



Rethinking the Meaning of Data for Integrated Science Problem Solving

Dalia E. Varanka

U.S. Department of the Interior
U.S. Geological Survey



Introduction

- **A vision of moving from compartmentalized databases to offering scientists a rich array of data for complex problem solving**
- **An example of semantic ambiguity hindering effective analysis and a possible type of solution**

Disciplines

- Geology
- Water Resources
- Biology
- Mapping

Strategies

- Energy, Minerals, and Environmental Health
- Water
- Ecosystems
- Climate and Land Use Change
- Core Science Systems
- Natural Hazards

Examples of USGS Datasets

Dataset	Geometry/ Format	Attribution/ Scaling	URL
National Hydrography Dataset (NHD)	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/nhd.html?p=nhd
National Transportation Dataset	Vector; tables	Discrete/nominal	http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
National Boundaries Dataset	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/
National Structures Dataset	Vector	Discrete/nominal	http://viewer.nationalmap.gov/viewer/
Geographic Names Information System (GNIS)	Vector	Discrete/nominal	http://geonames.usgs.gov/domestic/download_data.htm
National Elevation Dataset (NED)	Raster	Continuous/ratio	http://viewer.nationalmap.gov/viewer/ http://seamless.usgs.gov/website/seamless/viewer.htm
National Digital Orthophotos	Raster	Continuous/ interval	http://www.ndop.gov/data.html ; http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
National Land Cover Dataset (NLCD)	Raster	Discrete/nominal	http://viewer.nationalmap.gov/viewer/ http://gisdata.usgs.net/website/MRLC/viewer.htm
Global Land Cover Dataset	Raster	Discrete/nominal	http://landcover.usgs.gov/landcoverdata.php
LiDAR	Point	Continuous/ratio	http://viewer.nationalmap.gov/viewer/
Satellite images	Raster	Continuous/interval	http://edcsns17.cr.usgs.gov/NewEarthExplorer/ ; http://glovis.usgs.gov/
Hazards (Earthquakes, Volcanoes)	Graphics	Multiple forms	http://earthquake.usgs.gov/hazards/ ; http://volcanoes.usgs.gov/activity/status.php
Minerals	Vector; text	Discrete/nominal	http://mrdata.usgs.gov/ ; http://tin.er.usgs.gov/mrds/ http://tin.er.usgs.gov/geochem/ ; http://crustal.usgs.gov/geophysics/index.html
Energy	Vector; databases	Multiple forms	http://energy.usgs.gov/search.html
Landscapes and Coasts	Reports	Discrete/nominal	http://geochange.er.usgs.gov/info/holdings.html
Astrogeology	Databases	Discrete/nominal	http://astrogeology.usgs.gov/DataAndInformation/
Geologic Map Database	Vector; maps; text	Discrete/nominal	http://ngmdb.usgs.gov/
Geologic Data Digital Data Series	Maps; tables	Discrete/nominal	http://pubs.usgs.gov/dds/dds-060/
National Water Information System	Graphics; tables	Continuous/ratio	http://wdr.water.usgs.gov/nwisgmap/
Floods and High Flow	Graphics; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Drought	Graphics; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Monthly Stream Flow	Graphics; tables	Continuous/ratio	http://waterwatch.usgs.gov/new/index.php?id=ww
Ground Water	Vector; tables;	Continuous/ratio	http://waterdata.usgs.gov/nwis/gw/ ; http://groundwaterwatch.usgs.gov/
Water Quality	Graphics	Continuous/ratio	http://waterdata.usgs.gov/nwis/qw/ ; http://waterwatch.usgs.gov/wqwatch/
National Biological Information Infrastructure (NBII)	Graphics; vector; geodatabases	Multiple forms	http://www.nbii.gov/portal/server.pt/community/nbii_home/236
Vegetation Characterization	Vector; databases	Multiple forms	http://biology.usgs.gov/npsveg/
Wildlife	Vector; text; video	Multiple forms	http://www.nwhc.usgs.gov/
Invasive Species	Vector; databases; graphics, image	Multiple forms	http://www.nbii.gov/portal/server.pt/community/invasive_species/221

The Challenge

- **Structure – dizzying array of relational table data model designs**
- **Semantics – data representations rooted in ‘local’ needs**
- **Volume – redundancy of data value**
- **Access –**
 - **Physical / technical obstructions to data acquisition**
 - **Lack of models capturing human expertise**

USGS Community for Data Integration

A Semantic Web system that:

- **Acquires data from currently remote data sources**
- **Provides a single point of access to data**
- **Combines data into a single data set for download**

USGS databases

- **USGS fisheries records / State fisheries records**
- **Hydrography**
- **Geochemical resources**

National Hydrography Dataset (NHD)

- **Unite information from separate database/metadata/user sources**
- **Decrease redundancy by modeling certain attributes for reuse**
- **Create flexible models representing a variety of semantics**

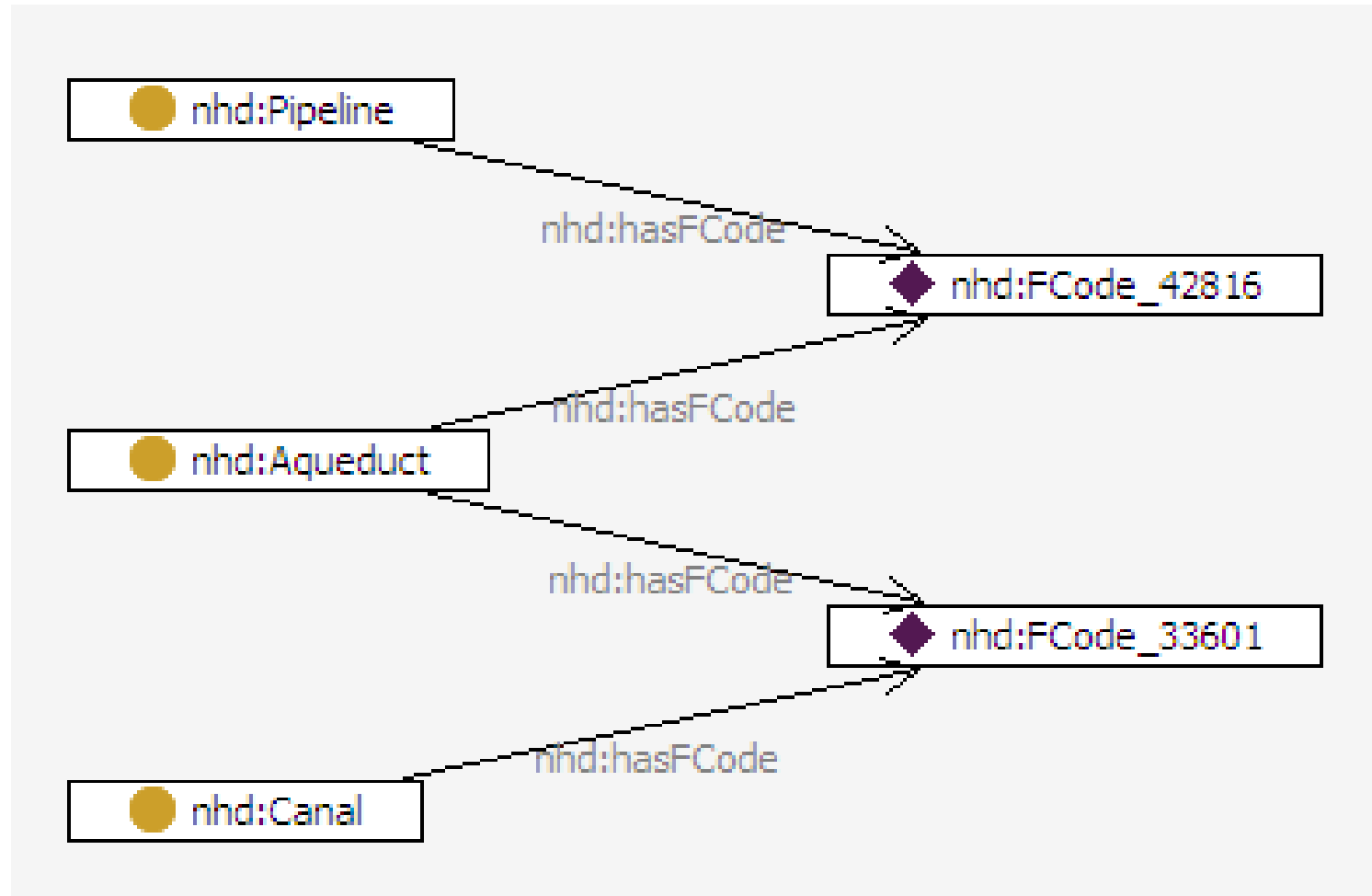
'Type' Attribute in GIS

Feature Type	FCode	Description
PIPELINE	42800	feature type only: no attributes
PIPELINE	42801	Pipeline Type: aqueduct ; Relationship to Surface: at or near
PIPELINE	42802	Pipeline Type: aqueduct ; Relationship to Surface: elevated
PIPELINE	42803	Pipeline Type: aqueduct ; Relationship to Surface: underground
PIPELINE	42804	Pipeline Type: aqueduct ; Relationship to Surface: underwater
PIPELINE	42805	Pipeline Type: general; Relationship to Surface: at or near
PIPELINE	42806	Pipeline Type: general; Relationship to Surface: elevated
PIPELINE	42807	Pipeline Type: general; Relationship to Surface: underground
PIPELINE	42808	Pipeline Type: general; Relationship to Surface: underwater
PIPELINE	42809	Pipeline Type: penstock; Relationship to Surface: at or near
PIPELINE	42810	Pipeline Type: penstock; Relationship to Surface: elevated
PIPELINE	42811	Pipeline Type: penstock; Relationship to Surface: underground
PIPELINE	42812	Pipeline Type: penstock; Relationship to Surface: underwater
PIPELINE	42813	Pipeline Type: siphon; Relationship to Surface: unspecified
PIPELINE	42814	Pipeline Type: general
PIPELINE	42815	Pipeline Type: penstock
PIPELINE	42816	Pipeline Type: aqueduct

Feature Type 'Canal'

Feature Type	FCode	Description
CANAL/DITCH	33600	feature type only: no attributes
CANAL/DITCH	33601	Canal/Ditch Type: aqueduct
CANAL/DITCH	33603	Canal/Ditch Type: stormwater

Conceptual Diagram



Semantics

Is 'Aqueduct' an attribute?

**Not as a noun; attributes are characteristics or qualities.
An aqueduct could be attributed (verb) to a pipeline.**

Is 'Aqueduct' a subclass (a member of a set) of 'Pipeline' or 'Canal?'

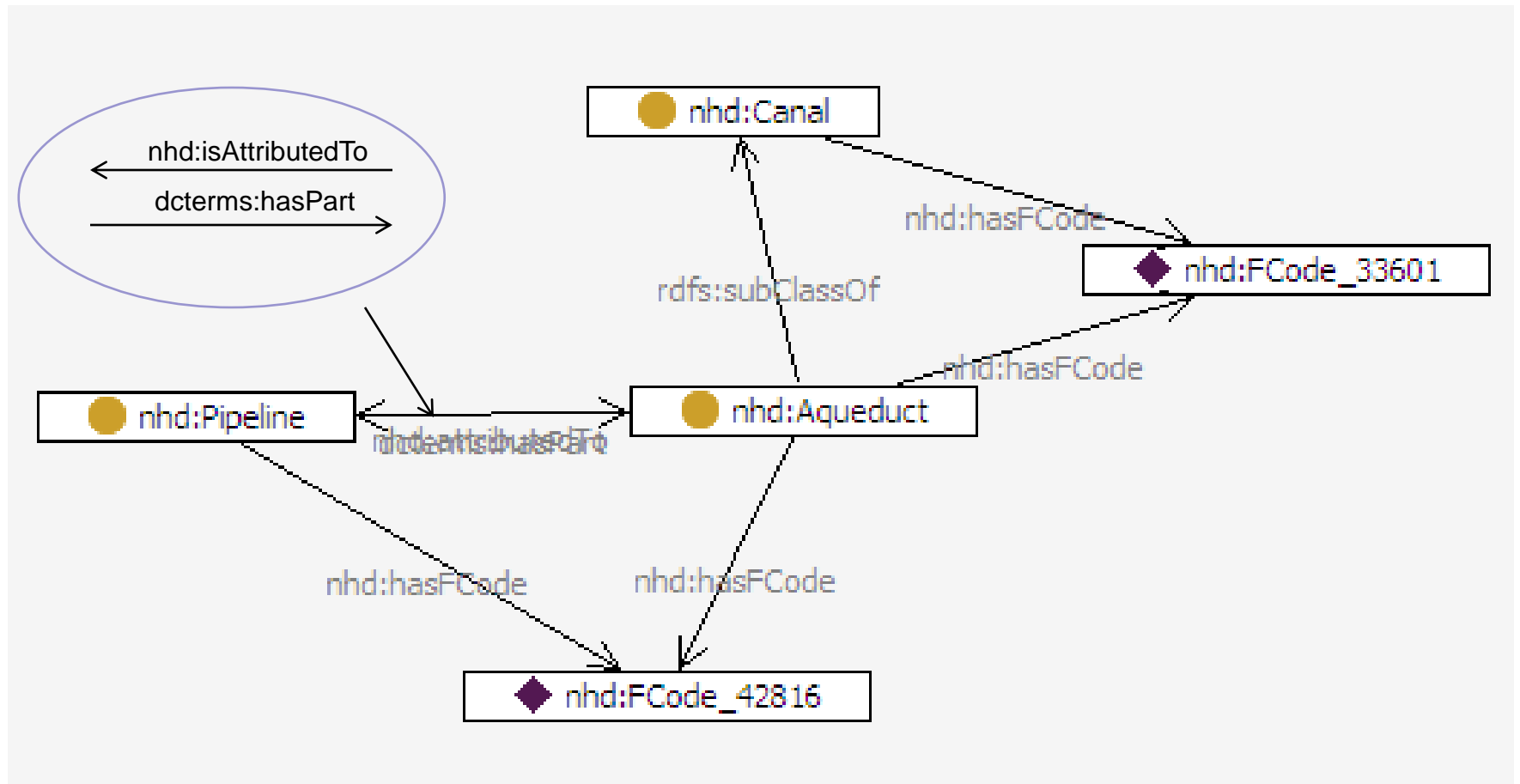
It could be a type of a pipeline or canal

Is 'Aqueduct' a part of a 'Pipeline' or 'Canal?'

It could be a feature along the pipeline or canal network

Can 'Aqueduct' be all of these?

Semantically Specific Model



Summary

- **Constrained data or fragmented specifications**
- **Ontology of seamlessly integrated semantics**
- **Sharper scientific focus**