Spatial Focus, Inc.

Tools for Practical Implementation of the United States Thoroughfare, Landmark and Postal Address Data Standard.

Table of Contents

Decatur Office: 436 Avery Street ● Decatur, GA 30030
Voice: 404.378.0989 ● FAX: 404.806.6257
Home Office: 6813 40th Avenue ●University Park, MD 20782

1 Introduction

This project is developing prototype implementation and training tools for the newly adopted United States Thoroughfare, Landmark and Postal Address Data Standard (Address Standard). The Address Standard is complex, and consists of four related parts: Address Data Content, Address Data Classification, Address Data Quality and Address Data Exchange. Implementation is expected at the Federal, State, and local government levels, and there are significant differences in the overall strategy of implementation at each of these levels. Private sector implementation by address aggregators and by software vendors is also expected, and will further support adoption and implementation by government entities.

Adoption of the standard depends on its accessibility. This project has chosen to produce prototype software tools to tackle the accessibility problem, and to further knowledge of the standard. The concept is one of creating materials for users of the standard to learn while doing. Designs and prototypes for two sets of tools are envisioned: desktop tools for data import, quality control and export, and a web application, targeted to smartphones but deployable on desktops as well, for quick information about the Address Standard. The former will help demystify the more technical aspects of the Standard, while the latter is geared towards helping users plan implementations. The tools will be called “desktop module” and “smartphone app”, respectively, for simplicity.

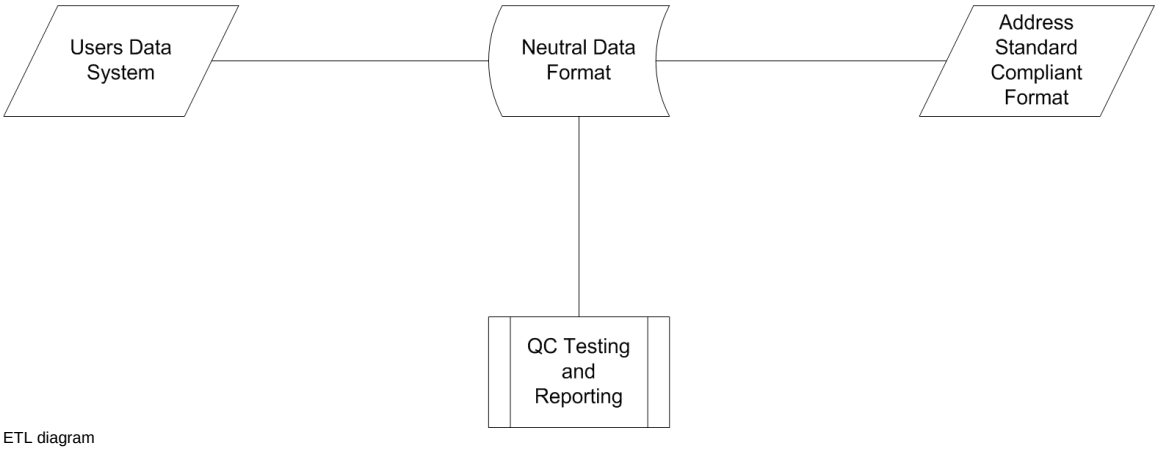
This document identifies requirements for each set of tools.

2 Desktop Module

2.1 Purpose

Tools to implement the Address Standard need to be flexible and extensible, to support the wide variety of both address data systems and address assignment systems in the United States. In the first phases the tools should support the most common cases of address data systems and address assignment systems. Likewise the implementation of Address Standard Quality Control reporting will need to be flexible enough to handle a wide variety of address data systems.

To maximize flexibility an ETL (Extract Transform and Load) style of information processing will be utilized. In this style, address data will be extracted from the user's data information system, from the user's data format, into a neutral data format for both QC testing and reporting and for publication into the Address Standard format. Incoming address data will be imported into the neutral format, QC tests will be run and reported and the data can be produced in a format similar to the users own data format for further importing steps.



ETL diagram

2.1.1 Profile of agencies and individuals implementing the Address Standard and their data.

Currently, agencies that are starting to implement the Address Standard are focused on publishing data that falls into one of a few categories:

- Geographic street centerline address data with number ranges associated with the left and right sides
- Geographic address point data representing a single aspect of an address or structure containing address(es)
- Tabular address data with 1 number range.

These same agencies are State, County and Local government and non-governmental organizations who possess address data and wish to share the data with their strategic data partners. The consumers of such data are state and regional data aggregators in addition to national data aggregators.

The agencies have a need to share address data in a way that uncertainties surrounding the data are minimized and data defects are known or communicated.

2.2 Requirements

Based on discussions with agencies the QC tests are most helpful when they:

- Compare the number of things that have addresses with the number of addresses in the database
- Compare the street names, parity and sequence of addresses to adjacent

roadways

- Compare the ranges to individual address points to check for consistency
- Check for completeness and agreement in range sources
- Compare Address tables for missing values

Initial Data Types to be Supported:

- Four Number Address Range - Thoroughfare Class
- Two Number Address Range - Thoroughfare Class
- Numbered Thoroughfare Address - Thoroughfare Class

Initial QC tests to be supported

- Address Completeness Measure
- Address Number Fishbones Measure
- Address Number Range Parity Consistency Measure
- Range Domain Measure
- Tabular Domain Measure

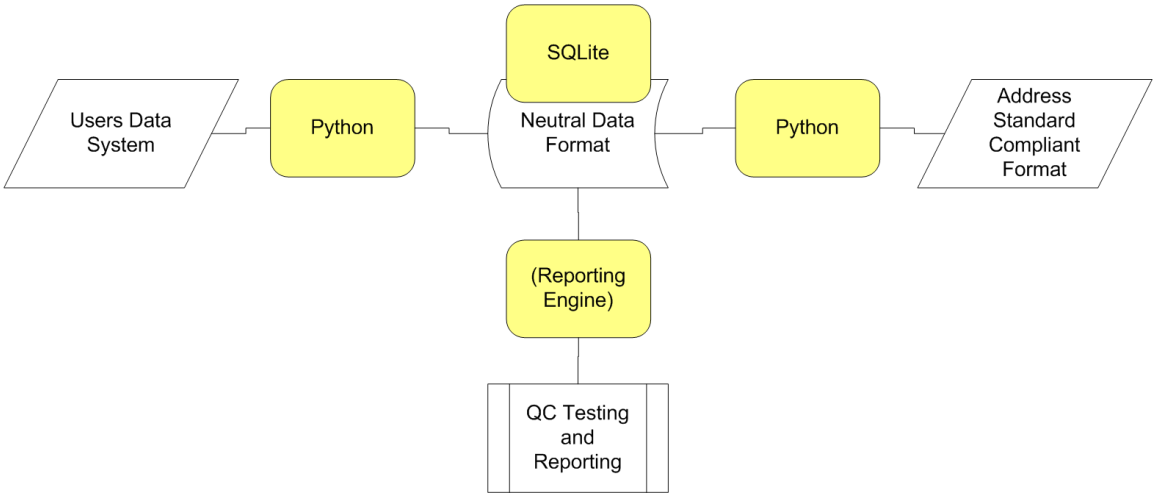
2.3 Software for Building the Module

To provide maximum reuse and extensibility a modular system is envisioned that will allow additional import, export and QC reporting paths to be developed and deployed over time. Such a system would build upon existing modular frameworks and be establish in such a way that integration into existing workflows and software would be possible. Three main components are required:

- Scripting language with the following capabilities
    - XML (eXtensible Markup Language) processing modules
    - XSLT (eXtensible Stylesheet Language Transformations) processing modules
    - DOM (Document Object Module) processing modules
    - SQL (Structures Query Language) database interoperability modules

- Embeddable SQL capable database
  - Spatialite geoprocessing modules
- Embeddable reporting engine

The three components are utilized to connect the points of the ETL diagram



To maximize use of the tools, open source or freely redistributable software was considered for each of these components. The initial development will utilize Python as the scripting language and SQLite as the embeddable database. The reporting engine has not yet been finalized but Dabo and BIRT are still being evaluated.

3 Smartphone App

3.1 Purpose

The purpose of this module is to create a smarter address standard document for both desktop and mobile web browsers. A smarter document is key to adoption of the standard, due to several serious barriers:

- The size of the standard
- Vocabulary defined in the document
- Parts of the standard that are inaccessible without familiarity with its vocabulary.

The standard is dauntingly large, and its emphasis on content led to a rich lexicon that can be unfamiliar to new users. A more accessible document can encourage use by presenting more digestible, user-selected portions of the standard. These smaller

portions of the document can familiarize users with the address standard vocabulary on a more manageable basis.

3.2 Requirements

3.2.1 Search Facilities

Hyperlinked terms and full text search capabilities in the wiki version of the standard makes it more approachable. Maintaining those facilities and adding menu-driven selections can make the document far more useful, especially to those outside the community that maintains the wiki. This approach focuses on address standard terms, helping new users with its basic vocabulary. Simply using the terms in a menu system gives users a more intuitive feel for the elements and attributes, their meanings and uses.

Greater familiarity with address standard vocabulary can bring more complete adoption, providing users with the tools for approach the Address Data Quality and Address Data Exchange sections of the document. Each of these sections assumes the user fully understands the address standard dictionary. A thorough grasp of the constituent parts of addresses provides users with the tools to select the Data Quality measures and Data Exchange schemas that apply to their situation.

3.2.2 Sections that Toggle Visibility

Each part of the standard is organized into subsections. Search results or menu-driven look-ups will be presented with only the section headings showing. Users can toggle each section on as needed. Particularly for long entries, this will promote consumption of the standard in smaller portions and avoid overwhelming users.

3.2.3 Index

Appendices relate various parts of the standard, illuminating element relationships, quality measures and XSDs. It lacks, however, a consolidated index that would not require the reader to classify a term before finding it. This would create another entryway to the document for an inexperienced user of the standard.

3.3 Software for Building the App

Software required to implement this application include:

- Mobile applications framework
    - Ability to create cross-platform apps using HTML and javascript
- Tool or method for HTML validation
- Javascript library
- Scripting language for user functions

4 Conclusion

The Address Standard has a number of barriers to adoption. A smarter document can provide multiple pathways into the document that do not currently exist. It will help build knowledge about the standard, and help users to better define tasks to perform with the proposed Desktop Module. Once users have defined tasks for implementing the Standard, the Desktop Module will simplify their execution.