Rethinking Gazetteers and Interoperability

Greg Janée
University of California at Santa Barbara
Outline

• ADL gazetteer protocol
• Interoperability use cases
• Gazetteer protocol, revisited
• Question for the workshop
Protocol history

• Motivation
  – define gazetteer’s role in ADL architecture
  – functional definition to complement Gazetteer Content Standard (GCS)

• Development
  – 2001: co-developed protocol with ESRI
  – 2003: minor additions, deletions
  – Retargetable server
    • full, GCS-compliant schema
    • “lite” schema
  – Web-based client
Protocol characteristics

• Defines gazetteer model
  – simplified; focus on interoperability
  – compatible with, mappable from GCS
• Seven query types

• Abstract specification
• HTTP+XML instantiation

• Separate thesaurus protocol
  – Z39.19 thesaurus model
Gazetteer model

each value has a status; one value must be designated “primary”

gazetteer
description,
...

entry

identifier
0+ codes
status
1+ names
1+ footprints
1+ types

named relationship

thesaurus references
Limitations/problems

• No support for qualified placename queries
  - e.g., find “Santa Barbara, CA”
  - onerous to implement using low-level facilities

• Conflicting and unpredictable query semantics
  - a tale of two queries:
    • spatially contained within California
    • has relationship PartOf to California
  - results are implementation-specific, unpredictable, and variable
Interoperability use cases

• Harvest
  – aggregate distributed, esp. local gazetteers
  – need
    • protocol (OAI–PMH)
    • representation standard(s) (GCS)

• Lookup
  – find place by name or other description
Example searches:

Go to a location
- kansas city
- 10 market st, san francisco

Find a business
- hotels near lax
- pizza

Get directions
- jfk to 350 5th ave, new york
- seattle to 98109

Drag the map with your mouse, or double-click to zoom. Take a tour »

Business Owners: Add/Edit Your Business
United States
You can search for an emergency contraception provider by entering state – or you can try using our interactive map of the United States.

Search by Zip Code
For the closest match enter your 5-digit zipcode:

Zip Code:  

Search

Search by Area Code
For a wider search, enter your 3-digit telephone area code:

Area Code:  

Search

Search by City and State
Alternatively, enter your city and state:
### Global Land Cover Facility

**Earth Science Data Data Interface**

<table>
<thead>
<tr>
<th>Sensor</th>
<th>WRS</th>
<th>Date</th>
<th>New Since</th>
<th>GeoCover</th>
<th>Level</th>
<th>Orthorectified</th>
<th>Terrain Corrected</th>
<th>Not Validated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETM+ TM</td>
<td>WRS-2</td>
<td>Start Date:</td>
<td></td>
<td>GeoCover</td>
<td>Level 1G</td>
<td>Orthorectified</td>
<td>Terrain Corrected</td>
<td>Not Validated</td>
</tr>
<tr>
<td>MSS</td>
<td>WRS-1</td>
<td>End Date:</td>
<td>Months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SRTM</td>
<td></td>
<td></td>
<td>ago</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enter dates as mm/dd/yy. 

<table>
<thead>
<tr>
<th>Start Path</th>
<th>Start Row</th>
<th>End Path</th>
<th>End Row</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A button to place a marker at the current coordinates.

The distance and direction to a few nearby place-names.

A Find box that accepts a wide variety of location specifiers:
- street address (USA only)
- city name (world wide)
- USA ZIP code
- latitude/longitude in any common format
- URL of a flickr photo that has geotags

A button to bring up the Markers Panel, where you can see all your current markers and
Use cases, cont.

• **Lookup**
  - find place by name or other description
  - broader than gazetteer: find place by *anything*

• **Reverse lookup**
  - find nearby places
  - find nearby places of a given type
  - find *nearest* place of a given type
**Smart Search**

Enter an airport name, city, state, country, or airport code, then press "Search."

---

**Browse by Country or Continent**

Select a country and press "View."

- Albania
- Algeria
- American Samoa
- Angola
- Anguilla
- Antigua and Barbuda
- Argentina

---

**Airport Code Alphabetical Listing**
Use cases, cont.

• Reverse lookup
  – find nearest place of a given type

• Geoparse
  – identify & geolocate place references in document
  – eg., GeoNames.org RSS-to-GeoRSS converter
Recent crime reports for Berkeley, CA

- 2006-12-04 08:00AM Burglary Res'l Cedar St
- 2006-12-02 05:00PM Burglary Auto Walnut St
- 2006-12-02 05:00PM Malicious Dam Block Piedmont Av
- 2006-12-02 02:00PM Malicious Dam Block Grayson St
- 2006-12-02 01:44PM Loud Noise @ Telegraph Av
- 2006-12-02 11:30AM Petty Theft @ Solano Av
- 2006-12-02 11:24AM Loud Noise @ Durant Av
- 2006-12-02 11:00AM Malicious Dam Block 8th St
- 2006-12-02 11:00AM Grand Theft @ Milvia St
- 2006-12-02 08:44AM Burglary Auto
Use cases, cont.

• Geoparse
  – uses gazetteers, but not protocol

• Ontology
  – inferencing over knowledge base of places
  – requirements:
    • unique Ids, ontology of relationships
    • unification of facts!
## Gazetteer protocol, revisited

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvest</td>
<td>already better supported by OAI–PMH</td>
</tr>
<tr>
<td>Lookup</td>
<td>too limited, rigid</td>
</tr>
<tr>
<td>Reverse lookup</td>
<td>supports near, but not nearest</td>
</tr>
<tr>
<td>Geoparse</td>
<td>N/A</td>
</tr>
<tr>
<td>Ontology</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Question

• Should we rethink gazetteer interoperability?

• From:
  – have an entity (gazetteer), have a protocol for accessing that entity

• To:
  – multiple protocols oriented around use cases, functionality
  – that various kinds of entities (gazetteers, geocoders, etc.) participate in, implement to varying degrees
Rethinking Gazetteers and Interoperability

Greg Janée
University of California at Santa Barbara
Outline

- ADL gazetteer protocol
- Interoperability use cases
- Gazetteer protocol, revisited
- Question for the workshop
Protocol history

• Motivation
  – define gazetteer’s role in ADL architecture
  – functional definition to complement Gazetteer Content Standard (GCS)

• Development
  – 2001: co-developed protocol with ESRI
  – 2003: minor additions, deletions
  – Retargetable server
    • full, GCS-compliant schema
    • “lite” schema
  – Web-based client
Protocol characteristics

• Defines gazetteer model
  – simplified; focus on interoperability
  – compatible with, mappable from GCS
• Seven query types
• Abstract specification
• HTTP+XML instantiation
• Separate thesaurus protocol
  – Z39.19 thesaurus model
Gazetteer model

each value has a status; one value must be designated “primary”
Limitations/problems

- No support for qualified placename queries
  - e.g., find “Santa Barbara, CA”
  - onerous to implement using low-level facilities

- Conflicting and unpredictable query semantics
  - a tale of two queries:
    - spatially contained within California
    - has relationship PartOf to California
  - results are implementation-specific, unpredictable, and variable
Interoperability use cases

• Harvest
  – aggregate distributed, esp. local gazetteers
  – need
    • protocol (OAI–PMH)
    • representation standard(s) (GCS)

• Lookup
  – find place by name or other description
Maps

Example searches:
Go to a location
kansas city
40 market st, san francisco

Find a business
hotels near lax
pizza

Get directions
 JFK to JFK, 7th avw, new york
Seattle to 10th

Drag the map with your mouse, or double-click to zoom. Take a tour

Business Owners: Add/Edit Your Business
United States
You can search for an emergency contraception provider by entering state – or you can try using our interactive map of the United States.

Search by Zip Code
For the closest match enter your 5-digit zipcode:

Zip Code: [Input Field] [Search Button]

Search by Area Code
For a wider search, enter your 3-digit telephone area code:

Area Code: [Input Field] [Search Button]

Search by City and State
Alternatively, enter your city and state:
- A button to place a marker at the current coordinates.
- The distance and direction to a few nearby place-names.
- A Find box that accepts a wide variety of location specifiers:
  - street address (USA only)
  - city name (world wide)
  - USA ZIP code
  - latitude/longitude in any common format
  - URL of a flickr photo that has geotags
- A button to bring up the Markers Panel, where you can see all your current markers and
Use cases, cont.

• Lookup
  – find place by name or other description
  – broader than gazetteer: find place by anything

• Reverse lookup
  – find nearby places
  – find nearby places of a given type
  – find nearest place of a given type
Use cases, cont.

• Reverse lookup
  – find nearest place of a given type

• Geoparse
  – identify & geolocate place references in document
  – eg., GeoNames.org RSS-to-GeoRSS converter
Use cases, cont.

• Geoparse
  – uses gazetteers, but not protocol

• Ontology
  – inferencing over knowledge base of places
  – requirements:
    • unique Ids, ontology of relationships
    • unification of facts!
## Gazetteer protocol, revisited

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvest</td>
<td>already better supported by OAI–PMH</td>
</tr>
<tr>
<td>Lookup</td>
<td>too limited, rigid</td>
</tr>
<tr>
<td>Reverse lookup</td>
<td>supports near, but not nearest</td>
</tr>
<tr>
<td>Geoparse</td>
<td>N/A</td>
</tr>
<tr>
<td>Ontology</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Question

• Should we rethink gazetteer interoperability?

• From:
  – have an entity (gazetteer), have a protocol for accessing that entity

• To:
  – multiple protocols oriented around use cases, functionality
  – that various kinds of entities (gazetteers, geocoders, etc.) participate in, implement to varying degrees