The ADL Gazetteer & Thesaurus Service Protocols

Greg Janée
gjanee@alexandria.ucsb.edu

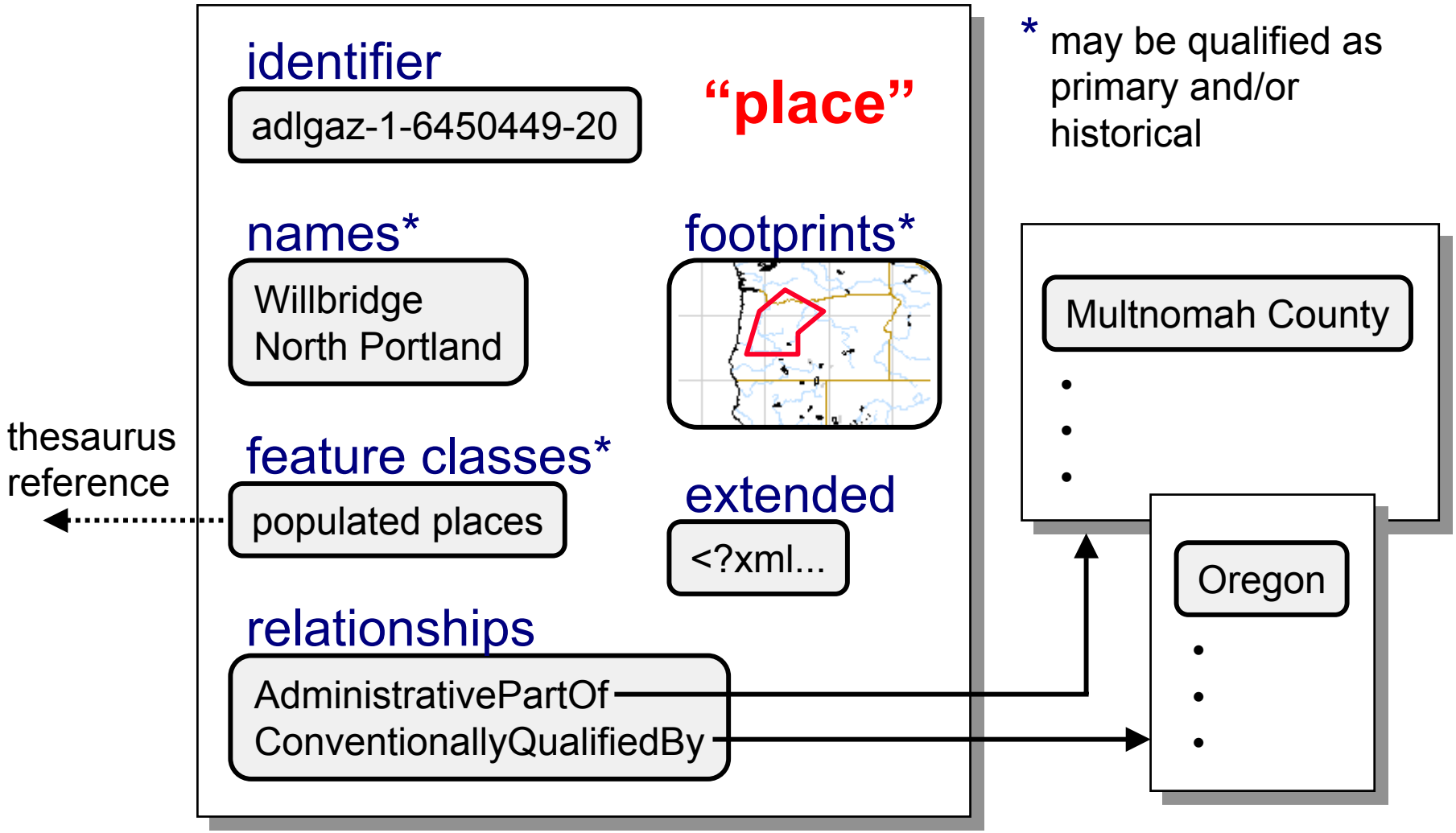
Outline

- Motivation
- ADL Gazetteer Protocol
  - model, services, query language, relationships
- ADL Thesaurus Protocol
  - model, services
- Summary

Motivation

- Gazetteer
  - more a service than a collection
  - streamlined, customized for gazetteers
- Thesaurus
  - referenced by gazetteers and DLs
- Style
  - lightweight protocols
  - buzzword-compliant

Gazetteer model



Service characteristics

- Stateless, independent, synchronous, optional
- SOAP-like, XML-over-HTTP formulation
  - HTTP POST to common URL
- Access control not addressed

Services

- get-capabilities() → *capabilities description*
  - what thesauri, services, query types, etc., are supported
- query(*query*, *report format* [, *geometry lang*]) → *reports*
  - returns all entries that match a query
  - two report formats: standard, extended
  - geometry language (GML, ArcXML, …) negotiable
- download(*report format* [, *geometry language*]) → *reports*
  - downloads entire gazetteer
- add-entry(*report*) → *identifier*
- relate-entries(*relationship*, $identifier_1$, $identifier_2$)
- remove-entry(*identifier*)

Query language

- Constraint types
  - identifier
    - **find gazetteer entry #314159**
  - name
    - **find placenames containing the phrase "San Diego"**
  - footprint
    - **find everything that overlaps a given region**
  - class
    - **find all cemeteries**
  - relationship
    - **find the capital of California**
- Booleans: AND, OR, AND NOT

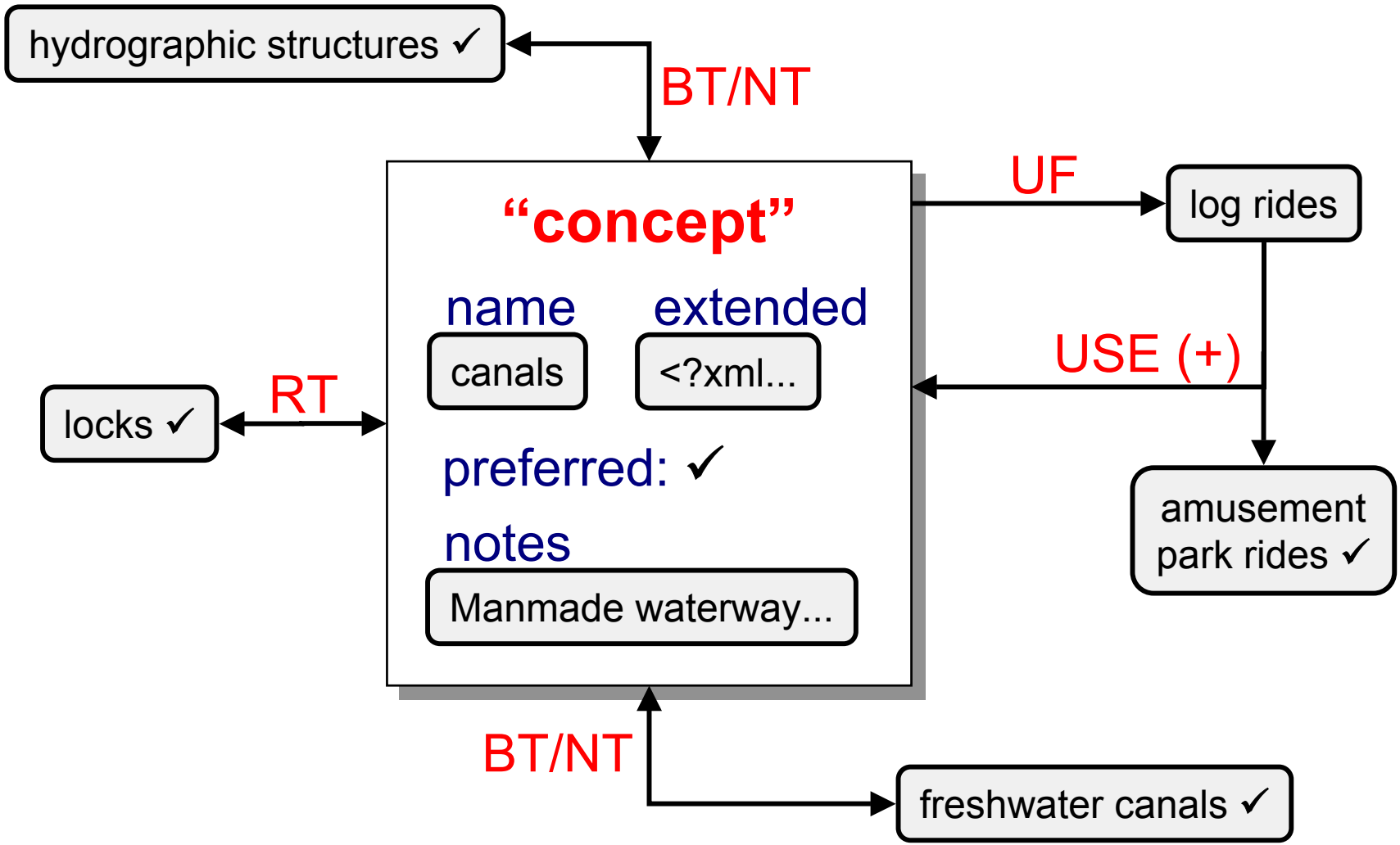
Relationships…

- PartOf
  - AdministrativePartOf
    - **AdministrativePartitionMemberOf**
    - **AdministrativeSeatOf**
    - **ConventionallyQualifiedBy**
  - SubfeatureOf
    - **GeophysicalPartitionMemberOf**
- PhysicallyConnectedTo
  - FlowsInto

Relationships… and the protocol

- Two semantically equivalent constraints
  - find place spatially within another place
  - find place PartOf another place
- Behavior depends on implementation
- Mandate equivalence?

Thesaurus model



Services

- get-properties() → *properties*
- download(*include-nonpreferred*, *format*) → *list*
  - returns list of terms
  - three formats: brief, standard, extended
- query(*operator*, *text*, *fuzzy*, *format*) → *list*
  - against term names
  - matching: word-based, equality, regular expression
- get-broader(*start-term*, *max-levels*, *format*) → *hierarchy*
  - returns DAG of terms
- get-narrower([*start-term*,] *max-levels*, *fmt*) → *hierarchy*
  - null start term yields whole thesaurus

Issues

- Identifiers
- Multilingual thesauri

Summary

- ADL Gazetteer & Thesaurus Protocols
  - lightweight, stateless, complementary
  - based on XML, HTTP
- Reference servers
  - generic, open source, J2EE-based
  - gazetteer server: maps to JDBC/SQL
  - thesaurus server: imports MultiTes
- For more information
  - http://www.alexandria.ucsb.edu/gazetteer/protocol
  - http://www.alexandria.ucsb.edu/thesaurus/protocol