

# *SIM* – *DL<sub>A</sub>*: A NOVEL SEMANTIC SIMILARITY MEASURE FOR DESCRIPTION LOGICS REDUCING INTER-CONCEPT TO INTER-INSTANCE SIMILARITY

Krzysztof Janowicz and Marc Wilkes

Münster Semantic Interoperability Lab (MUSIL)  
Institute for Geoinformatics  
University of Muenster, Germany  
{janowicz|marc.wilkes}@uni-muenster.de

June 4, 2009

## 1 SEMANTICS IN GISCIENCE

- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

## 2 SIMILARITY (REASONING) IN GISCIENCE

- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

## 3 SIM – DL<sub>A</sub>

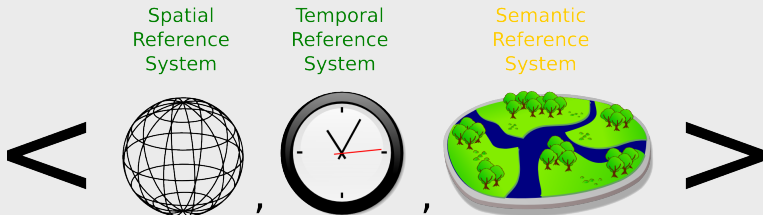
- Why SIM – DL<sub>A</sub>?
- The Similarity Tableau for *SHI*
- A Simplified Example

## 4 DISCUSSION & OUTLOOK

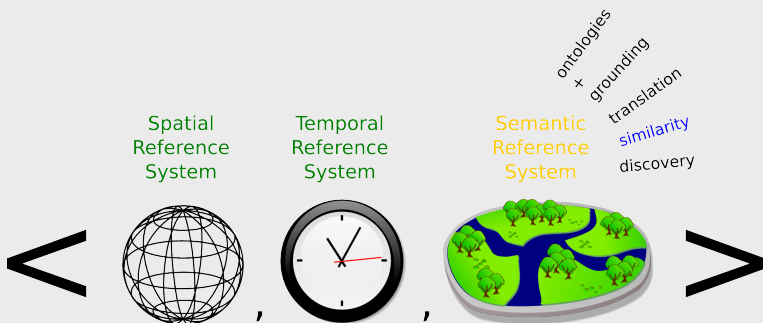
## GEOINFORMATION



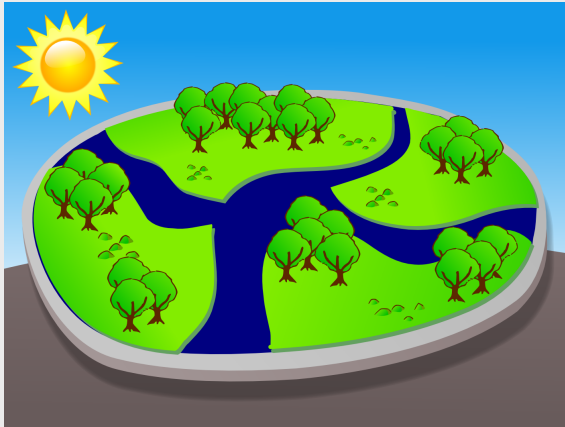
# GEOINFORMATION



## GEOINFORMATION

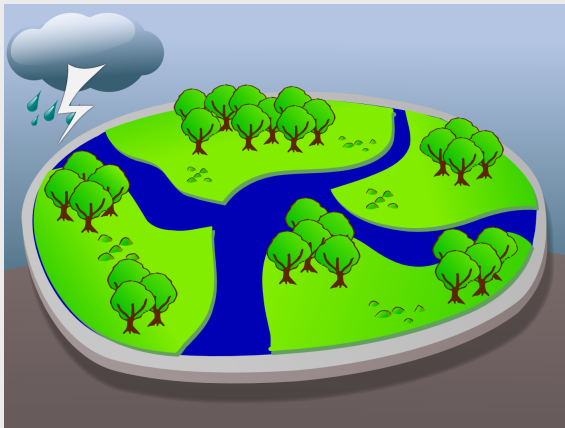


# SPATIAL DATA INFRASTRUCTURE: OGC SERVICES



- Show me Hills, Forests, Rivers, *[Feature Type]* ... at *[Loc]* → [WFS](#)
- Show me a map of this area → [WMS](#)

# SPATIAL DATA INFRASTRUCTURE: OGC SERVICES



- Water gauge? *[Stimulus Type]* ... at *[Loc][Time]* → SOS → WPS
- *[Sensor]* move to *[Loc][Time]* → SPS

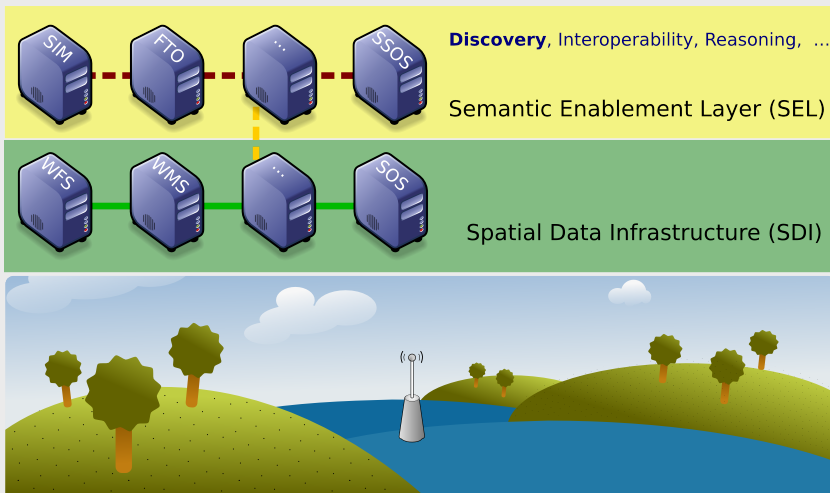
# SPATIAL DATA INFRASTRUCTURE: OGC SERVICES



■ Flooding! [Event Type] ... at [Loc][Time] → SAS → WNS



# YET ANOTHER LAYER CAKE



## 1 SEMANTICS IN GISCIENCE

- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

## 2 SIMILARITY (REASONING) IN GISCIENCE

- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

## 3 *SIM – DL<sub>A</sub>*

- Why *SIM – DL<sub>A</sub>*?
- The Similarity Tableau for *SHI*
- A Simplified Example

## 4 DISCUSSION & OUTLOOK

## ADL GAZETTEER INTERFACE

Alexandria Digital Library Gazetteer Server Client. Using ADL Server [Use ESRI Server](#)

Search Result: 19 matches (Problem Report)

#	names	Class
1	<a href="#">Winquipin Lake - Levy County - Florida - United States</a>	reservoirs
2	<a href="#">Lake Rousseau - Levy County - Florida - United States</a>	reservoirs
3	<a href="#">Meeting Pond - Alachua County - Florida - United States</a>	reservoirs
4	<a href="#">Spring Pond - Gilchrist County - Florida - United States</a>	reservoirs
5	<a href="#">Frank Taucher Reservoir - Highlands County - Florida - United States</a>	reservoirs
6	<a href="#">Little Tucker Lake - Franklin County - Florida - United States</a>	reservoirs
7	<a href="#">Venetian Pool - Dade County - Florida - United States</a>	reservoirs
8	<a href="#">Lake Manatee - Manatee County - Florida - United States</a>	reservoirs
9	<a href="#">Fox Lake - Levy County - Florida - United States</a>	reservoirs
10	<a href="#">Lake Laurence East - Dade County - Florida - United States</a>	reservoirs
11	<a href="#">Mineral Feeder Pond - Levy County - Florida - United States</a>	reservoirs
12	<a href="#">Ward Lake - Manatee County - Florida - United States</a>	reservoirs
13	<a href="#">Tucker Lake - Pinellas County - Florida - United States</a>	reservoirs

My last query was :

query request:

Footprint(within):  
 Box: ( -86.66666666666666, 23.33333333333333 -73.33333333333333, 23.33333333333333)

Class: reservoirs (from thesaurus: ADL Feature Type Thesaurus)

Zoom In Zoom Out Full Extent

Please set at least one search condition.

Location:  within map  overlays map  anywhere

Place Name:

Feature Type: ( Feature Type Thesaurus)

- levees
- offshore platforms
- piers
- reservoirs**
- waterworks
- landmarks
- launch facilities
- mine sites
- monuments
- oil fields

Place Status:  ANY  Current  Former  Proposed

Identification Code:

[help](#)

- Based on a Feature Type Thesaurus (> 1000 types)

## SIMCAT GAZETTEER INTERFACE

place name

restrict search to this area.

place type

place type suggestions

place type suggestions	supertype(s)	similar types
<a href="#">Canal</a>	<a href="#">Watercourse</a> , <a href="#">Navigable</a> , <a href="#">ManMade</a>	<a href="#">River</a> , <a href="#">Irrigation Canal</a> , <a href="#">Reservoir</a> , <a href="#">Lake</a> , <a href="#">Oasis</a>
<a href="#">Cape</a>	<i>not implemented yet</i>	<i>not implemented yet</i>
<a href="#">Capital</a>	<i>not implemented yet</i>	<i>not implemented yet</i>

place type suggestions

supertype(s)

similar types

[Canal](#)

[Watercourse](#), [Navigable](#), [ManMade](#)

[River](#), [Irrigation Canal](#), [Reservoir](#), [Lake](#), [Oasis](#)

[Cape](#)

*not implemented yet*

*not implemented yet*

[Capital](#)

*not implemented yet*

*not implemented yet*

■ Based on a Feature Type Ontology

# SIM-DL SERVER AND PLUGIN

The screenshot displays the Protégé 3.3.1 interface with the following components:

- CLASS BROWSER:** Shows a hierarchy for 'FTO\_hst' with categories like Construction, Hydrographic, Landmass, ManMade, and Navigable. Under 'Hydrographic', 'Waterbody' is expanded to show 'Canal', 'Inlet', 'IrrigationCanal', 'Lake', 'Reservoir', 'River', and 'Sea'.
- SIMILARITY REQUEST:** Shows a search for 'Canal' with options for visualization mode, similarity values, font size, and categories. The 'Number of categories' is set to 3.
- SIMILARITY RESULTS:** A table showing the results of the similarity measurement.
 

concept name	similarity (%)
River	75
IrrigationCanal	69
Reservoir	61
Lake	50
Inlet	42
Sea	33
- Reasoner log:** A window showing the performance of the similarity reasoner:
  - Synchronize reasoner
    - Time to clear knowledgebase = 0.0060 seconds
    - Time for DIG conversion = 0.022 seconds
    - Time to update reasoner = 0.026 seconds
    - Time to synchronize = 0.074 seconds
  - Computing concept similarity
    - Time to build query = less than 0.001 seconds
    - Time to perform similarity measurement = 1.37 seconds
  - Total time: 1.556 seconds
- Release Beta 2.2 (Balou); Version 2.2:** A dialog box stating 'Similarity Reasoner for the ALCQH Description Logic. Server started. Running on port 8085'.

## 1 SEMANTICS IN GISCIENCE

- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

## 2 SIMILARITY (REASONING) IN GISCIENCE

- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

## 3 *SIM – DL<sub>A</sub>*

- Why *SIM – DL<sub>A</sub>*?
- The Similarity Tableau for *SHI*
- A Simplified Example

## 4 DISCUSSION & OUTLOOK

## THE CHALLENGE

- SIM-DL is a **structural measure**; it requires the **normalization** of concept descriptions and recursively compares the (syntactic) structure of the normalized descriptions for similarity.
- Feature type ontologies require **expressive** description logics such as OWL-DL; a structural approach **fails** to support these DLs.
  
- Why not use an **ABox-based** similarity measure?
- Most feature type ontologies cannot be populated so far. Features are stored in shape-files, geo-databases, GML files, satellite images,...
  
- **Can we learn from the history of subsumption reasoning and switch to a **tableau-based** approach?**

## THE IDEA

→ Reduce the problem of **inter-concept** similarity to **inter-instance** similarity?

- How to get or compute the **instances required for comparison**?
- From the **completion tree** of a modified tableau algorithm as used for satisfiability checking. Instead of trying to find one (clash-free) model, generate a set of **proxy individuals** for comparison.
  
- How to compute similarity between proxy individuals?
- Reuse existing similarity framework, i.e., functions, contexts, similarity modes, and the alignment procedure from SIM-DL.
  - Jaccard similarity coefficient of co-occurrence
  - Network measures for role hierarchies and neighborhood
  - ...



THE SIMILARITY TABLEAU FOR *SHI*

...

**The  $\sqcup$ -rule:****Condition:**  $C_1 \sqcup C_2 \in \mathcal{L}(x)$  and  $x$  is not indirectly blocked.**Action:** Create three  $\sqcup$ -successors  $w, y, z$  of  $x$  with:

$$\mathcal{L}(w) := (\mathcal{L}(x) \setminus \{C_1 \sqcup C_2\}) \cup \{C_1\}$$

$$\mathcal{L}(y) := (\mathcal{L}(x) \setminus \{C_1 \sqcup C_2\}) \cup \{C_2\}$$

$$\mathcal{L}(z) := (\mathcal{L}(x) \setminus \{C_1 \sqcup C_2\}) \cup \{C_1, C_2\}$$

...

**The  $\forall$ -rule:****Condition:**  $(\forall R.C) \in \mathcal{L}(x)$ ,  $x$  is not indirectly blocked.**Action:**If there is an  $R$ -neighbor  $y$  of  $x$  and  $C \notin \mathcal{L}(y)$ :

$$\mathcal{L}(y) := \mathcal{L}(y) \cup \{C\}.$$

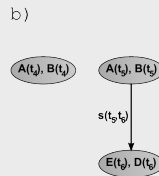
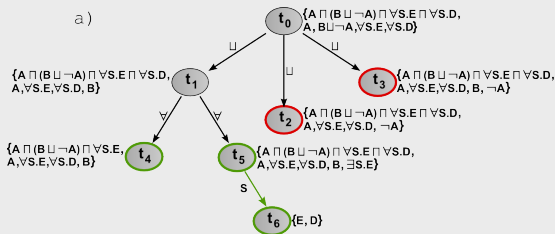
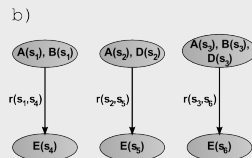
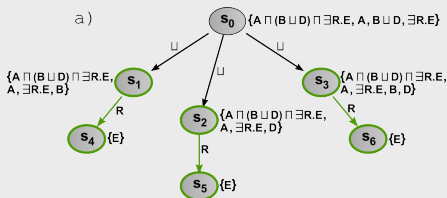
If there is no  $R$ -neighbor  $y$  of  $x$ , create two  $\forall$ -successors  $y, z$  of  $x$  with:

$$\mathcal{L}(y) := \mathcal{L}(x) \text{ (} y \text{ will then be blocked)}$$

$$\mathcal{L}(z) := \mathcal{L}(x) \cup \{\exists R.C\}.$$

...

 **$\geq$  and  $\leq$ -rules ...**

SIM – DL<sub>A</sub>: A SIMPLIFIED EXAMPLE

# SIM – DL<sub>A</sub>: ALIGNMENT MATRIX (FOR MAX. SIMILARITY)

a)

Proxy Models of  $C_s$  and  $C_t$ 

$C_s \backslash C_t$	$s_1$	$s_2$	$s_3$
$t_4$	0.67	0.5	0.67
$t_5$	0.83	0.67	0.63

b)

Assertions of  $s_1$  and  $t_5$ 

$s_1 \backslash t_5$	A	B	$r(s_1, s_4)$
A	1	0.5	-
B	0.5	1	-
$s(t_5, t_6)$	-	-	0.5

c)

Assertions of  $s_4$  and  $t_6$ 

$s_4 \backslash t_6$	E
E	1
D	0

- a) selecting **individuals** for comparison
- b) selecting **assertions** for comparison
- c) ...

## 1 SEMANTICS IN GISCIENCE

- Geoinformation
- Spatial Data Infrastructure
- Semantic Enablement Layer

## 2 SIMILARITY (REASONING) IN GISCIENCE

- ADL Gazetteer Interface
- SimCat Gazetteer Interface
- SIM-DL Server and Plugin

## 3 SIM – DL<sub>A</sub>

- Why SIM – DL<sub>A</sub>?
- The Similarity Tableau for *SHI*
- A Simplified Example

## 4 DISCUSSION & OUTLOOK

## DISCUSSION & OUTLOOK

- Support more **expressive DL** (e.g., cardinality restrictions)
- **Integration** into existing reasoners (e.g. FaCT, Pellet)
- **Approximation** and optimization (thresholds → roles)
- Evaluation of *SIM – DL<sub>A</sub>* (Human Participants Test)
  
- Conceptualization of and reasoning about **events**
- Concepts change over time (**ecology** of concepts)

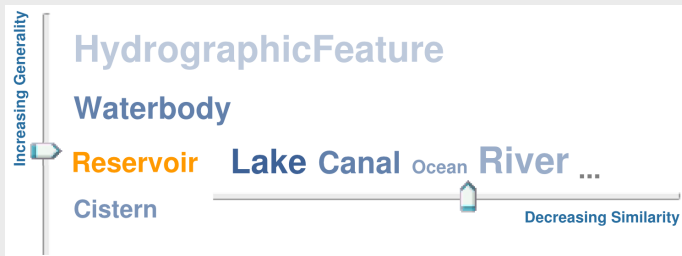
## QUESTIONS?

# Thank you!

- SimCat Project Website: [sim-dl.sourceforge.net](http://sim-dl.sourceforge.net)
- Münster Semantic Interoperability Lab: [musil.uni-muenster.de](http://musil.uni-muenster.de)
- International Conference on Geospatial Semantics: [www.geosco.org](http://www.geosco.org)
- Semantic Enablement Community at 52° North: [52north.org/semantics](http://52north.org/semantics)

The SIM-DL server, plug-in and user interfaces are free and open source software and can be downloaded at the SimCat project website.

## NAVIGATION



## QUERY-BY-EXAMPLE

Add reference feature



Selected reference features

Esnagami River - Ontario, CDN

Rideau Canal - Ontario, CDN

Marmion Lake - Ontario, CDN

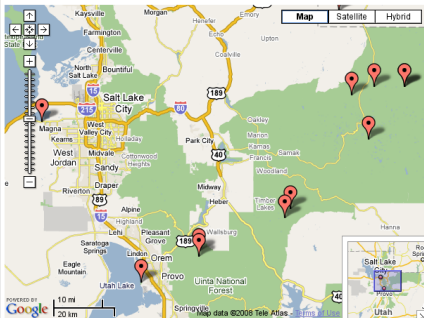
1

similar types

Canal, River, IrrigationCanal, Lake, Reservoir, Tan

2

Search within this region



14 features found

Lake Fork River - Utah, US

3

Right Hand Fork East Bear River - Utah, US

Little South Fork Provo River - Utah, US

Duchesne River - Utah, US

East Fork Bear River - Utah, US

Left Hand Fork East Fork Bear River - Utah, US

Provo River - Utah, US

Weber River - Utah, US

East Fork Duchesne River - Utah, US

West Fork Duchesne River - Utah, US

&lt;&lt;back 1-10 11-14 next&gt;&gt;



# SIMILARITY REASONING

- 1 Definition of application area and intended audience
- 2 Selection of search (query) and target concepts
- 3 Transformation of concepts to canonical form
- 4 Definition of an alignment matrix for concept descriptors
- 5 Application of constructor specific similarity functions
- 6 Determination of standardized overall similarity
- 7 Interpretation of the resulting similarity value(s)

# COMPLEXITY?

## ■ EXPTIME-complete, but

*'[...]This shows that, at the beginning of the new millennium, even an EXPTIME-algorithm is no longer automatically considered to be impractical in the DL community.'*

(Baader & Sattler; An Overview of Tableau Algorithms for Description Logics (2000))

→ Optimization and Approximation

## THE ADL GAZETTEER WEB INTERFACE &amp; THESAURUS

Alexandria Digital Library Gazetteer Server Client. Using ADL Server [Use ESRI Server](#)

Search Result: 19 matches (Problem Report)


#	names	Class
1	Winquipin Lake - Levy County - Florida - United States	reservoirs
2	Lake Rousseau - Levy County - Florida - United States	reservoirs
3	Meeting Pond - Alachua County - Florida - United States	reservoirs
4	Spring Pond - Gilchrist County - Florida - United States	reservoirs
5	Frank Taucher Reservoir - Highlands County - Florida - United States	reservoirs
6	Little Tucker Lake - Franklin County - Florida - United States	reservoirs
7	Venetian Pool - Dade County - Florida - United States	reservoirs
8	Lake Manatee - Manatee County - Florida - United States	reservoirs
9	Fox Lake - Levy County - Florida - United States	reservoirs
10	Lake Laurence East - Dade County - Florida - United States	reservoirs
11	Mineral Feeder Pond - Levy County - Florida - United States	reservoirs
12	Ward Lake - Manatee County - Florida - United States	reservoirs
13	Taylor Lake - Dinlark County - Florida - United States	reservoirs

My last query was :

query request:

Footprint(within):  
Box: ( -86.66666666666666, 23.333333333333332 -73.33333333333333, ...

Class: reservoirs (from thesaurus: ADL Feature Type Thesaurus)



Please set at least one search condition.

Location:  within map  overlaps map  anywhere

Place Name: has all words

Feature Type: (Feature Type Thesaurus)

- levees
- offshore platforms
- piers
- reservoirs
- waterworks
- landmarks
- launch facilities
- mine sites
- monuments
- oil fields

Place Status:  ANY  Current  Former  Proposed

Identification Code:

help

# THE ADL GAZETTEER WEB INTERFACE & THESAURUS

Search USGS Search Google Main Index TopTerm Index Term Index

Prev Term: [reserves](#)  
Next Term: [residential sites](#)

## reservoirs

**Used for:**

[covered reservoirs](#)  
[intermittent reservoirs](#)

**Broader Terms:**

[hydrographic structures](#)

**Related Terms:**

[lakes](#)  
[waterworks](#)

**Definition:**

Artificially impounded bodies of water. [USGS Feature Class Definitions ]

---

Linda Hill [Comments/questions about the Feature Type Thesaurus](#)

---

## THE ADL GAZETTEER WEB INTERFACE &amp; THESAURUS

Search USGS Search Google Main Index TopTerm Index Term Index

Prev Term: [reserves](#)  
Next Term: [residential sites](#)

## reservoirs

**Used for:**  
[covered reservoirs](#)  
[intermittent reservoirs](#)

**Broader Terms:**  
[hydrographic structures](#)

**Related Terms:**  
[lakes](#)  
[waterworks](#)

**Definition:**  
Artificially impounded bodies of water. [USGS Feature Class Definitions ]

---

Linda Hill [Comments/questions about the Feature Type Thesaurus](#)

```
graph TD
    HF((hydrographic features)) --- L((lakes))
    MF((manmade features)) --- HS((hydrographic structures))
    HS --- R((reservoirs))
```

# THE ADL GAZETTEER WEB INTERFACE & THESAURUS

[Search USGS](#) [Search Google](#) [Main Index](#) [TopTerm Index](#) [Term Index](#)

Prev Term: [offshore areas](#)  
Next Term: [oil camps](#)

## offshore platforms

**Used for:**

[artificial islands](#)  
[oil platforms](#)  
[platforms \(offshore\)](#)

**Broader Terms:**

[hydrographic structures](#)

**Related Terms:**

[oil fields](#)

**Definition:**

Structures erected in a sea primarily, but not exclusively, for the extraction of petroleum products. [Adapted from USGS Circ 1048]

---

Linda Hill [Comments/questions about the Feature Type Thesaurus](#)

# THE ADL GAZETTEER WEB INTERFACE & THESAURUS

## hydrographic structures

### Used for:

[boat landings](#)  
[boat launches](#)  
[boat ramps](#)  
[sluices](#)  
[water mills](#)

### Broader Terms:

[manmade features](#)

### Narrower Terms:

[breakwaters](#)  
[canals](#)  
[dam sites](#)  
[gaging stations](#)  
[harbors](#)  
[levees](#)  
[offshore platforms](#)  
[piers](#)  
[reservoirs](#)  
[waterworks](#)

### Related Terms:

[hydrographic features](#)  
[seaplane bases](#)  
[transportation features](#)  
[wells](#)

### Scope Note:

For constructed bodies of water. For natural water bodies, use 'hydrographic features'.

# THE ADL GAZETTEER WEB INTERFACE & THESAURUS

## hydrographic structures

### Used for:

[boat landings](#)  
[boat launches](#)  
[boat ramps](#)  
[sluices](#)  
[water mills](#)

### Broader Terms:

[manmade features](#)

### Narrower Terms:

[breakwaters](#)  
[canals](#)  
[dam sites](#)  
[gaging stations](#)  
[harbors](#)  
[levees](#)  
[offshore platforms](#)  
[piers](#)  
[reservoirs](#)  
[waterworks](#)

### Related Terms:

[hydrographic features](#)  
[seaplane bases](#)  
[transportation features](#)  
[wells](#)

### Scope Note:

For constructed bodies of water. For natural water bodies, use 'hydrographic features'.

