Design of a Pilot SRW-compliant Terminologies Mapping Service (HILT)

Terminologies Workshop, ECDL, Alicante, 2006

Dennis Nicholson, CDLR, Strathclyde University

HILT: Background and Overview

- Funded JISC; Support: OCLC; Collaborative
- Aim: provide subject interoperability in a multi-scheme environment via inter-scheme mapping
- Ideally by identifying a generic approach, able to be built up through distributed collaborative action…
- Originally: intellectual mapping, but now see how model can include range of interoperability services
- Phase I, II, M2M FS; Now: Phase III (main focus)

HILT Phase III [Nov '05 – Jan '07]

✦ Aim: an M2M pilot that:

- Offers terminology services via SRW but is open to later extension to (Z39.50; SRU)

- Uses SKOS-Core as the 'mark-up' for sending out terminology sets and classification data but open to later extension to other formats (MARC; Zthes)

- Is open to the possibility of a distributed approach to building a full service up via wide collaboration

- Extends the user-accessible (non-M2M) Phase II pilot beyond inter-scheme mapping

Phase III: How Phase II Pilot works

✦ Offers mapping based subject interoperability via a DDC spine, and works like this:
  - The user enters a subject term, which is used to search the database for DDC captions that may fit the user's topic
  - Captions / numbers returned; user chooses best match
  - The DDC number chosen is used to find collections covering the user's subject and the subject schemes they use in a collections database, and the best term for the user's topic in any given scheme
  - Sample retrieval is provided where possible

✦ Screen shots to illustrate

The HILT terminologies server aims to identify JISC services and/or collections likely to have resources relevant to any subject query you ma process has three steps:

1. You enter your search term (or browse the subject hierarchies).
2. HILT looks for the best matches for your subject and asks you to clarify which is most appropriate.
3. You choose the most appropriate subject.
4. HILT tells you about possible services or collections that may interest you, tells you what subject schemes they use, and tells you the use (if it can). It also allows you to connect through and do a search of the services and collections identified.

Enter your search term here...

Teeth   [Search]

Use Double Quotes for a phrase search (Eg; "biology and life science") See search tips for details

...or browse by category

Arts & recreation → Architecture , Arts , Drawing & decorative arts , Graphic arts , Landscaping & area planning , Music ...

Computers, information & general reference → Associations, organizations & museums , Bibliographies , Computers, Internet & systems , Encyclopedias & books of Journalism, publishing & news media , Library & information science ...

History & geography → Biography & genealogy , Geography & travel , History , History of Africa , History of Asia , History of Europe (ca. 5

Language → Classical & modern Greek languages , English & Old English languages , French & related languages , German & related Italian, Romanian & related languages , Language ...

Literature → American literature in English , Classical & modern Greek literatures , English & Old English literatures , French & rel literatures , German & related literatures , Italian, Romanian & related literatures ...

Philosophy & psychology → Ancient, medieval & eastern philosophy , Astrology, parapsychology & the occult , Epistemology , Ethics , Logic , M

Christian denominations , Christian pastoral practice & religious orders , Christian practice & observance , Christianit

- **Description**
- **Top levels browse hierarchy**
- **Search box – common - teeth**

Search Terms

Searched for:

`teeth`   [Search]

Change search terms or start a new search

Click on a link to retrieve collections relevant to that term

1. Prosthetic dentistry (Prosthodontics)

Technology ... Medicine ... Miscellaneous branches of medicine Surgery ... Dentistry ... Prosthetic dentistry (Prosthodontics)

2. Cavities (Caries)

Technology ... Medicine ... Miscellaneous branches of medicine Surgery ... Dentistry ... Cavities (Caries)

3. Dental diseases

Technology ... Medicine ... Forensic medicine, incidence of disease, public preventive medicine ... Incidence of and public measures to prevent specific d
kinds of diseases ... Diseases of regions, systems, organs; other diseases ... Other diseases ... Dental diseases

4. Teeth

Science ... Biology & life sciences ... Specific physiological systems in animals, regional histology and physiology ... Digestive system ... Mouth and eso
Teeth

- **Dental diseases 614.5996 in DDC, used for 3 things**
- **Truncation used to find these relevant collections at DDC 610**
- **This identifies the subject schemes they use (MeSH here)**
- **Best term in scheme via mapping to DDC 614.5996 (dentition)**

Search Terms

Searched for:

Dental diseases [Search]

Change search terms or start a new search

Selected subject:

Dewey 21 > Technology > Medicine > Forensic medicine, incidence of disease, public preventive medicine > Incidence of and public measures to prevent specific diseases and kinds of diseases > Diseases of regions, systems, organs; other diseases > Other diseases > Dental diseases

Results from JISC collections / services database

Searched for: Dental diseases (Dewey No: **614.5996**)

1. Bristol Biomedical Image Archive

A collection of medical, dental and veterinary images for use in teaching.

Dewey No   : 610
Associated Taxonomy   : MESH

5. OMNI: Organising Medical Networked Information

OMNI is a gateway to evaluated Internet resources in health and medicine, aimed at students, researchers, academics and practitioners in the health and medical sciences. OMNI is created by a core team of information specialists and subject experts based at the University of Nottingham Greenfield Medical Library, in partnership with key organisations throughout the UK and further afield. OMNI is one of the gateways within the BIOME service.

Dewey No   : 610
Associated Taxonomy   : MESH
Subject Terms in Mesh and their relationship   : **Dentition** (broader term match)

Click on the subject term to search the collection directly

OMNI

NMAP

VETGATE

BIORES

NATURAL

AGRIFOR

a service of the
R D N

THE
WELLCOME
GATEWAYS

BioethicsWeb

MedHist

psci-com

No. of results: 4.

dentition   [Search]   Advanced | Help

Dental health

A public-oriented document published by the British Nutrition Foundation, a UK charity partially funded by the food industry and government. The guide addresses tooth development and structure, demineralisation and remineralisation, dental caries and factors that affect them. In PDF requiring Adobe Acrobat Reader.

Tooth; Popular Works [Publication Type]; Oral Hygiene; Oral Health; **Dentition**; Dental Caries;

Document/report / Other

Mouth and teeth : how to keep them healthy

A patient education leaflet providing information on the mouth and teeth. Topics covered include children's teeth, dental care for adolescents and adults, taking care of bad breath, how to stop smoking smokeless tobacco, and how to keep your mouth and teeth healthy. Published by the American Academy of Family Physicians (AAFP).

Patient Education Handout [Publication Type]; Oral Health; Mouth Diseases; Halitosis; **Dentition**; Dental Health Services;

Document/report / Patient information leaflet

Treating mandibular edentulism

An informal exploration of the treatment of mandibular edentulism. This tutorial, which is

• **Finally, dentition used to search last of these collections via OpenURL; send back relevant hits**

HILT M2M Pilot Architecture
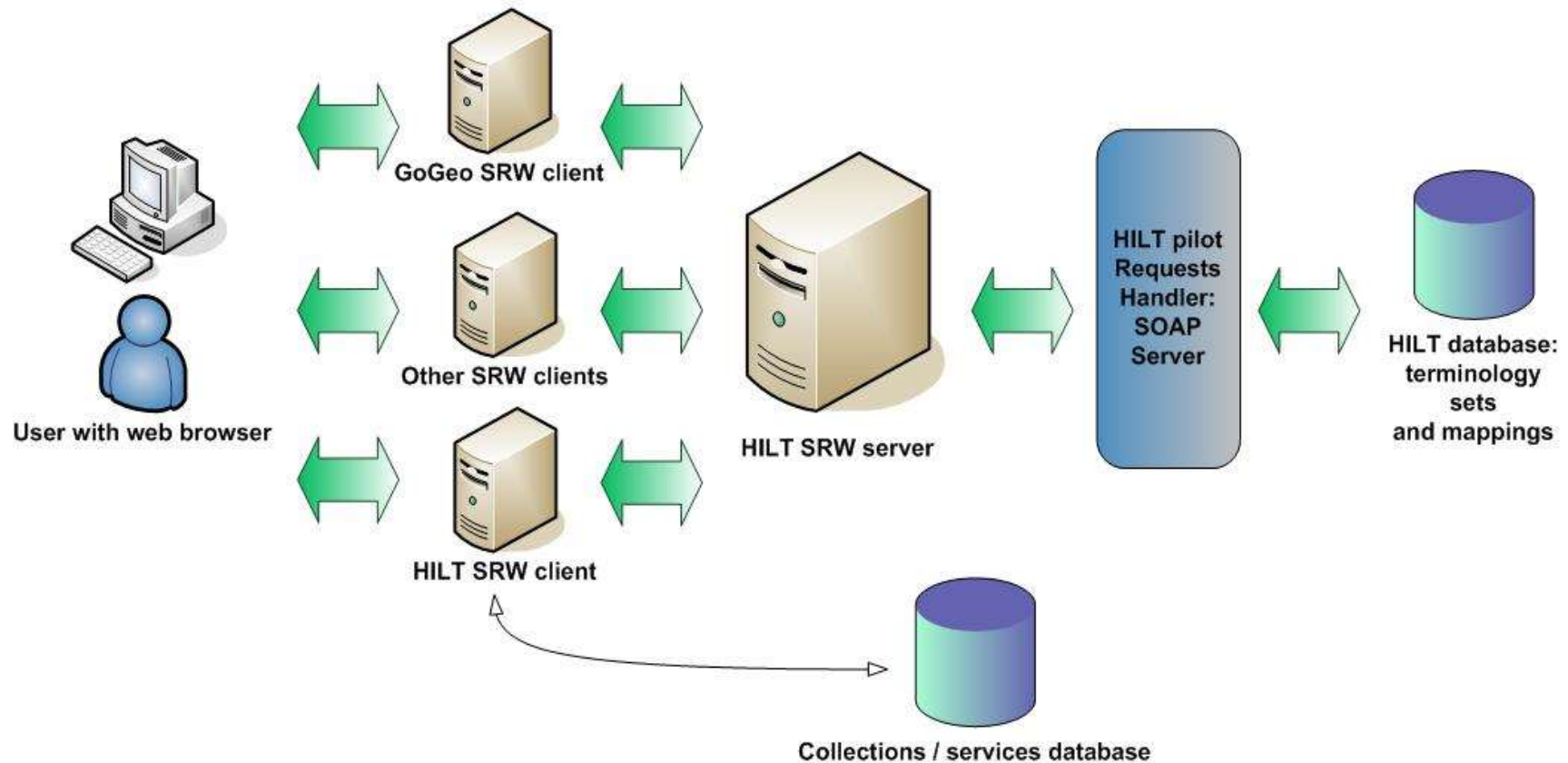
- **Diagram: architecture SRW version; Blue (II): Users, browsers, HILT RH PHP/web service, screens**
- **Grey: additional SRW elements: users, services, embedded clients, Two: HILT Phase II +, GoGeo**
- **service specific; Collections/services dbase via client/ RH SOAP server queries database; SKOS**

```
HTTP/1.1 200 OK
Connection: close
Date: Wed, 16 Aug 2006 15:27:32 GMT
Server: Microsoft-IIS/6.0
X-Powered-By: PHP/5.1.4
Server: NuSOAP Server v0.7.2
X-SOAP-Server: NuSOAP/0.7.2 (1.95)
Content-Type: text/xml; charset=ISO-8859-1
Content-Length: 7209


<?xml version="1.0" encoding="ISO-8859-1"?>
<SOAP-ENV:Envelope SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/" xmlns
<SOAP-ENV:Body>
<ns1:get_filtered_setResponse xmlns:ns1="http://tempuri.org">
<return xsi:type="SOAP-ENC:Array" SOAP-ENC:arrayType="xsd:string[1]">
<item xsi:type="xsd:string">
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xml:base="http://hiltm2m.cdlr.strath.ac.uk/hiltm2m/concepts.php">
<rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
<skos:ConceptScheme rdf:about="http://hiltm2m.cdlr.strath.ac.uk/GCMD"/>

<skos:Concept rdf:about#Armed forces">
<skos:prefLabel>Armed forces</skos:prefLabel>
<skos:broader rdf:resource="#Defence"/>
<skos:narrower rdf:resource="#Cadet forces"/>
<skos:narrower rdf:resource="#Royal Air Force"/>
<skos:narrower rdf:resource="#British Army"/>
<skos:narrower rdf:resource="#Royal Navy"/>
<skos:narrower rdf:resource="#Territorial Army"/>
<skos:related rdf:resource="#Service associations"/>
<skos:related rdf:resource="#Mercenaries"/>
```
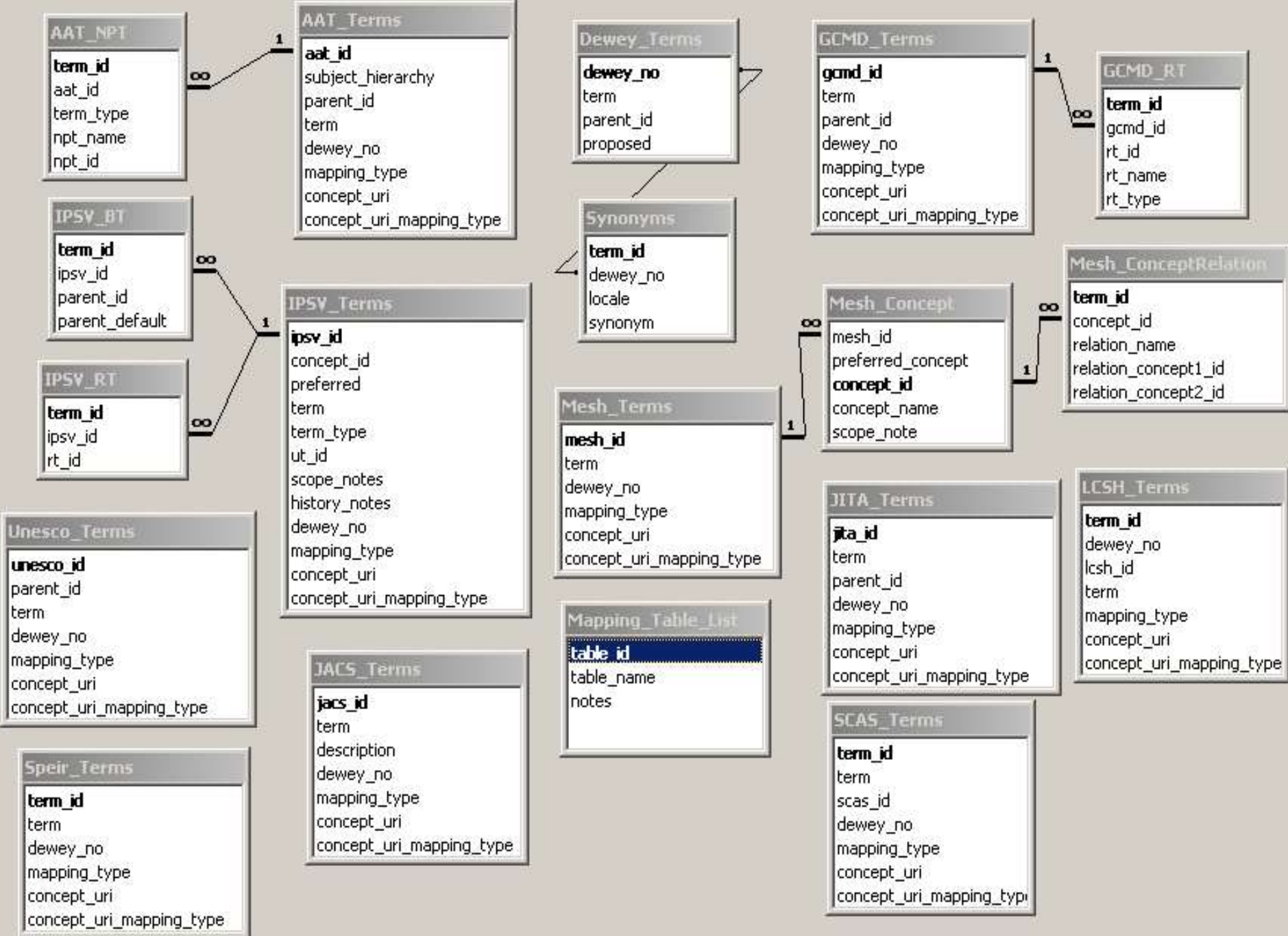
AAT_NPT
- term_id
- aat_id
- term_type
- npt_name
- npt_id

AAT_Terms
- aat_id
- subject_hierarchy
- parent_id
- term
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

IPSV_BT
- term_id
- ipsv_id
- parent_id
- parent_default

IPSV_RT
- term_id
- ipsv_id
- rt_id

IPSV_Terms
- ipsv_id
- concept_id
- preferred
- term
- term_type
- ut_id
- scope_notes
- history_notes
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

Dewey_Terms
- dewey_no
- term
- parent_id
- proposed

Synonyms
- term_id
- dewey_no
- locale
- synonym

GCMD_Terms
- gcmd_id
- term
- parent_id
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

GCMD_RT
- term_id
- gcmd_id
- rt_id
- rt_name
- rt_type

Mesh_Concept
- mesh_id
- preferred_concept
- concept_id
- concept_name
- scope_note

Mesh_Terms
- mesh_id
- term
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

Mesh_ConceptRelation
- term_id
- concept_id
- relation_name
- relation_concept1_id
- relation_concept2_id

Unesco_Terms
- unesco_id
- parent_id
- term
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

JACS_Terms
- jacs_id
- term
- description
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

Mapping_Table_List
- table_id
- table_name
- notes

JITA_Terms
- jita_id
- term
- parent_id
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

LCSH_Terms
- term_id
- dewey_no
- lcsh_id
- term
- mapping_type
- concept_uri
- concept_uri_mapping_type

SCAS_Terms
- term_id
- term
- scas_id
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

Speir_Terms
- term_id
- term
- dewey_no
- mapping_type
- concept_uri
- concept_uri_mapping_type

• **Database structure**

Data; 6 Functions; Emulate Phase II

✦ Get_DDC_records
- Returns DDC captions and numbers related to an input term; list for user to choose best fit caption/no.

✦ Get_collections
- Returns collections classified under a specified DDC number or its stem, including subject scheme used

✦ Get_non_DDC_records
- Returns mappings to other schemes from a specified (untruncated) DDC number

Functions: Phase II and beyond

- **Get_all_records**
  - Combines the functions of get_DDC_records and get_non_DDC_records seen above
- **Get_explain**
  - Provides information to feed SRW Explain requests
- **Get_filtered_set**
  - Allows specified fields from specific terminologies or combinations of terminologies to be searched – enabling functionality beyond phase II to be added

Caries (Cavities)
Dental diseases
Dental diseases
Dental radiography-dentistry, . . .
Dentistry
Diseases of gums and tooth sockets
Diseases of tooth tissues
Head
Oral hygiene and preventive dentistry
Physical adaptation
Prosthodontics
Scaphopoda
Teeth
Teeth
Teeth
Teeth
Tooth diseases-animals-veterinary medicine
Tooth diseases-humans-cancer
Tooth diseases-humans-diagnosis
Tooth diseases-humans-etiology
Tooth diseases-humans-therapy

- Next: how SRW clients use functions to emulate pilot
- Top:SOAP get_DDC_records indexed under 'teeth'
- Middle: DDC captions returned (with numbers)
- User picks best fit caption: dental diseases; 614.5996
- Bottom: SOAP get_collections - 614.5996; stem (610)

http://hiltpilot.cdlr.strath.ac.uk/soap/hiltsoapdemo.php?request=get_collections&ddc=614.5996

Bristol Biomedical Image Archive (ddc: 610)
Cochrane Library (ddc: 610)
EMBASE (ddc: 610)
MIDRIB: Medical Images Digitised Reference Information Bank (ddc: 610)
OMNI: Organising Medical Networked Information (ddc: 610)

(Schemes)

- **Top: repeat of request; get_collections&DDC=614.5996**
- **Upper middle: collections classified at 610; schemes used**

http://hiltpilot.cdlr.strath.ac.uk/soap/hiltsoapdemo.php?request=get_non_ddc_records&ddc=614.5996

Dentition(Mesh)

- **Lower middle: get_non_ddc_records for terms mapped to 614.5996**
- **Bottom: one returned term (dentition best term in MeSH)**

BIOME

OMNI

The UK's gateway to high quality Internet resources in Health and Medicine

SEARCH BROWSE   ABOUT WHAT'S NEW SUBMIT TRAINING SITE MAP

OMNI
NMAP
VETGATE
BIORES
NATURAL
AGRIFOR

a service of the
R—D—N

THE
WELLCOME
GATEWAYS

BioethicsWeb

MedHist

psci-com

No. of results: 4.

• **Dentition used to search last collection for relevant hits**

dentition   Search   Advanced | Help

Dental health

A public-oriented document published by the British Nutrition Foundation, a UK charity partially funded by the food industry and government. The guide addresses tooth development and structure, demineralisation and remineralisation, dental caries and factors that affect them. In PDF requiring Adobe Acrobat Reader.

Tooth; Popular Works [Publication Type]; Oral Hygiene; Oral Health; **Dentition**; Dental Caries;

Document/report / Other

Mouth and teeth : how to keep them healthy

A patient education leaflet providing information on the mouth and teeth. Topics covered include children's teeth, dental care for adolescents and adults, taking care of bad breath, how to stop smoking smokeless tobacco, and how to keep your mouth and teeth healthy. Published by the American Academy of Family Physicians (AAFP).

Patient Education Handout [Publication Type]; Oral Health; Mouth Diseases; Halitosis; **Dentition**; Dental Health Services;

Document/report / Patient information leaflet

Treating mandibular edentulism

An informal exploration of the treatment of mandibular edentulism. This tutorial, which is

Switch to M2M: Advantages

✦ Services can use a standard web services protocol and query language to interact with HILT and other terminology services
✦ They can use HILT services selectively
✦ They can offer enhanced services that are transparent to their users
✦ They can utilise what they know about their users and their behaviour to interact more usefully with HILT

Phase III pilot: Not Just Phase II

✦ Extended Features:
  ▪ A baseline SRW open source client that includes the DDC collections-finding code
  ▪ More schemes: DDC, LCSH, IPSV, AAT, GCMD HASSET, MeSH, NMR, JACS, UNESCO
  ▪ Additional – but still illustrative - mappings
  ▪ Detailed data on terms in schemes: BT, NT,RT, synonyms, scope note etc.

✦ Last should allow clients to generate – and allow users to navigate - scheme hierarchies

Future Possibility: A Baseline Service?

- ✦ Basis of generic collaborative distributed solution?
- ✦ If clients can generate scheme hierarchies; initial but extendable service based on top level mappings and hierarchy based collection retrieval looks feasible
- ✦ Deeper levels of mapping as and when possible
- ✦ Distributed approach could allow faster progress on scheme expansion and deeper mapping
- ✦ Model open to external interoperability/terminology services, not just local intellectual mappings (CSD)

Beyond Phase III – A Possible Path?

✦ Could be used for distributed collaborative work; research into retrieval effectiveness; user needs
✦ Open question; Obvious practical issues

- Local hits can improve user topic identification

- But can hierarchies offer effective retrieval?

- Is DDC practical for deeper mapping? (costs)

- If not, is mapping via SKOS concept URIs?

- How can services database best categorise mapping and terminology services, schemes and versions for clients?

Other Issues Impacting on Futures

+ Currently a probable path, but too early to be sure
+ Within Phase III, still have to:
  - Complete illustrative mappings, SOAP, SRW server/clients
  - Look at the feasibility and design of a distributed approach
+ Baseline service possibility looks most attractive if distributed, but implications of this yet to be explored
+ Unknowns: JISC review of shared services (like HILT) and a report to JISC on the terminologies area generally

Further Information

- Website (all HILT phases):
  - http://hilt.cdlr.strath.ac.uk/
- E-mails:
  - d.m.nicholson@strath.ac.uk
  - george.macgregor@strath.ac.uk
  - e.mcculloch@strath.ac.uk
- HILT (SOAP) Demonstrator:
  - http://hiltm2m.cdlr.strath.ac.uk/hiltm2m/hiltsoapclient.php