Developing a Comprehensive Address Data Standard for the United States

Address Standard Working Group:

Martha McCart Wells, GISP, Spatial Focus, Inc.
Carl Anderson, Fulton County, Georgia
Hilary Perkins, GISP, Data Transfer Solutions, Inc.
Ed Wells, GISP, Washington (DC) Metro Area Transit Authority
Sara Yurman, Spatial Focus, Inc.

ISO TC 211 U.S. Technical Advisory Group:

Julie Maitra, Federal Geographic Data Committee
Randy Fusaro, Census Bureau

ISO/TC211 Workshop on Address Standards

Copenhagen, Denmark

May, 2008

Purpose and Scope

- *"United States Street, Landmark, and Postal Address Data Standard"*
- Draft data standard for U.S. addresses
- In preparation by the Address Standard Working Group
- For submittal to U.S. Federal Geographic Data Committee
- One standard in four parts:
  - Data Content
  - Data Classification
  - Data Quality
  - Data Exchange

Organizing Principles

- Definition of an address:
  - "An address specifies a location by reference to a thoroughfare, or a landmark; or it specifies a point of postal delivery."
- Syntactical approach to address classification.
  - The standard classifies addresses according to their address elements and the order in which the elements are arranged.
- Address assignment and daily usage are local in nature.
  - Addresses are based on local schemes for naming and numbering.
  - There is information about the address that is vital to its many uses.
- The quality of address data must be measured and recorded.
- Address data must be able to be seamlessly exchanged between different users.
- All of these must be incorporated into a comprehensive address data standard.

Address Elements

- **Address numbers**
- **Street names**
- **Occupancies**
- **Landmark names**
- **Larger areas** (place names, states, postal codes, and country names)
- **USPS postal address elements**
- **USPS address lines**
- **Address Scheme Elements** (grid, axis, numbering rules)

Address Attributes

Purpose: documentation, mapping and quality control

Key attributes include:

- **Address identifier**
- **The address authority, dataset, and start and end dates**
- **Geographic coordinates and linear referencing**
- **Lifecycle and official status**
- **Class**
- **Feature type**
- **Attributes for quality control** (parity, sequence, relationships, etc.)

Taxonomy of U.S. Address Classes

Thoroughfare Address Classes

- Numbered Thoroughfare Address: 123 Main Street
- Intersection Address: Fifth Avenue and Main Street
- Two-Number Address Range: 405-411 West Green Street
- Four-Number Address Range: 900-962, 901-963 Milton Street
- Unnumbered Thoroughfare Address: Forest Service Road 698

Landmark Address Classes

- Landmark Address: Statue of Liberty
- Community Address: 123 Urbanizacion Los Olmos

Postal Delivery Address Classes

- USPS Postal Delivery Box: PO Box 16953
- USPS Postal Delivery Route: RR 1, Box 100
- USPS General Delivery Office: General Delivery

Data Quality

- Address schemes
  - The local geographic framework and business rules for address assignment.
  - Basis for testing the validity of an address
- Data quality
  - A complete suite of data quality tests.
  - Tests of data relative to business rules
  - Tests of address location relative to geographic scheme and known features

Data Exchange

- Requires open, standardized format:
  - XML Schema Document (XSD) and XML.
  - Protects data producers and consumers
  - Allows localizations, but provides standard form for exchange.
- A data model, but not a database model.
  - Organizational database requirements and relationships vary considerably.

Standard Development Process

- Sought broad awareness and participation
  - Wiki collaborative website
  - Teleconferences
- Posted drafts for public comment via web form
- Focused on practical needs and usefulness
  - Local emphasis: Where addresses are created and used the most
  - Quality is integral to address use, must be built in to process
- Included both tabular and geospatial data

Findings and Results

- In the US, addresses are primarily created and maintained locally.
- In the US, address data is often poorly understood, badly recorded, and badly documented.
- The ASWG standards development process has sought to develop a comprehensive view of addresses to support the organization, documentation, validation, and exchange of information.
  - Broad participation provided information about differing address practices throughout the U.S.
  - Testing of the standard in real situations insured its usefulness and viability.
- The standard is intended to support the full range of address data needs, at all levels of government and in the private sector.

What Is An Address?

Which of These Are Addresses?

- ***Rentemestervej 8, 2400 Copenhagen NV***
- ***55.704698°N, 12.535380°E***
- ***Copenhagen, Rentemestervej 180m +10***

Which Are Excluded by These Definitions of "Address":

- **ASWG:** *"An address specifies a location by reference to a thoroughfare or a landmark, or it specifies a point of postal delivery."*
- **BS7666:** *"means of referencing an object for the purposes of identification and location"*
- **OASIS v3:** *"A physical location or a mail delivery point"*
- **SANS 1883-1:** *"an unambiguous specification of a point of service delivery"*

Proposed Typology of Address Standards

Three Broad Classes:

1. Postal or Delivery Address Standards
2. Address Gazetteer Standards
3. Address Data Management Standards

Note: Based on review of BS 7666, draft SANS 1883, ASWG draft, UPU v.2006, USPS Pub. 28, and OASIS CIQ v.3

Postal or Delivery Address Standards

Purpose: To support standardizing and formatting addresses to ensure correct delivery of mail, parcels, etc.
Premise: Addresses are text strings to be parsed, standardized, matched against an internal master list, and formatted for mailpiece labels.

Characteristics:
1. Geographic coordinate location is of secondary importance or omitted.
2. Typically one simple, general record structure is defined for all addresses (UPU is an exception).
3. There are clear and detailed rules about how the address must be formatted on a mailing label (OASIS is an exception).
4. Metadata and data quality tests are minimized or omitted.

Examples: *UPU, USPS Pub 28, OASIS*

Address Gazetteer Standards

Purpose: To govern the construction of lists of geographic features and their locations.

Premise: Addresses are not strings to be parsed, but features to listed and mapped.

Characteristics:

1. Coordinate locations are mandatory for every address.
2. One record structure is prescribed for all addresses.
4. Comparatively few address elements are defined (only a dozen or so).
5. Metadata are mandated, including unique record IDs for each address, and careful attention is given to both gazetteer (file-level) metadata and record-level metadata.

Example: *BS7666*

Address Data Management Standards

Purpose: To support the creation and administration of authoritative address databases used to assign and administer address repositories.
Premise: Addresses are text strings to be parsed into one of several syntaxes for classification and entry into a geographical relational database.

Characteristics:
1. Defines all simple data elements needed to parse any address into a normalized data record.
2. Also defines how simple elements may be combined into complex elements to represent larger components of an address (e.g. complete street name)—to support less-detailed parsing.
3. Classifies addresses by their internal syntaxes, and provides a complete taxonomy of address syntaxes.
4. Supports, but does not require, provision of coordinate locations for addresses.
5. Defines attributes needed to document addresses (record-level metadata).

Examples: ASWG draft, SANS 1883 draft

Framing a New International Address Standard

The differences between the three types of standards highlight some important questions to be resolved in framing a new standard:

- **Addresses:** Are they strings to be parsed or features to mapped?
- **Elements, Parsing, and Record Structure:**
  - One flexible record structure for all addresses, with only a limited set of address elements?
  - Or a taxonomy of record structures and a detailed set of address elements to support parsing for entry into a normalized relational database?
- **XY coordinates:** Required, optional, or unnecessary for addresses?
- **Metadata:** What attributes should be defined for address documentation?
- **Data Quality:** Should an international standard specify address data quality tests?