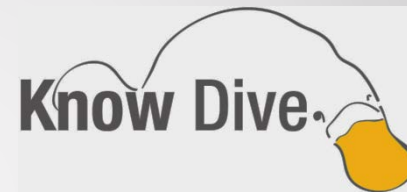# A Faceted Ontology for a Semantic Geo-catalogue

**F. Farazi**, *University of Trento*
V. Maltese, *University of Trento*
F. Giunchiglia, *University of Trento*
A. Ivanyukovich, *Trient Consulting Group*

Feroz Farazi – University of Trento

# Outline

❑ Introduction

❑ Semantic Geo-catalogue

❑ Dataset Processing

❑ Ontology Population

❑ Integration with GeoWordNet

❑ Conclusion

Feroz Farazi – University of Trento

# Introduction

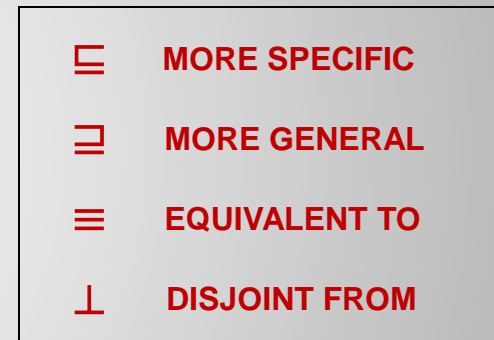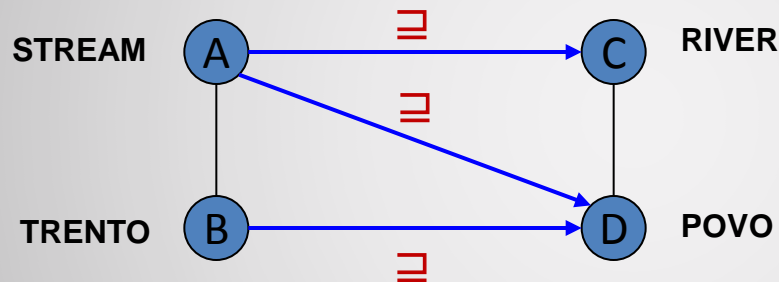## SDI in Trentino

Feroz Farazi – University of Trento
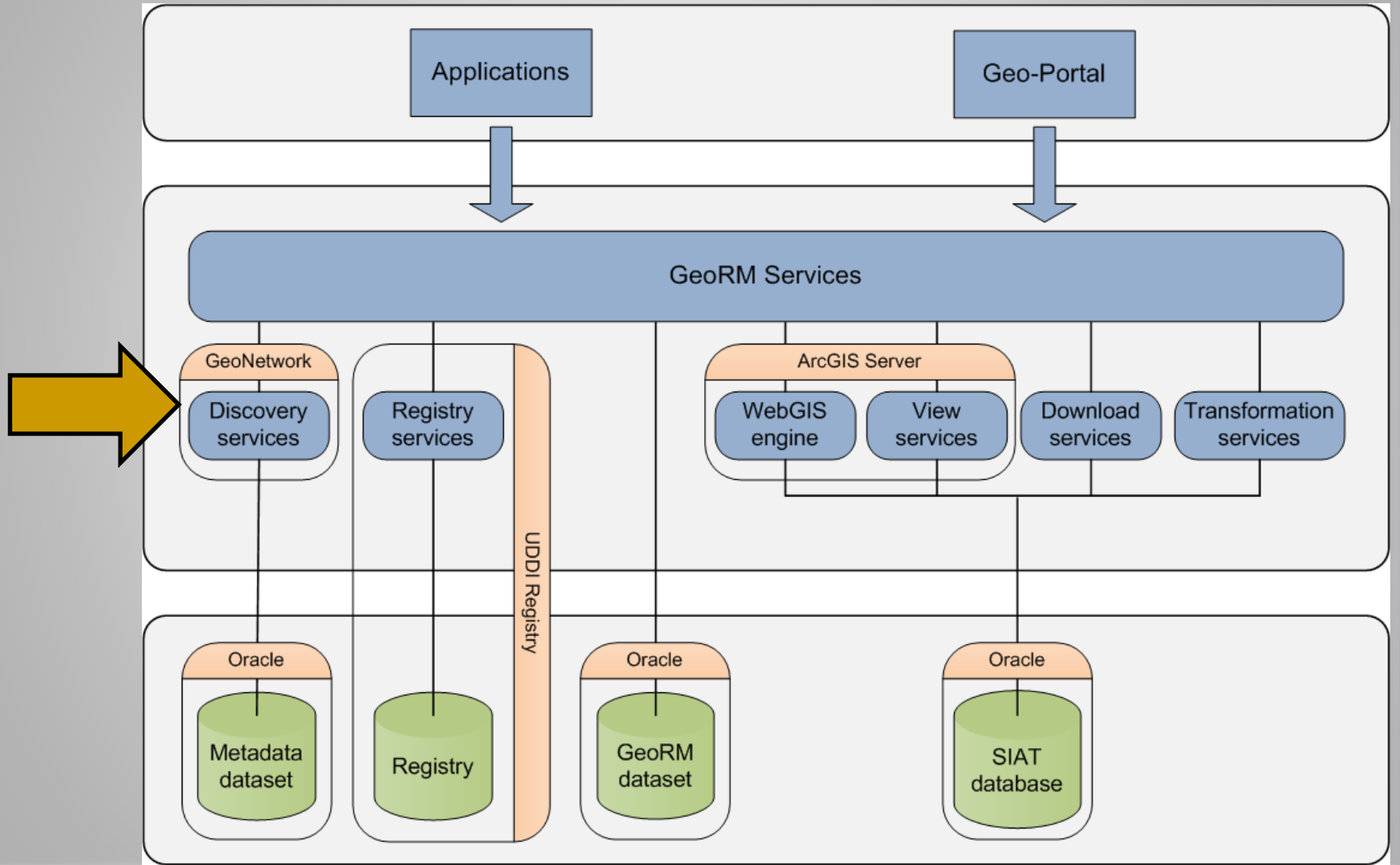
# Introduction

❑ Problems with current tools for search (e.g. keyword search)

   ❑ Different terminology used ⇨ low recall

   ❑ Poor expressivity ⇨ low recall

❑ As a solution we propose intelligent query expansion to improve user experience

   ❑ Based on Semantic Matching techniques: the S-Match system

Feroz Farazi – University of Trento

# Introduction

## Semantic matching

❑ Semantic matching is a technique for identifying semantic correspondences between ontologies

❑ A set of matches is called alignment

STREAM **A** ⊒ **C** RIVER

⊒

**B** TRENTO **D** POVO

⊒

| ⊑ | MORE SPECIFIC |
|---|---|
| ⊒ | MORE GENERAL |
| ≡ | EQUIVALENT TO |
| ⊥ | DISJOINT FROM |

❑ A tool for semantic matching called S-Match developed at the University of Trento

❑ S-Match is an open source tool, available at: http://semanticmatching.org/

Feroz Farazi – University of Trento

# Semantic Geo-catalogue (SGC)

## Architecture

Feroz Farazi – University of Trento

# SGC
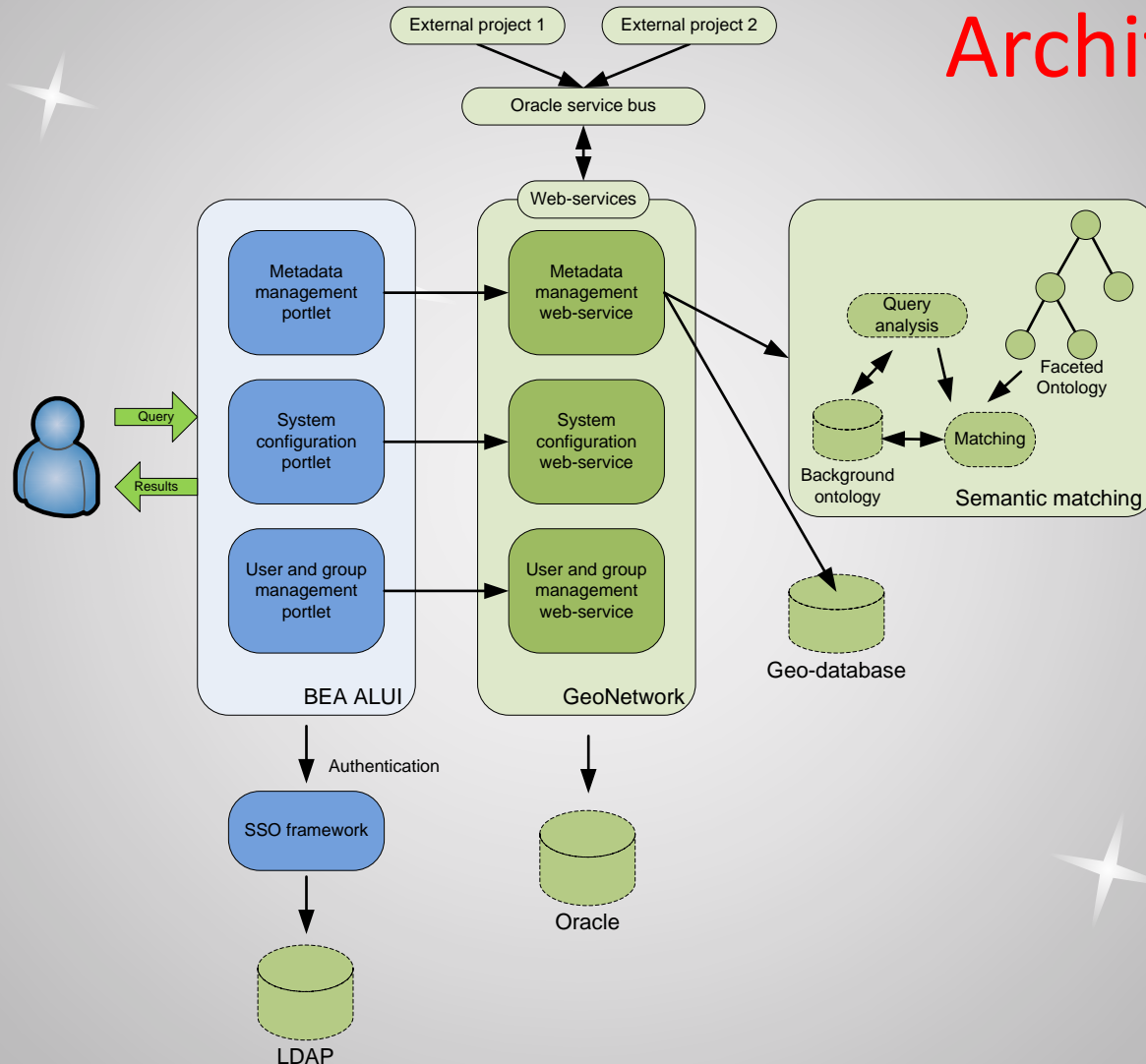
## S-Match usage example

| Query | Query expansion with S-Match |
|---|---|
| watercourse | Rivulet, Stream, River |
| falls | Cascade, Waterfall |
| water | Rivulet, Waterfall, Cascade, River, Body of water, Stream, Spring, Canal, Group of lakes, Lake |
| elevation | Natural elevation, Mountain, Highland, Glacier, Mountain range, Peak, Hill |
| installation | Milestone, Hut, Farm, Highway, Railway, Road, Street, Transportation system, Provincial Road, Facility, Shelter |
| ice mass | Glacier |
| district | Administrative division, Province, Municipality, Ward, Populated place |
| transportation facility | Transportation system, Road, Street, Provincial Road, Milestone, Railway, Highway |
| reef | |

Feroz Farazi – University of Trento

# SGC

## Query expansion

**Query**

watercourse

**Faceted Ontology**

stream

river   rivulet

S-Match

**Expanded Query**

watercourse   stream
river   rivulet

Geo
WordNet

Feroz Farazi – University of Trento

# Dataset Processing

A global view of the dataset processing

PHASE1: ANALYSIS

| Analysis of the dataset |

PHASE2: DATA EXTRACTION

| Extraction of the macro-classes | Extraction of the locations |

PHASE4: BUILDING THE ONTOLOGY

| Define classes and semantic relations between them |

PHASE3: FILTERING

| Provide missing data and remove duplicates | Bilingual issues: double names |

Feroz Farazi – University of Trento

# Dataset Processing

Analysis → Extraction → Building Ontology
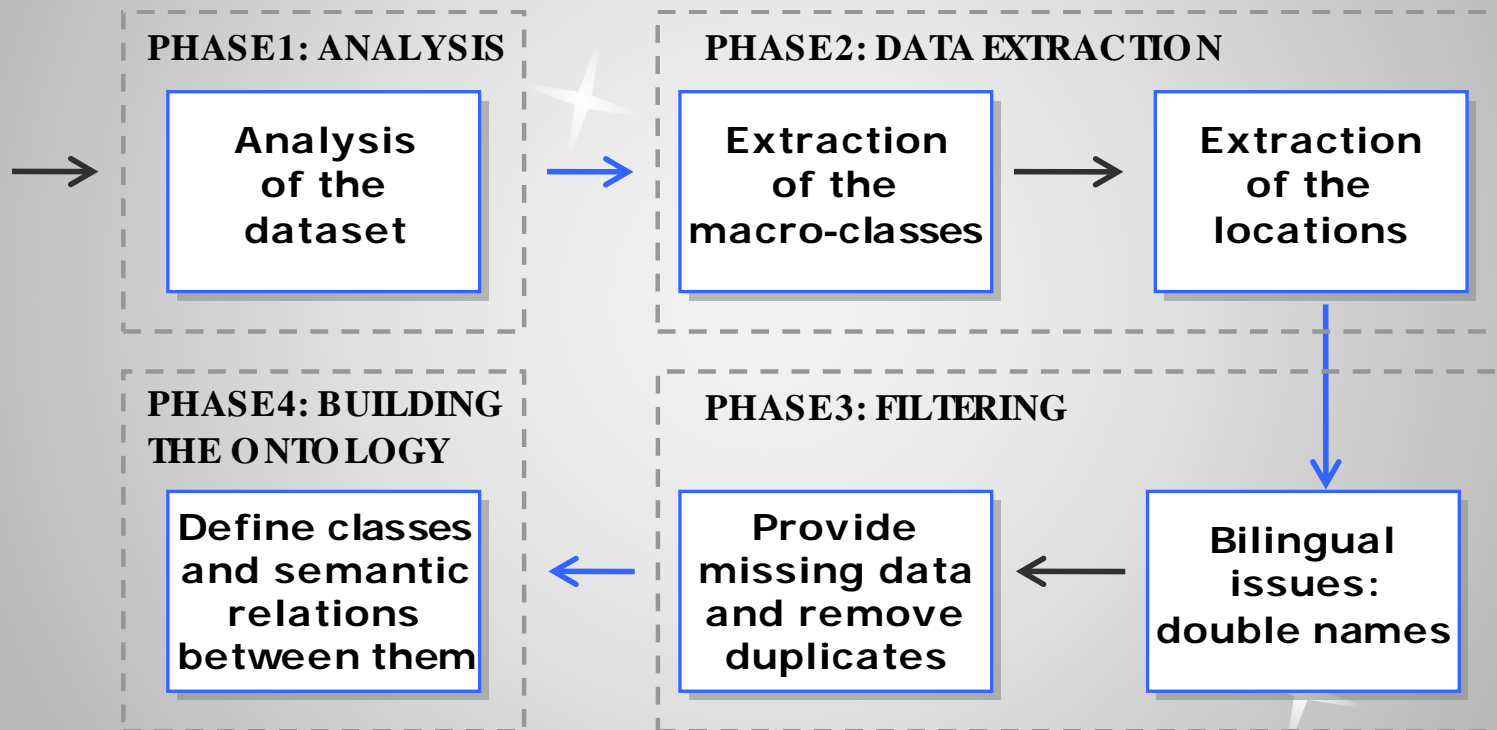
❑ Feature Classes:

| In Italian | Corsi dacqua/laghi |
|------------|-------------------|
| In English | Watercourses/lakes |

❑ These are too broad and not homogeneous classes of entities

❑ Therefore, need refinement: from macro to micro classes

Feroz Farazi – University of Trento

# Dataset Processing

Analysis → Extraction → Building Ontology

Province ← Instance-of ─ Autonomous Province of Trento

part-of ↑        part-of ↑

Municipality ← Instance-of ─ Municipality of Povo

part-of ↑        part-of ↑

Ward and Populated Place ← Instance-of ─ Povo

part-of ↑        part-of ↑

(Other) Location ← Instance-of ─ Gabbiolo

Feroz Farazi – University of Trento

# Dataset Processing

Analysis ⟶ Extraction ⟶ Building Ontology
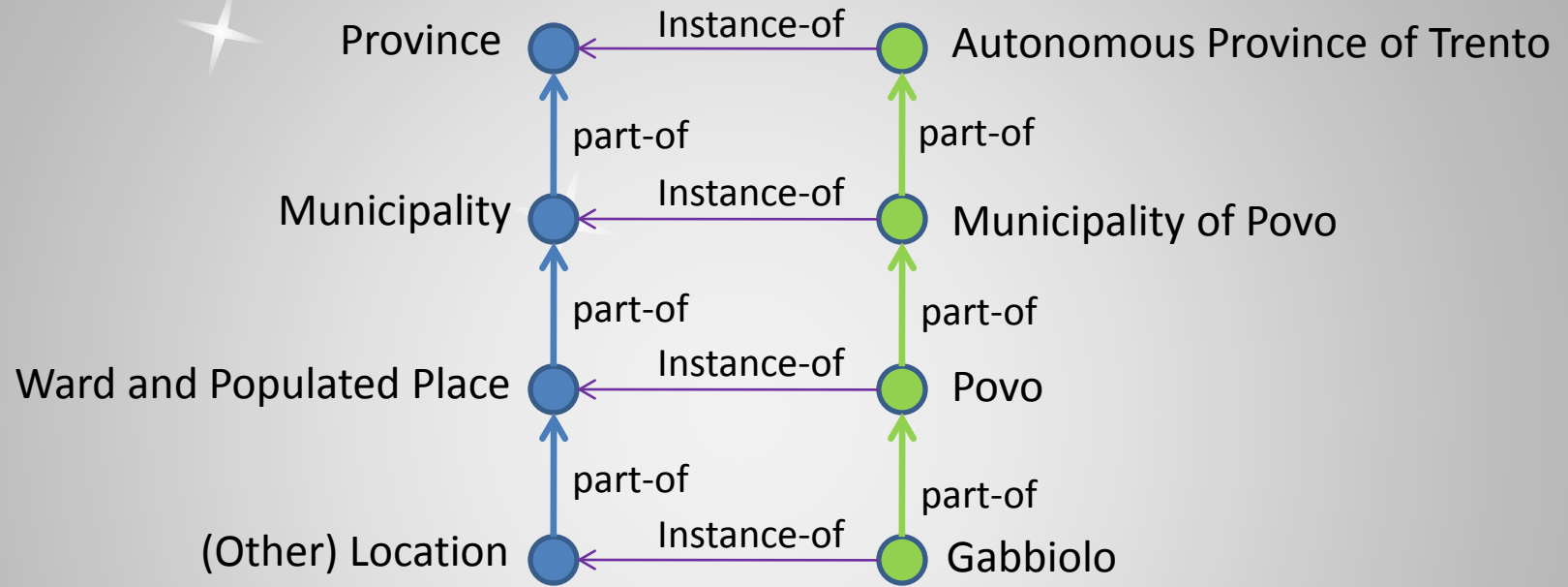
❑ Objective: Create an ontology that reflects the specificity of Trentino and respects the canons of the analytico-synthetic approach for the generation of a faceted lightweight ontology

❑ Analytico-synthetic approach:

  ❑ Step 1 (Analysis): Disambiguation of the classes

  ❑ Step 2 (Synthesis): Hierarchical organization: semantic relations

Feroz Farazi – University of Trento

# Dataset Processing

Analysis → Extraction → Building Ontology

## ❑ Step 1: Disambiguation result mapping to atomic concepts

| Feature Class | Atomic Concept |
|---|---|
| P510 Antichita importanti<br>P520 Antichita di importanza minore | Antiquity |
| P210 Corsi dacqua/laghi (1ord.)<br>P220 Corsi dacqua/laghi (2 ord.)<br>P230 Corsi dacqua/Canali/Fosse/Cond. forz./Laghi (3 ord.)<br>P240 Corsi dacqua/Canali/Fosse/Cond. forz./Laghi (>3 ord. 25.000)<br>P241 Corsi dacqua/Canali/Fosse/Cond. forz./Laghi (>3 ord.) | Lake<br>Group of lakes<br>Stream<br>River<br>Rivulet<br>Canal |

Feroz Farazi – University of Trento

# Dataset Processing

→ | Analysis | → | Extraction | → | Building Ontology |

☐ Step 2: Hierarchical organization of atomic concepts into facets, e.g., body of water

**Body of water (Idrografia)**

Lake (Lago)

Group of lakes (Gruppo di laghi)

Stream (Corso dacqua)

River (Fiume)

Rivulet (Torrente)

Canal (Canale)

Feroz Farazi – University of Trento

# Ontology Population

❑ Each location is connected to a feature class

❑ We identify a suitable atomic concept

❑ Example heuristics: applied to natural Elevation (NE), block of classes P110 – P142

| # | Heuristic used | Class name in English | Class name in Italian |
|---|---|---|---|
| 1 | If the name starts with "Cima", we map to Peak | Peak | Cima |
| 2 | If the name starts with "Monte", we map to Mountain in the cases where the entity is in P110-115-120, we map to NE otherwise | Mountain | Montagna, Monte |
| 3 | If the name contains "Passo", "Pas" or "Forcella" we map to Pass | Pass | Passo |

Feroz Farazi – University of Trento

# Integration with GeoWordNet

❑ Concept Integration

    ❑ Facet concept identification

    ❑ Concept identification

    ❑ Parent identification

❑ Entity matching for entity integration

    ❑ A set of rules were used

| Name | Class | Coordinate | Parent |
|------|-------|-----------|--------|
| 1380 | 1167 | 2(exact match) | 2 |
|      |       | 11(using the offset +/-0.0001) | 11 |
|      |       | 354 (using the offset +/-0.001) | 244 |
|      |       | 727 (using the offset +/-0.01) | 244 |

Feroz Farazi – University of Trento

# Conclusion

❑ A Geo-catalogue is extended with semantic search capability through the use of the S-Match semantic matching tool and a domain specific faceted lightweight ontology

❑ We presented how we built and populated the faceted lightweight ontology and its use in the Semantic Geo-catalogue

❑ We briefly described how the faceted ontology was integrated into the GeoWordNet, which is used as background knowledge of the S-Match

Feroz Farazi – University of Trento

# Thank you!
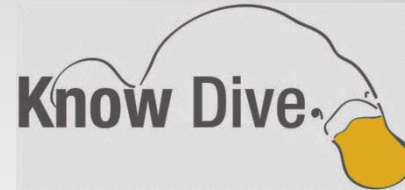


If you have some question?

❑ The S-Match open source tool is available at:
http://semanticmatching.org/

❑ The GeoWordNet is available at:
http://geowordnet.semanticmatching.org/

Feroz Farazi – University of Trento