



STEINBEIS-STIFTUNG



**Modelling and Implementation of
WebService based Vocabularies**

Walter Koch

**Technical University of Graz &
Steinbeis Innovation Transfercentre for
Information Management and
Cultural Heritage Informatics (IMCHI)**

Wir machen Innovationen schneller



Steinbeis Foundation

☒ **Service Areas**

Consultancy, Research & Development, Expert Assessments,
International Technology Transfer, Education, Steinbeis University Berlin

☒ **Network**

765 Steinbeis enterprises formed in 2008 the „**Steinbeis Network**“

☒ **Turnover 2008**

124 Million Euro

☒ **Employees (2008)**

1.383 Staff Members, **3.338** Freelancers, **801** Professors

☒ **Headquarter**

Stuttgart, Germany



“Perspectives on Metadata“, Vienna, 2009-11-13

“Modelling and Implementation of WebService Based Vocabularies“

Content:

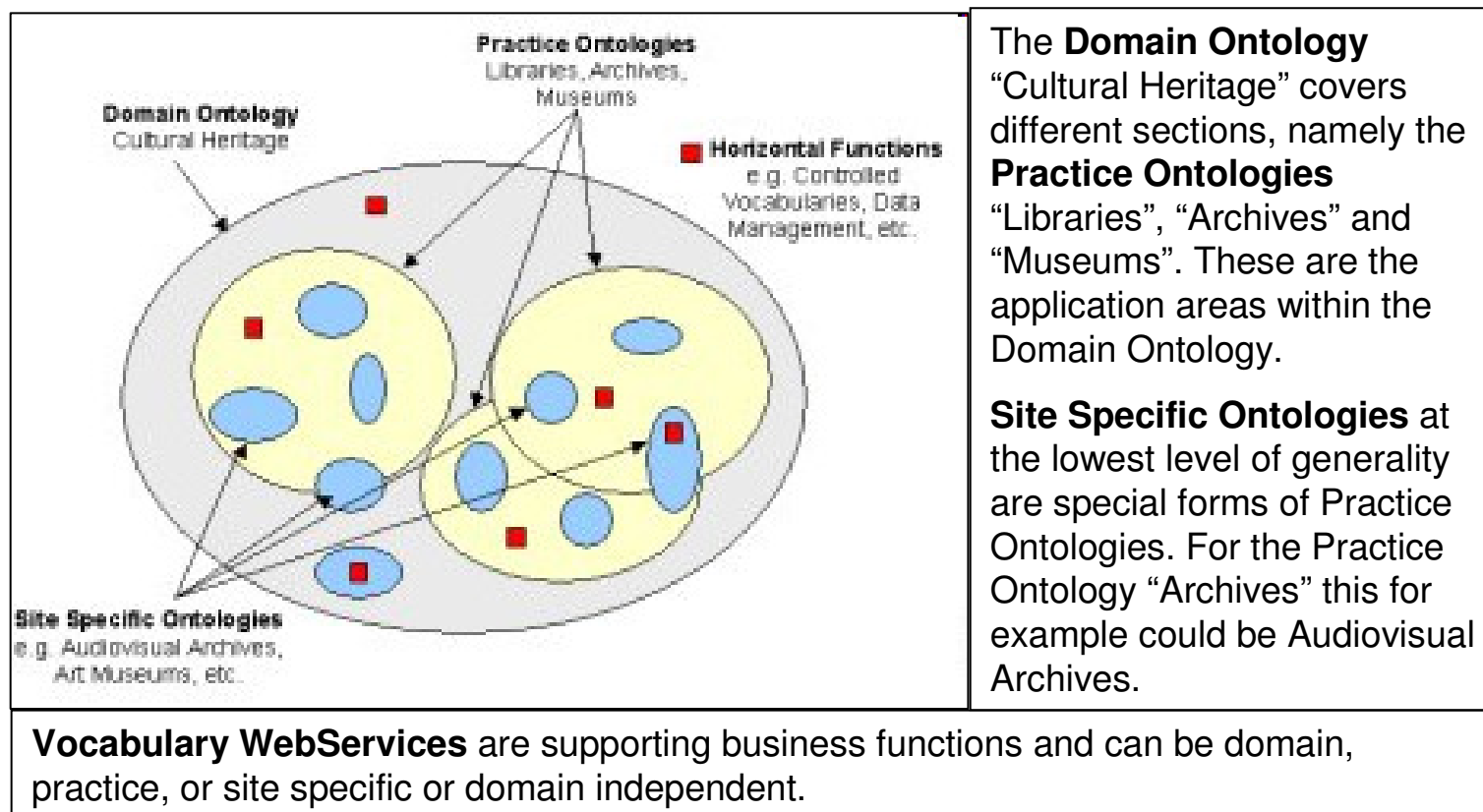
- ☒ Methods and Technologies
- ☒ Implementation Issues
- ☒ Workflows
- ☒ Business Model
- ☒ Examples

SOOM¹⁾ - Reference Implementation: The Vocserv²⁾-System

- 1) Service Oriented Object Management:
<http://scrum.ait.co.at/projects/soom/>
- 2) Vocabulary Services:
<http://scrum.ait.co.at/projects/soom/wiki/Vocabulary-Issues>

☒ Ontology Engineering: IDEF5 (Integrated Definition 5)

<http://www.idef.com/IDEF5.html>



☒ Construction of Controlled Vocabularies

B.1. Properties of Lists, Synonym rings, Taxonomies, and Thesauri

Table 3: Controlled vocabulary properties

Property	List	Synonym Ring	Taxonomy	Thesaurus
Types of Terms				
Preferred terms	Yes	No	Yes	Yes
Entry terms	No	Yes	No	Yes
Candidate terms	No	No	No	Optional
Provisional terms	No	No	No	Optional
Deleted terms	No	No	No	Optional
Relationships				
Equivalence		Yes	No	Yes
Hierarchy		No	Yes	Yes
Part/Whole		No	Yes	Yes
IsA		No	Yes	Yes
HasA		No	Yes	Yes
Classification		No	Optional	Optional
Related terms		No	No	Yes
Facet		No	No	Optional
Notes				
Scope note	No	No	No	Optional
History note			No	Optional
Other notes			No	Optional

ANSI/NISO Z39.19-2005

Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

Vocabulary independent
(metadata)

Vocabulary independent
(structural elements)

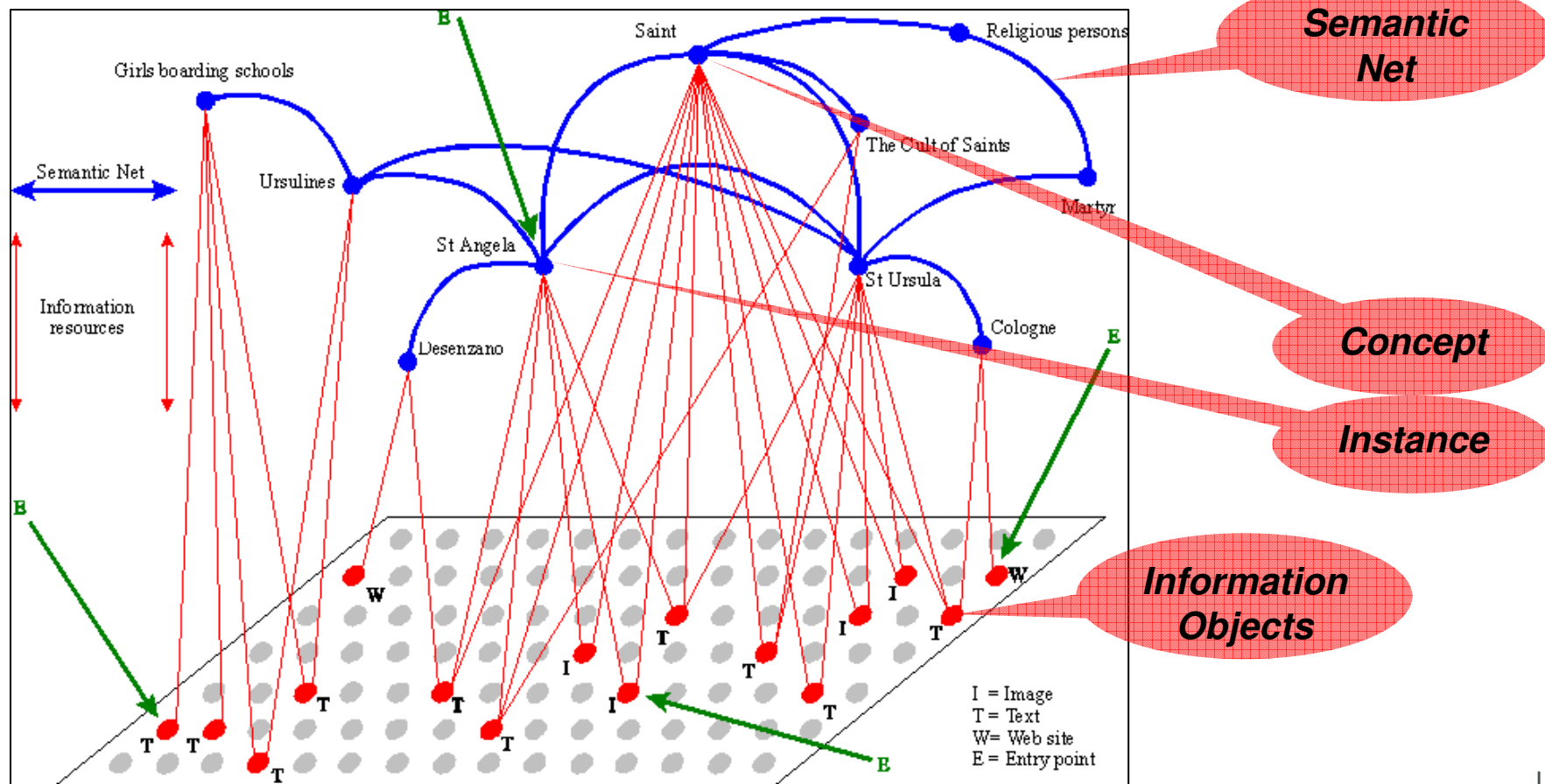
Vocabulary dependent
(metadata)



Methods and Technologies

☒ Semantic Technologies

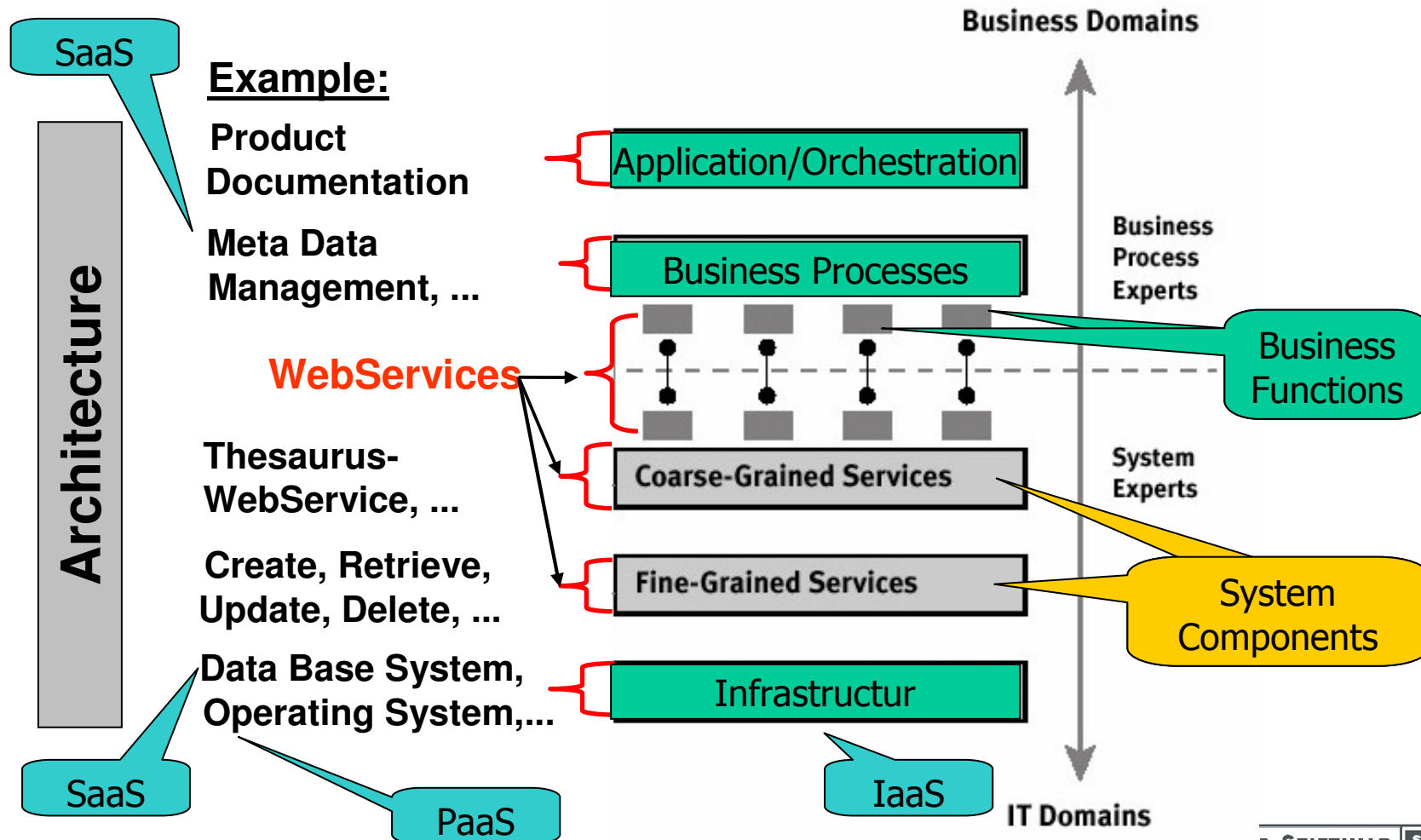
XTM-Topicmap (ISO/IEC 13250:2000), **SKOS** Simple Knowledge Organisation System





Implementation Issues

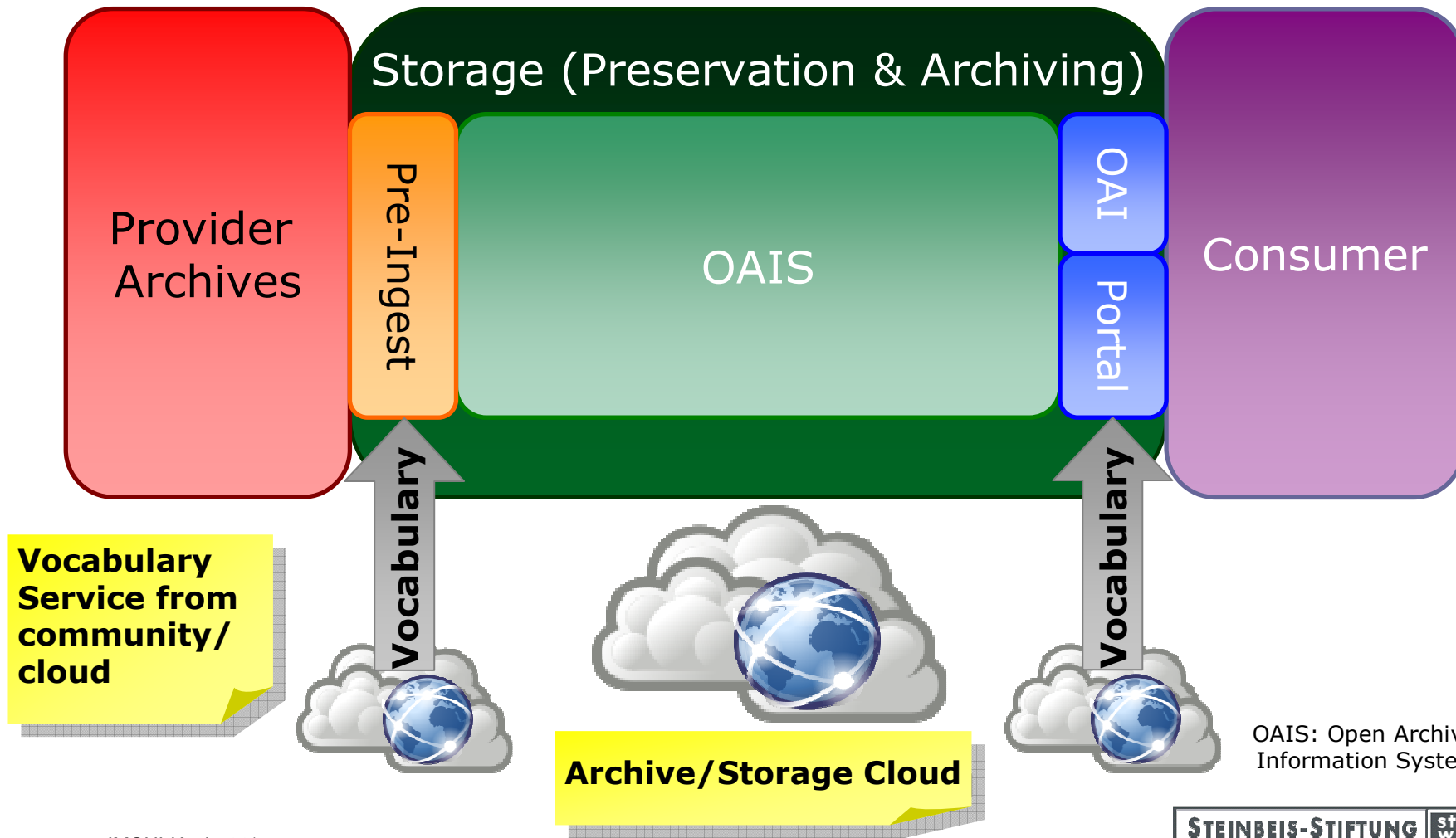
☒ Service Oriented Architecture (SOA) - WebServices





Implementation Issues

☒ Cloud Computing/Storage



OAI: Open Archival Information System



Implementation Issues

☒ Data Structure

Follows ANSI/NISO Z39.19-2005 Standard with extensions to provide multilinguality

☒ Structur of a Node

<Thesaurus>	(1)	Name of Controlled Vocabulary
ID	(1)	Unique Identifier (Numeric)
LANG	(1)	Language
TRM	(1)	Name of Term
BT	(0-n)	Broader Term (Parent Node)
NT	(0-n)	Narrower Term (Child Node)
USE	(0-n)	Use
UF	(0-n)	Use for
RT	(0-n)	Related Term
GS	(1)	Generic Structure (Path)
GS_LEVEL	(1)	Depth within GS (Sub Path)
NOTATION	(0-1)	Alternate Unique ID (Text/Label)
SCOPE	(0-1)	Container for Meta Data (vocabulary dependent)
<ATTR.>	(1-n)	Metadata Element



Implementation Issues

⊞ AIT Vocabulary Service Description V1 (API)

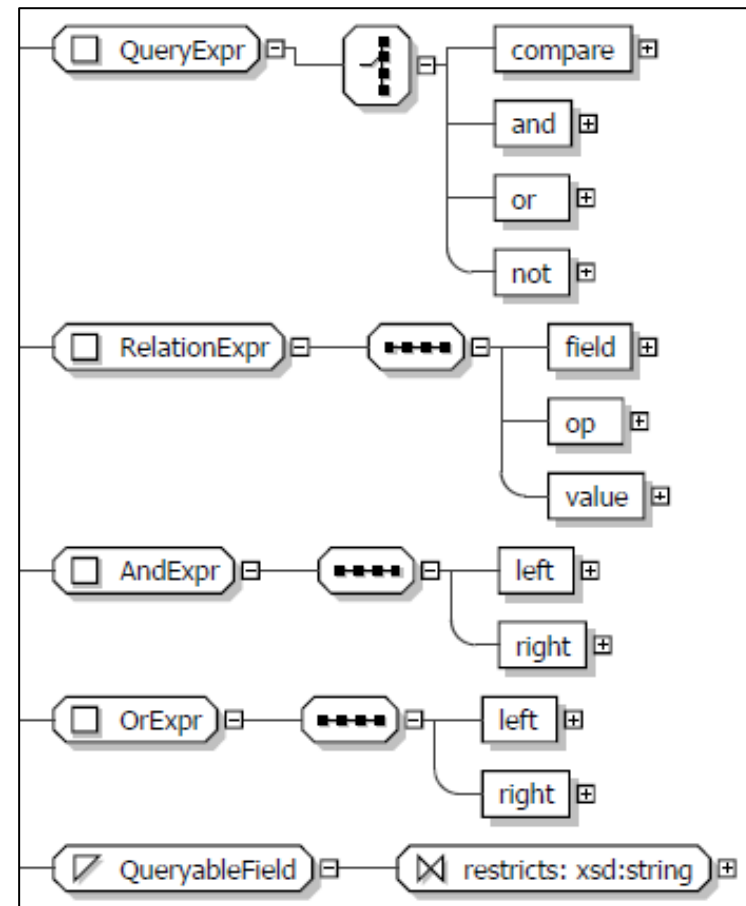
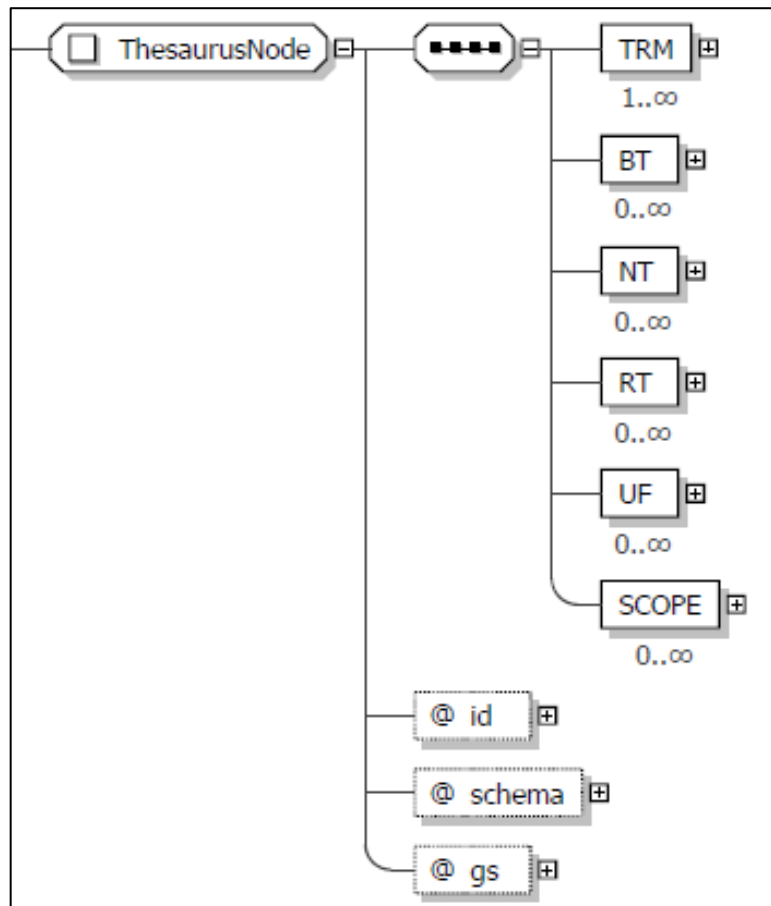
[http://demo.ait.co.at/thesaurus/AIT Vocabulary Service Description V1.pdf](http://demo.ait.co.at/thesaurus/AIT_Vocabulary_Service_Description_V1.pdf)

<http://demo.ait.co.at/thesaurus/Thesauri.wSDL>

wsdl:message		xsd:element		
	@name		@name	xsd:complexType
(30 rows)		(14 r...		
	1 getAllThesauriRequest		1 getLanguages	> xsd:complexType
	2 getAllThesauriResponse		2 getLanguagesResponse	> xsd:complexType
	3 getLanguagesRequest		3 query	> xsd:complexType
	4 getLanguagesResponse		4 queryResponse	> xsd:complexType
	5 queryRequest		5 getNodes	> xsd:complexType
	6 queryResponse		6 getMetaAttribute	> xsd:complexType
	7 getNodesRequest		7 getMetaAttributeResponse	> xsd:complexType
	8 getNodesResponse		8 login	> xsd:complexType
	9 getMetaAttributeRequest		9 loginResponse	> xsd:complexType
	10 getMetaAttributeResponse		10 logout	> xsd:complexType
	11 loginRequest		11 logoutResponse	> xsd:complexType
	12 loginResponse		12 thesaurusFault	> xsd:complexType
	13 logoutRequest		13 query_fault1	> xsd:complexType
	14 logoutResponse		14 getLanguages_fault	> xsd:complexType
	15 createNodeRequest			

Implementation Issues

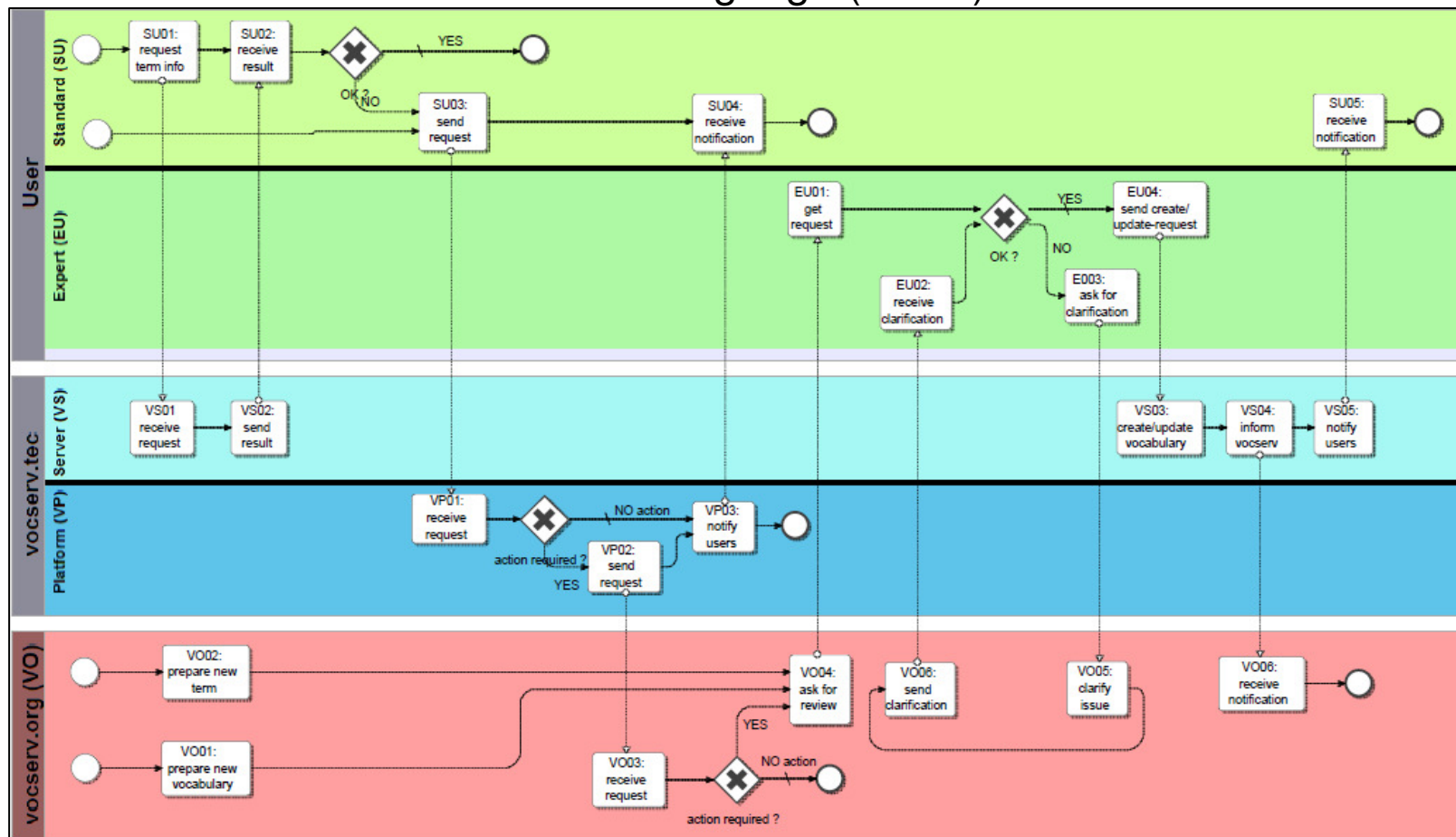
⊞ AIT Vocabulary Service - Schema





Workflows

- Business Process Modelling Notation (BPMN)
- Business Process Execution Language (BPEL)





Business Model

- ☒ As Webservice (Software as a Service): Pay Per View (ePoints):

costs	duration	0.248
	bandwidth	234838
	hits	100
	license	0
	e_points	97.22
	e_points_left	8630661.47

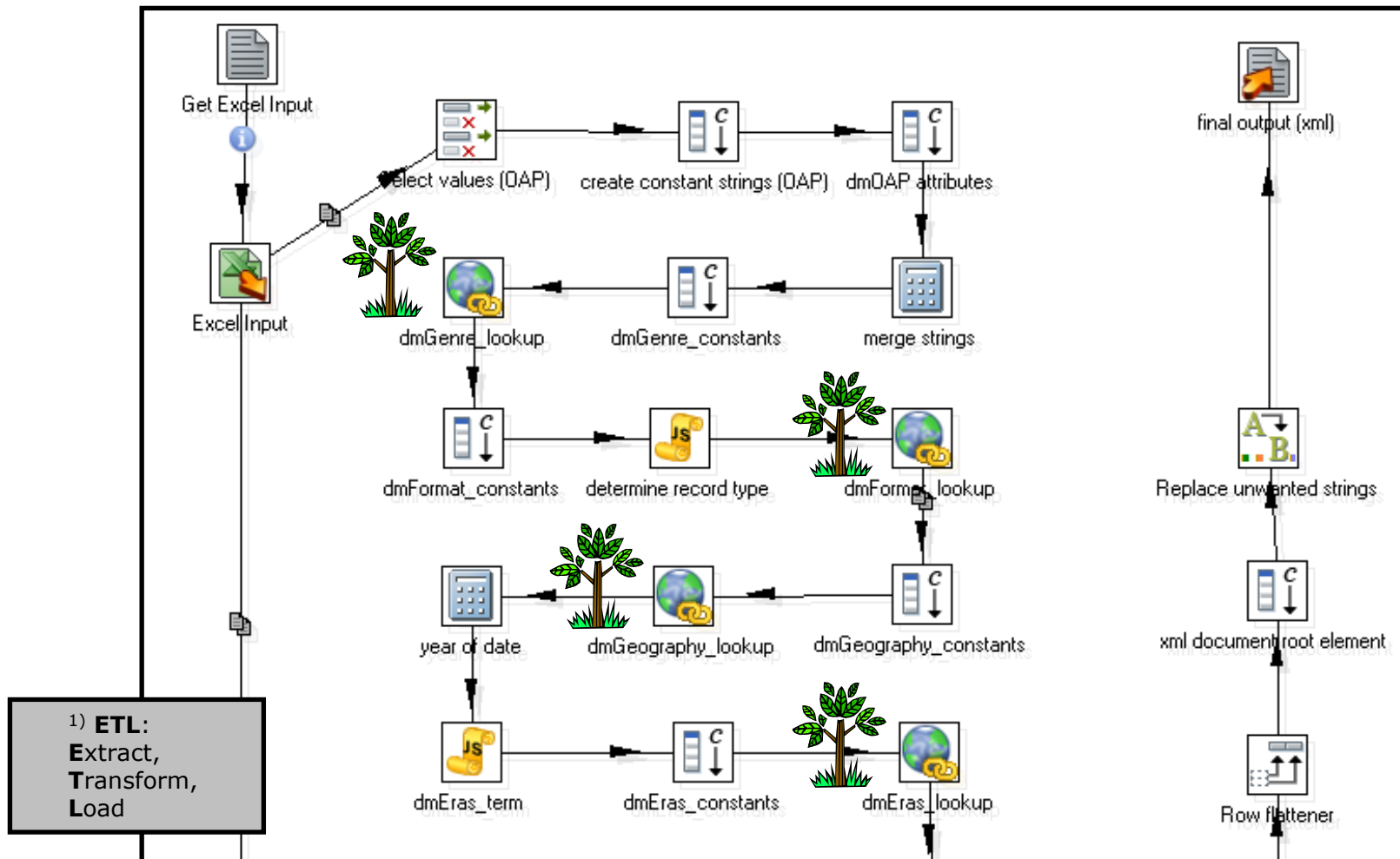
- ☒ Virtual Machine Image (TGN¹)-example: 16GB including database and all system components)
- ☒ Infrastructure as a Service (integrated in a Virtual Machine in a Cloud)

¹) Thesaurus of Geographic Names (2008 ~ 1 mio place names)



Example: Data Aggregation (semantic enrichment)

Usage of ETL¹⁾-Technology: DISMARC case (<http://www.dismarc.eu>)





Example: Page Level Indexing (1)

- ☒ BHL-Europe: TaxonFinder lookup on page level.

- ☒ (This function is an attempt to merge the TaxonGrab and FindIT algorithms.) It will identify **taxonomic name strings** within documents.
- ☒ **input arguments:**
 - ☒ url: URL of document to be searched
 - ☒ freeText: block of text to be searched.
(Will only be looked at if URL is empty.)
 - ☒ includeLinks: returns external URLs associated with returned name strings if set to 1

- ☒ **Output:** Taxon/Taxa

- ☒ **WSDL:** <http://names.ubio.org/soap/>




Example: Page Level Indexing (2)

- ☒ BHL-Europe: Query := “quercus alba”
(<http://bhl.ait.co.at/>)



Search | Advanced Search | Browse | Collections | Login | OAI | Help Wik







Search for OBJECTS > ? 






quercus alba in the field Any Field + -

Items with Digital Objects

1 items found The query was ("quercus alba") IN (ENG) Search result list (bibliographic)

Items 1 to 1 #1 [1] 10

#	Sel.	Organisation / Collection	Type Of Resource	Bibliographic citation	Resource
1.	<input type="checkbox"/>	CSIC		Contributions to morphometrics / Monografias (Museo Nacional de Ciencias Naturales), 8. Museo Nacional de Ciencias Naturales, C.S.I.C.; 264 p..	   

1 items found The query was ("quercus alba") IN (ENG) Search result list (bibliographic)

Items 1 to 1 #1 [1] 10

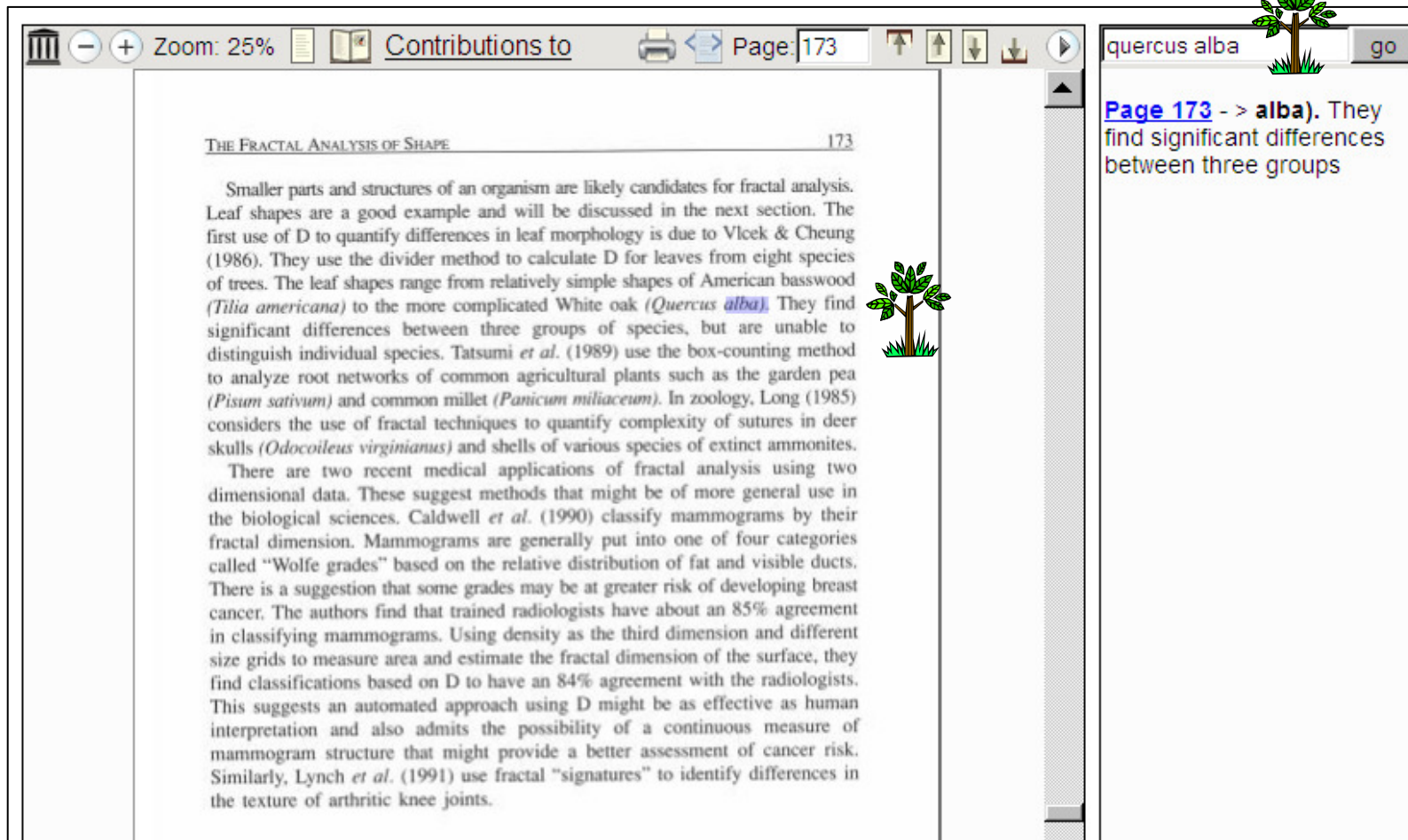




Example: Page Level Indexing (3)

Page Look Up

Subject:  **Quercus alba** -- [Link to page](#)



The screenshot shows a web browser window. The main content area displays a document titled "THE FRACTAL ANALYSIS OF SHAPE" on page 173. The text discusses fractal analysis of leaf shapes and mentions *Quercus alba*. A search sidebar on the right shows the search term "quercus alba" and a result for "Page 173 - > alba). They find significant differences between three groups".

THE FRACTAL ANALYSIS OF SHAPE 173

Smaller parts and structures of an organism are likely candidates for fractal analysis. Leaf shapes are a good example and will be discussed in the next section. The first use of D to quantify differences in leaf morphology is due to Vlcek & Cheung (1986). They use the divider method to calculate D for leaves from eight species of trees. The leaf shapes range from relatively simple shapes of American basswood (*Tilia americana*) to the more complicated White oak (*Quercus alba*). They find significant differences between three groups of species, but are unable to distinguish individual species. Tatsumi *et al.* (1989) use the box-counting method to analyze root networks of common agricultural plants such as the garden pea (*Pisum sativum*) and common millet (*Panicum miliaceum*). In zoology, Long (1985) considers the use of fractal techniques to quantify complexity of sutures in deer skulls (*Odocoileus virginianus*) and shells of various species of extinct ammonites.

There are two recent medical applications of fractal analysis using two dimensional data. These suggest methods that might be of more general use in the biological sciences. Caldwell *et al.* (1990) classify mammograms by their fractal dimension. Mammograms are generally put into one of four categories called "Wolfe grades" based on the relative distribution of fat and visible ducts. There is a suggestion that some grades may be at greater risk of developing breast cancer. The authors find that trained radiologists have about an 85% agreement in classifying mammograms. Using density as the third dimension and different size grids to measure area and estimate the fractal dimension of the surface, they find classifications based on D to have an 84% agreement with the radiologists. This suggests an automated approach using D might be as effective as human interpretation and also admits the possibility of a continuous measure of mammogram structure that might provide a better assessment of cancer risk. Similarly, Lynch *et al.* (1991) use fractal "signatures" to identify differences in the texture of arthritic knee joints.

quercus alba go

Page 173 - > alba). They find significant differences between three groups



Example: Search Term Selection (1)

☒ BHL-Europe/DISMARC

The screenshot displays the BHL-Europe/DISMARC search interface. At the top, it says "Search for OBJECTS >". Below this, there are radio buttons for "Free" and "Thesaurus", with "Thesaurus" selected. A search input field contains "XIXe siècle". There is also a checkbox for "Items with Digital Objects" and a "GO" button. A tree icon is visible next to the search input.

Below the search area, there is a section for "Eras" with a list of centuries from "1er siècle de notre ère" to "XXIe siècle". The "XIXe siècle" is selected, and a tree icon is next to it.

A dropdown menu is open, showing search fields. The "Date (Origin Info)" field is selected and highlighted in blue. The menu includes options like "Any Field", "My Favorites", "MODS", and various fields such as "Title Info", "Name", "Type Of Resource", "Genre", "Place (Origin Info)", "Publisher (Origin Info)", "Date (Origin Info)", "Language", "Physical Description", "Note", "Subject", and "- mods-Subject-url~".

On the right side of the interface, there is a "Multilingual options" button with a flag icon.



Example: Search Term Selection (2)

BHL-Europe/DISMARC – Result Record
 BHL

Search | Advance

Main Menu

BHL-Europe

User Menu

Login

My Basket

Select Language

English

Select Template

bhle

Detail

MODS Info

Title Info: Histoire de la lutte entre la science et la théologie / p Varigny,... et G. Adam,...

Name: White, Andrew Dickson (Type: personal) (Role Type: ...)

Type Of Resource: Text

Genre: monographie imprimée

Genre: printed monograph

Publisher (Origin Info): Guillaumin (Paris)

Date (Origin Info): 1899 (Type: dateOther)

Date (Origin Info): XIXe siècle (Type: dateOther) (Vocabulary: dmEras)

Language: français (Vocabulary: dmLanguages) (Type: code)

Physical Description: 1 vol. (536 p.) ; 23 cm

Physical Description: application/pdf

Related Item: Notice du catalogue : <http://catalogue.bnf.fr/ark:/12148/ark:/12148/cb316436749/description>

Related Item: Bibliothèque nationale de France (Type: original)

Location-Url: <http://catalogue.bnf.fr/ark:/12148/cb316436749/description> (hypertext)

Translations:

[eng] 19th century CE

[fin] 19. vuosisata jKr.

[bul] 19-ти век от н.е.

[zho] 19世紀

[ces] 19. století

[dan] 19. århundrede

[nld] de 19ste eeuw

[est] 19. sajand pKr

[fra] XIXe siècle

[deu] 19. Jahrhundert

[ell] 19ος αιώνας μ.Χ.

[hin] 19TH CENTURY CE

[hun] 19. század

[ita] XIX secolo d.C.

[jpn] 19世紀

[kor] 19세기

[lav] m.ē. 19. gadsimts

[lit] XIX amžius

[ltz] 19t Jorhonnert no Christus

[mit] DSATAH-IL SEKLU TA' L-ERA KOMUNI

[nob] 19. århundre

[pol] XIX wiek

[ron] secolul XIX e.n.

[rus] 19 век от Рождества Христова

[hrv] 19. stoljeće

[slv] 19. stoletje n.l.

[spa] Siglo XIX A.D.

[swe] 1800-talet evt

[tur] M.S. 19. Yüzyıl

[ukr] дев'ятнадцяте сторіччя після різдва Христова

co-funded by the European Commission
 Copyright © 2007/08 The BHL Consortium
 No part of this website may be reproduced, in any form, or by any means, without prior written permission of the BHL Consortium.
 Servvertime: 0.956 sec | Clienttime: 0.033 sec



Example: Collection Management Integration (1)

☒ TGN – Lookup in IMDAS (Screenshot provided by J.Sieglerschmidt)

The screenshot shows the 'Thesaurus' window in the IMDAS system. The search term 'Konstanz' is entered in the 'Thesaurusbegriff' field. A table below shows search results, with 'Konstanz' highlighted. A yellow callout bubble labeled 'Requested term' points to the search input. Another yellow callout bubble labeled 'Generic structure' points to the hierarchical path 'World/Europe/Deutschland/Baden-Württemberg/Freiburg/Konstanz' in the table. A third yellow callout bubble labeled 'SCOPE/tgn:note' points to the detailed description of Konstanz in the 'TGN-Ortsthesaurus' section. The interface includes navigation buttons like 'Hierarchisch', 'Alphabetisch', and 'Freie', and a 'Fundort/Herkunft' section at the bottom.

Geändert	Thesaurusbegriff	Benutzer	
	Bodensee		World/Europe/Bodensee □ □ The Bode... is a lake that border
	Konstanz		World/Europe/Deutschland/Baden-Württemberg/Freiburg/Konstanz

TGN-Ortsthesaurus
World/Europe/Deutschland/Baden-Württemberg/Freiburg/Konstanz
The city of Konstanz is located on Lake Constance, where the Rhein River exits it. It was the site of a Roman fort, which was taken by the Alemanni in the 3rd century BCE, and made an episcopal see in the 6th century. It prospered in the linen trade from 13th century, threw off episcopal rule in the 14th century, and became the leader of a confederacy of towns. It joined the Protestant Schmalkaldic League 1531, but became Roman Catholic again in 1547. It came under Austrian rule and remained there until ceded to the duchy of Baden in 1805. The city remains a cultural and economic center, and has much historic architecture, including the Kaufhaus built in 1388, the Renaissance town hall, the Gothic Rosgarten Museum, and the Insel Hotel, a former medieval Dominican monastery. There are also a university and an art museum. It is a popular tourist destination. The 2004 estimated population was 80,100.



Beispiel: Collection Management Integration (2)

SOAP-Result

SOAP-Request

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
  http://www.aitbiz.com/ns/Thesauri">
  <soapenv:Header/>
  <soapenv:Body>
    <thes:query>
      <session_ticket>4d2e0428b1c
      session_ticket>
      <thesaurus_name>tgn</thesa
      <simpleQuery>
        <term>Konstanz</term>
      </simpleQuery>
    </thes:query>
  </soapenv:Body>
</soapenv:Envelope>

```

```

<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<soapenv:Envelope xmlns:soapenv="http://schemas.xmlsoap.org/soap/envelope/"
  http://www.aitbiz.com/ns/Thesauri">
  <soapenv:Body>
    <ns1:queryResponse xmlns:ns1="http://www.aitbiz.com/ns/Thesauri">
      <message>Everything seems to be ok</message>
      <node gs="TGN:/World/Europe/Deutschland/Baden-Württemberg/Freiburg/Konstanz" id="TGN:7005174"
      schema="TGN">
        <BT href="1003143">
          <TRM xml:lang="eng">Freiburg</TRM>
        </BT>
        <NT href="4012654">
          <TRM xml:lang="eng">Altstadt</TRM>
        </NT>
        <TRM xml:lang="eng">Konstanz</TRM>
        <TRM xml:lang="eng">Constantia</TRM>
        <TRM xml:lang="eng">Kostnitz</TRM>
        <TRM xml:lang="eng">Konstanz</TRM>
        <TRM xml:lang="eng">Constance</TRM>
        <SCOPE xmlns:tgn="urn:tgn">
          <tgn:note>The city of Konstanz is located on Lake Constance, where the Rhein River exits it. It was
          the site of a Roman fort, which was taken by the Alemanni in the 3rd century BCE, and made an episcopal see in
          the 6th century. It prospered in the linen trade from 13th century, threw off episcopal rule in the 14th century, and
          became the leader of a confederacy of towns. It joined the Protestant Schmalkaldic League 1531, but became
          Roman Catholic again in 1547. It came under Austrian rule and remained there until ceded to the duchy of Baden
          in 1805. The city remains a cultural and economic center, and has much historic architecture, including the
          Kaufhaus built in 1388, the Renaissance town hall, the Gothic Rosgarten Museum, and the Insel Hotel, a former
          medieval Dominican monastery. There are also a university and an art museum. It is a popular tourist destination.
          The 2004 estimated population was 80,100.</tgn:note>
          <tgn:Term name="Constantia" other_flags="NA" preferred="V" vernacular="V">
            <tgn:DisplayDate>ancient name, possibly referring to reputed founding by Constantius Chlorus ca.
            300 AD</tgn:DisplayDate>
            <tgn:Qualifier/>
            <tgn:StartDate>Thu Jan 01 01:00:00 CET 1970</tgn:StartDate>
            <tgn:EndDate>Thu Jan 01 01:00:01 CET 1970</tgn:EndDate>

```



Generic structure

Requested term

SCOPE/tgn:note



Example: Collection Management Integration (3)

SOAP Test Client Test-Client (previous version of API)

tgn_v04.wsdl [\[Download\]](#) [\[Change\]](#)

Result

[SOAP Request \(XML\)](#)

[SOAP Response \(XML\)](#)

[Call statistics](#)

Methods

[GetTreeNodes](#)

[GetTreeNodesQualified](#)

[IndexLookup](#)

[Login](#)

[Logout](#)

Result of GefTreeNodes

Called:

```
GetTreeNodes( session_ticket = "95b80a4b39b008158d253ad645b24c9b84c934ca", TRM = "Konstanz*", SCOPE = {{{ 'field' => 'SCOPE', 'query' => " " }}, language = "", BT = "", GS_LEVEL = "", max_subtree_depth = "2", max_node_count = "30")
```

Returned:

⊖ [Konstanz* + 2 + 30

- ⊖ World
 - ⊖ Europe
 - ⊖ Deutschland
 - ⊖ Baden-Württemberg
 - ⊖ Freiburg
 - ⊖ Konstanz
 - Constantia
 - Constanz
 - Kostnitz
 - Altstadt
 - Constance

[\[Show result list\]](#)

costs	duration	0.324
	bandwidth	16792
	hits	7
	license	0
	e_points	7.18
	e_points_left	7635126.5

Konstanz

LANG : V

TRM : Konstanz

BT : Freiburg

NT : Altstadt

GS LEVEL : 6

GS : TGN:/World/Europe/Deutschland/Baden-Württemberg/Freiburg/Konstanz

RT :

UF : Constance, Konstanz, Kostnitz, Constantia

level : 0

identifier : TGN:40312

SCOPE : description : Located on Bodensee (Lake Constance) where Rhein river exits; was site of Roman fort; taken by Alemanni in 3rd cen. AD; prospered in linen trade from 13th cen.; headed confederacy of towns; joined Protestant Schmalkaldic League 1531, Catholic again 1547.



Thank You

Steinbeis Innovation Transferring Centre

Information Management and Cultural Heritage Informatics (IMCHI)

Leiter: Univ.Prof.Dr.**Walter Koch** (TU-Graz)

A-8010 Graz, Klosterwiesgasse 32/1

Fon: +43.316.8182228

Fax: +43.316.8182228-30

E-Mail: walter.koch@stw.de (walter.koch@tugraz.at)

Internet: <http://www.stw.de/>