



Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



Lafayette Ragsdale - Limber Up ! A Limited Edition Civil War Print

Outline

Introductions - Myself, Project, Thesauri, RDF

**The Challenge - To Facilitate Cross-European
Comparative Data Analysis**

The Dream - A Computer Data-ing Service

The Reality - The Limber Approach

The Progress - Metadata, Thesaurus, Workshop

The End - A Possible Sponsor and Questions ???

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Introductions

Ken Miller - Head of Information Systems

Development, The UK Data Archive, The University of
Essex, England

**EU Human Language Technologies IST
Project - Language Independent Metadata**

Browsing of European Resources.

CLRC Rutherford Appleton Laboratory, Intrasoft, UK
Data Archive , Norwegian Data Archive, user
group:- other European national archives

To Breakdown Linguistic and Discipline Barriers to
interoperability between European Resources

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Searching

[Go back to the](#)

Softness. [324](#)

Adj. soft,
lissom, limber,
extensile,
remollient



**Language
Independent
Metadata
Browsing of
European
Resources**

the, lithesome;
malleable,

Hierarchical Thesauri

List of specific synonyms

Broader and narrower relationships

Subject concept

Concept tree - hierarchies

Shift focus within concept

Ensured relevance

Browsing and KWIC listings

Apply to different parts of metadata

Ranking

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

NT

EMPLOYMENT HISTORY,
EMPLOYMENT PROGRAMMES,
FULL-TIME EMPLOYMENT,
HOME-BASED WORK,
IRREGULAR ECONOMIC
ACTIVITY, JOB CHANGING,
JOB LOSSES,
JOB SHARING,
JOB VACANCIES,
MILITARY SERVICE,
OFF-SHORE EMPLOYMENT,
PART-TIME EMPLOYMENT,
SEASONAL EMPLOYMENT,
SHELTERED EMPLOYMENT,
SPOUSES' EMPLOYMENT,
STUDENT EMPLOYMENT,
SUBSIDIARY EMPLOYMENT,
TEMPORARY EMPLOYMENT,
UNEMPLOYMENT,
WOMEN'S EMPLOYMENT,
YOUTH EMPLOYMENT

UF

JOBS, WORK

BT

LABOUR AND EMPLOYMENT

RT

CONDITIONS OF EMPLOYMENT,
EMPLOYMENT ABROAD,
EMPLOYMENT
OPPORTUNITIES,
EMPLOYMENT POLICY,
EMPLOYMENT SERVICES,
EMPLOYMENT STATUS,
LABOUR (WORK),
LABOUR ECONOMICS,
OCCUPATIONS,
PERSONNEL,
PERSONNEL MANAGEMENT,
RIGHT TO WORK,
WORKING CONDITIONS

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

RDF - Resource Description Framework

World-Wide Web Consortium (W3C)

**A framework within which statements about
Web resources can be recorded**

Uses XML as its physical syntax

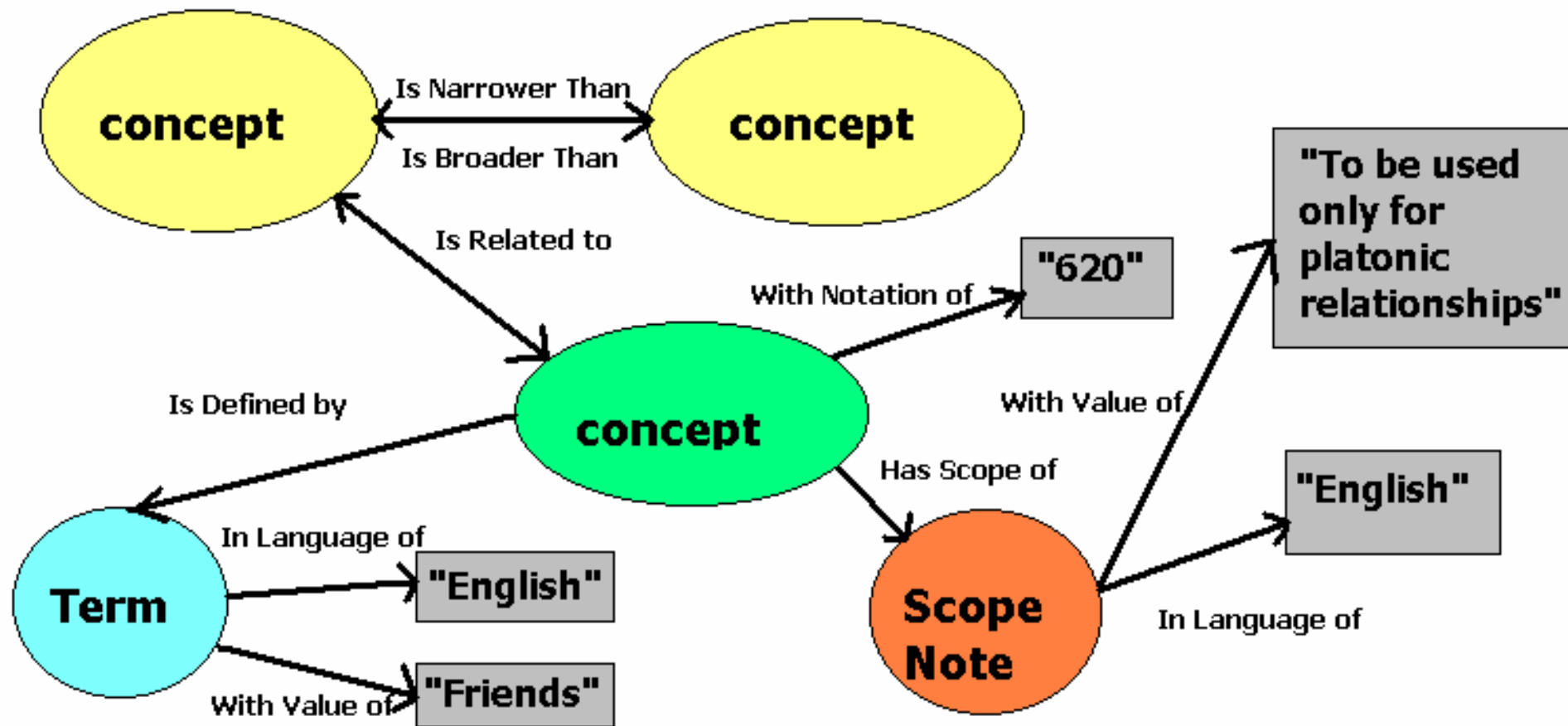
Based around directed labelled graphs

Triples - Resource, Property, Value

Can construct more complex models

Statements about statements

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

```
<web:Description about="http://elsst/concepts/CID_6">  
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#Concept"/>  
<rdfs:isDefinedBy web:resource="http://elsst/concepts/" />  
  
<thes:indicator web:resource="http://elsst/terms/TID_3"/>  
  
<thes:notation>620</thes:notation>  
  
<thes:scope web:resource="http://elsst/scopenotes/SN_12"/>  
  
<thes:relatedConcept web:resource="http://elsst/concepts/CID_15"/>  
</web:Description>
```

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

```
<web:Description about="http://elsst/terms/TID_3">  
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#Term"/>  
<thes:lang>en</thes:lang>  
<web:value>Friends</web:value>  
<thes:termType web:resource="http://rdf-dot/Thes/Thes.xrdf#preferred"/>  
</web:Description>
```

```
<web:Description about="http://elsst/scopenotes/SN_12">  
<web:type resource="http://rdf-dot/Thes/Thes.xrdf#ScopeNote"/>  
<thes:lang>en</thes:lang>  
<web:value>To be used only for platonic relationships</web:value>  
</web:Description>
```

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

The Challenge - To Facilitate Cross-European Comparative Data Analysis

- A metadata model and representation to allow integration within and across data archives.
- A multilingual thesaurus to index and access social science datasets in data archives.
- A multilingual query and retrieval tool to allow queries to datasets in archives to be made in several languages with keyword and phrase translation in the retrieved metadata.
- Tools to support the construction and maintenance of datasets in an archive using automatic indexing.
- An XML metadata server using the metadata model and representation.

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

“Effect of Diet on Children’s Teeth”

**Greenhouse effect, Diet, Diet and nutrition, Diet
therapy, Children’s rights, False teeth, teeth**

TEETH

nt DENTURES

rt DENTAL DISEASES

rt DENTAL HEALTH

“National Diet, Nutrition and Dental Survey”

Show me more

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Computer dating service for datasets

Metadata elements to search for a perfect match

Metadata to display to confirm a perfect match

Analysing tools to ensure a perfect match

Resource Independent

Language Independent

Thesaurus Independent

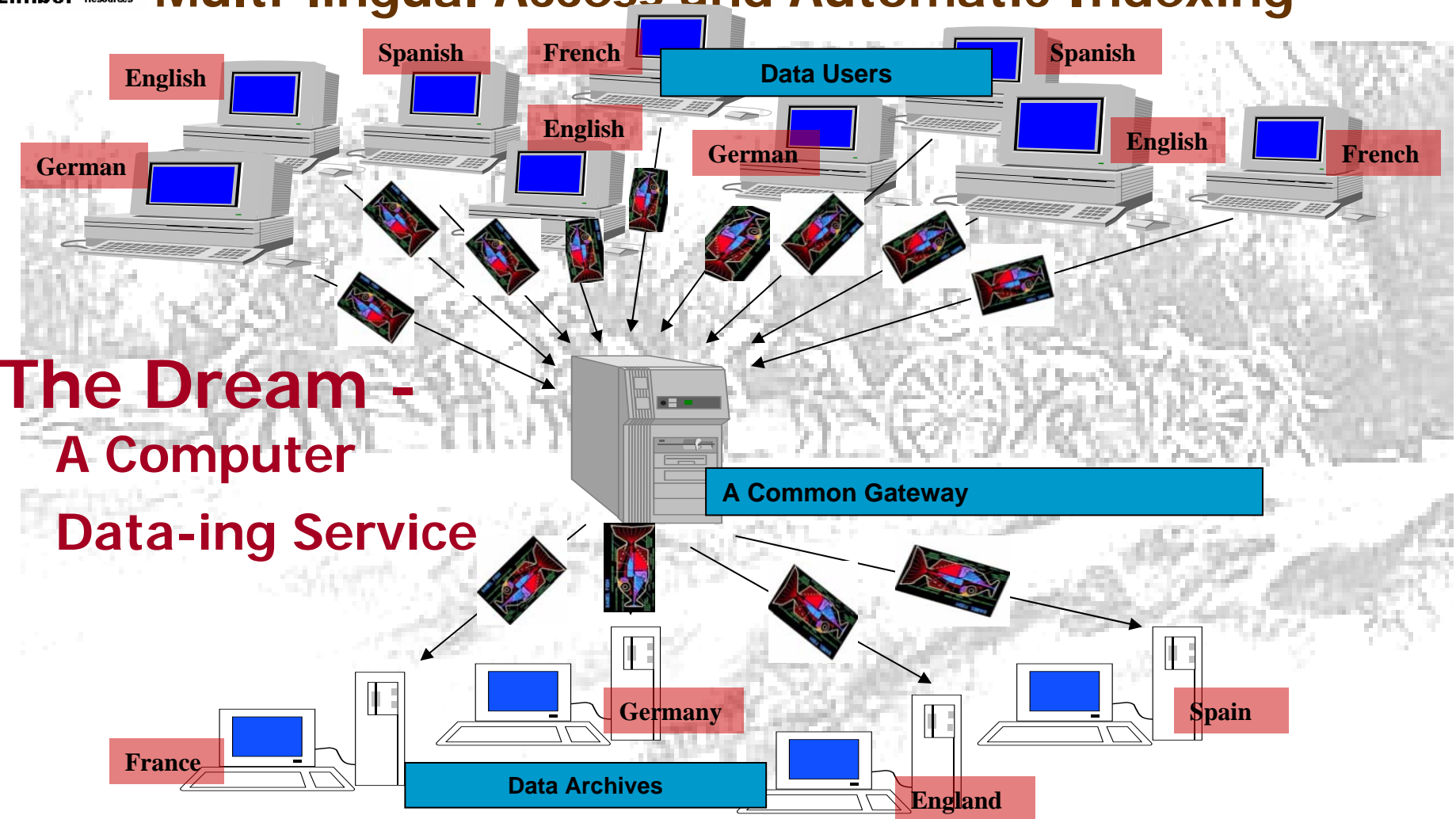
Dream on

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



**The Dream -
A Computer
Data-ing Service**

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Women's Employment and Family Life in the United Kingdom, Norway, France, Czech Republic and Russia

Documentation Table Graphics

Table options

42: DDL_NOW: Domestic division of labour: now
7: MUMOCC: Mother's occupation

	semi/uns	Skilled	low pro	higher pro	entrepre	housewif	Total
traditional	41.7	60.6	53.8	73.9	60.0	50.0	55.9
shared	33.3	24.2	25.6	13.0	0.0	32.1	25.0
men do most	8.3	0.0	5.1	0.0	0.0	3.6	3.3
third party	4.2	0.0	7.7	4.3	20.0	7.1	5.3
lives alone	8.3	15.2	7.7	8.7	20.0	7.1	9.9
6	4.2	0.0	0.0	0.0	0.0	0.0	0.7
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
N=	24	33	39	23	5	28	152

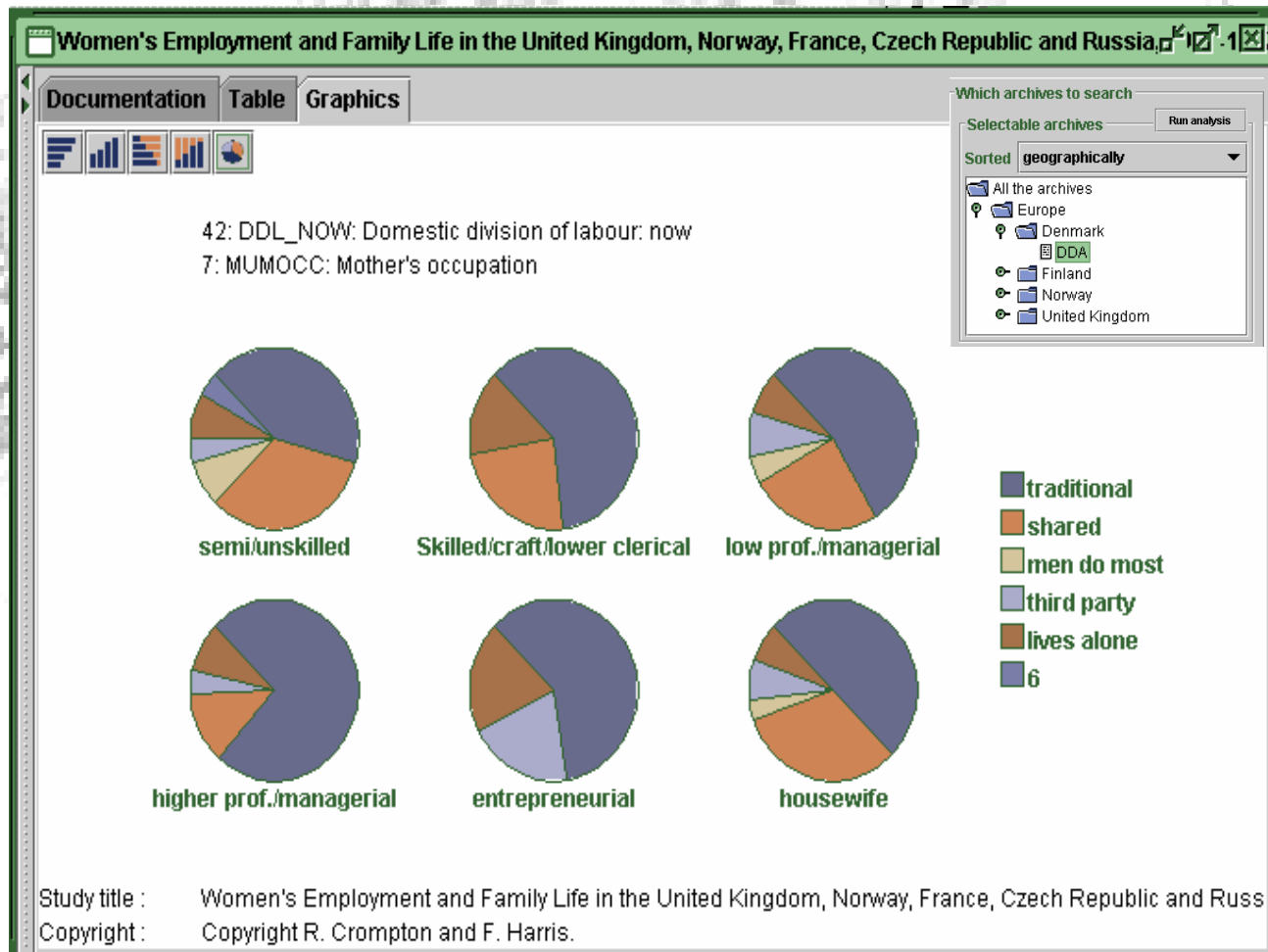
Study title : Women's Employment and Family Life in the United Kingdom, Norway, France, Czech Republic and Russ
Copyright : Copyright R. Crompton and F. Harris.

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

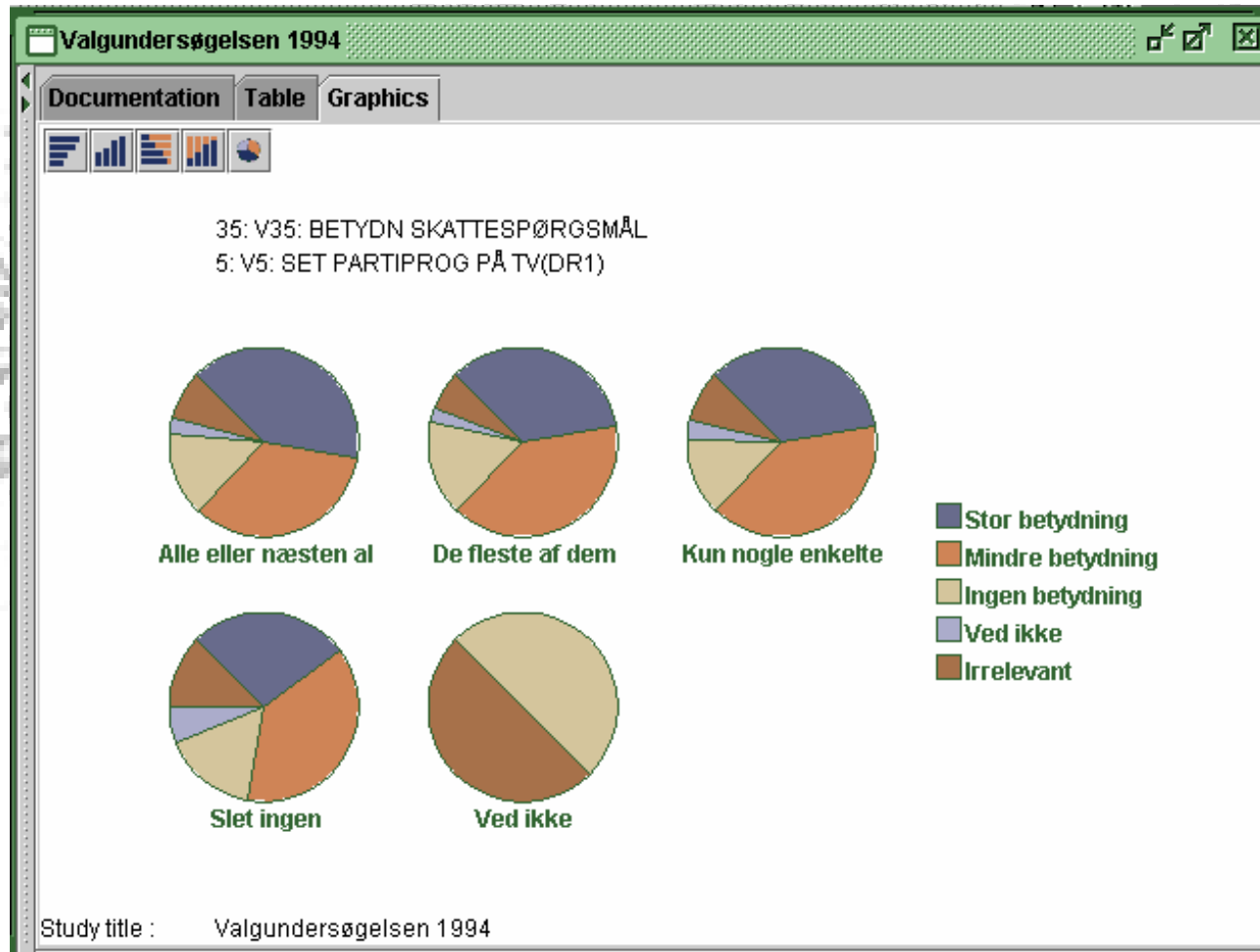


IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

The Reality - The Limber Approach

Discipline boundaries

RDF

Mappings between metadata standards

Linguistic boundaries

Multi-lingual Thesaurus

Interface

Relevance feedback

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Convert DDI codeBook to RDF
Translate DDI Headings/Tag Library
Reduce HASSET
Add Methodology Terms
Translate Hierarchies
Check individual languages
Convert to XML/RDF
Multi-mapped 4 Language Thesaurus

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

The Progress - Metadata, Thesaurus, Workshop

Reduction of Thesaurus

Automation NOT possible

Removal of Cultural specificity

Removal of Institutional specificity

Removal of hierarchies in existing thesauri

Initial reduction of major 20 hierarchies

A top down reduction to very broad base

User group evaluation of reductions

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Metadata Additions to Thesaurus

methodology

kind of data

universe

spatial unit

file structure, format, type

access conditions

age groups

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Translation of Thesaurus

Reduced monolingual thesaurus

Specialist teams with social science background

Backwards and sideways translations

Additions of new terms

Allow for different structures

Allow for non-equivalence

Extensive use of synonyms and scope notes

Other EU projects (PACO, CHINTEX etc)

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Thesaurus defined in XML and RDF

```
<thesTerm classCode="L10.10" type="pref">  
<termName xml:lang="en">ADDICTION</termName>  
<termName xml:lang="fr">DÉPENDANCE</termName>  
<scopeNote xml:lang="en" type="ambig">Use a more  
specific term if possible</scopeNote>  
<ufTerm fterm="L10.10NP"/>  
<ntTerm fterm="L10.10.10"/>  
<btTerm fterm="L10"/>  
<ttTerm fterm="L"/>  
<rtTerm fterm="R70.40"/>
```

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Limber Workshop - April 17-18th 2000

Alta-vista style searches

Free text / controlled vocabulary option

Logic of multi-word combination explained

Interface and hit list in chosen language

Hit list ranked by where word found

Customised prioritising of ranking

Search history to refine and combine searches

Other EU projects (e.g. PACO, EPAG etc)

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Search via classification code

Automatic use of thesaurus only for synonyms

**Complexity of thesaurus hidden unless
requested**

Simple alphabetic listings except for browsing

Searches performed first with suggestion list

**Failed searches employ stemming and
truncation**

Keywords are useful for searching and ranking

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

- Keywords are NOT so useful for relevance**
- Serious analysis requires translation**
- Title and summary in English**
- Controlled vocabularies on specific elements**
access conditions, methodology, kind of data, universe and spatial unit
- Languages of metadata displayed and selectable**
- Translation of DDI element headings**

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Great interest in Automatic Indexing Tool
Great scepticism as to if it would work
To work from DDI metadata
Use linguistic techniques to draw concepts out
Assign keywords that express those concepts
Learn from metadata with manual keywords
Other EU projects (RENARDUS) + OCLC

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing

Some doubts over RDF

Some doubts over DDI / NESSTAR

Some doubts over imbalances

Some doubts over multiple resources

Some doubts over mapping

IASSIST conference 7-10th June 2000

New doubts ?????

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University

Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



**The End -
A Possible
Sponsor and
Questions ???**

Mother Earth Herbals



Limber Up – Ache and Pain Salve

Here's help for all you weekend warriors! Apply this all natural healing rub to sore joints, muscles, tendons, ligaments or sprains. 2 oz. Jar.

Ingredients: Olive oil, St John's Wort, arnica, beeswax and essential oil.

Regular price: \$6.50 **Sale price: \$5.50**

[Order](#)

IASSIST 2000 - DATA IN THE DIGITAL LIBRARY:

Charting the Future for Social, Spatial and Government Data

June 7-10, Northwestern University



Limber Up: For the Race to Provide Multi-lingual Access and Automatic Indexing



Lafayette Ragsdale - Limber Up ! A Limited Edition Civil War Print