 (record indexing processes). We have given significant thought to high-level conceptual designs for #5 (citation templates) and #6 (research microdata).

GEDCOM X is partitioned in such a way that concerns are divided into separate specifications. The GEDCOM X project is currently developing the following specification set:

- The **conceptual model** specification establishes a set of abstract data types and concepts that are used to model the genealogical research process. It also defines a basic set of constrained vocabularies to provide a framework for exchanging data.
- The **date format** specification defines a way to exchange historical date information.
- The **XML media type** specification defines an XML representation of the conceptual model.
- The **JSON media type** specification defines an JSON representation of the conceptual model.
- The **file format** specification defines how to bundle genealogical data into a single file on a filesystem.

There are also other GEDCOM X specifications being developed for Web service APIs and record indexing.

GEDCOM X is the best option among many to address the needs of the genealogical information community. GEDCOM X is being actively implemented at FamilySearch, and FamilySearch will continue to integrate GEDCOM X into its products and services. The FamilySearch developer platform uses GEDCOM X to exchange data with the partners and affiliates of FamilySearch, so there are already a significant number of record managers and data providers that have at least partially implemented it.

By validating and ratifying GEDCOM X, FHISO can position itself as a leader in genealogical information standardization. It will be able to show a major accomplishment in providing a clear successor to GEDCOM 5 to which the industry can migrate, and FHISO will be in a strong position to lead additional efforts to address standardization needs that extend beyond the basic genealogical research process.

There is much work to be done. By leveraging the significant amount of work and effort that have been applied to GEDCOM X, the genealogical information community can take a massive step forward with a minimal amount of additional resources. From there, resources can be applied to drive real and exciting change in innovative areas of genealogical data exchange, bringing the exciting and noble work of genealogical research to an entirely new level.

Thank you for your consideration.

Ryan Heaton