



# Databases of Women Scientists

Overview, Best Practice Guideline  
and Future Perspectives





# Databases of Women Scientists

Overview, Best Practice Guideline  
and Future Perspectives



#### LEGAL NOTICE:

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

The views expressed in this publication are the sole responsibility of the consortium partners of the project and do not necessarily reflect the views of the European Commission.

#### IMPRESSUM

##### Contact:

Center of Excellence Women and Science CEWS

University of Bonn

Poppelsdorfer Allee 15

D – 53115 Bonn

Phone: ++49 - (0)228 - 73 - 4835

Fax: ++49 - (0)228 - 73 - 4840

mailto: [cews-info@cews.uni-bonn.de](mailto:cews-info@cews.uni-bonn.de)

<http://www.cews.org>

##### Editorial staff :

Almuthe Schlueter

##### Design:

[wogo.de/sign](http://wogo.de/sign), Overath

##### Printing:

Warlich Druck Ahrweiler GMBH, Bad Neuenahr-Ahrweiler

This project was funded by the European Commission within the Sixth Framework Programme.

Bonn, February 2005

# Content

Foreword .....	7
Chapter 1 – Introduction .....	9
Chapter 2 – Methodology .....	13
Chapter 3 – Women Scientists in Europe – Databases and other Resources.....	17
<b>1 Databases in Europe.....</b>	<b>17</b>
1.1 Specific situation in the EU countries .....	18
1.2 Other clusters and database categories .....	19
1.3 Short summary .....	22
<b>2 Listing of selected databases .....</b>	<b>23</b>
2.1 Grids of analysis of the selected databases .....	24
2.2 Austria.....	26
2.3 Czech Republic .....	31
2.4 Denmark.....	32
2.5 Finland .....	34
2.6 Germany .....	35
2.7 Hungary .....	46
2.8 Iceland.....	47
2.9 Ireland.....	49
2.10 Malta .....	50
2.11 Netherlands .....	51
2.12 Norway.....	52
2.13 Sweden.....	53
2.14 Switzerland.....	54
2.15 United Kingdom.....	55
<b>3 List of selected organisations, institutions and networks .....</b>	<b>57</b>
3.1 Czech Republic .....	58
3.2 Denmark.....	58
3.3 Estonia.....	59
3.4 European Commission .....	59
3.5 European wide .....	60
3.6 France.....	61
3.7 Italy.....	63
3.8 Lithuania.....	63
3.9 Portugal.....	64
3.10 Romania.....	64

3.11	Slovakia.....	66
3.12	Spain.....	66

**Chapter 4 – Best Practice Guideline for Developing and Revising  
Women Scientists Databases ..... 67**

<b>1</b>	<b>Pre-Development or Pre-Revision Quality Criteria.....</b>	<b>68</b>
1.1	Institutional or organisational qualifications.....	68
1.2	Analysis of the context .....	68
1.3	Financial resources.....	69
<b>2</b>	<b>Data Collection Quality Criteria.....</b>	<b>70</b>
2.1	Objectives of a database.....	70
2.2	Target groups .....	70
2.3	Distinguishing women scientists databases from other data resources or collections.....	71
2.4	Online database, search and registration .....	71
2.5	Geographical and language coverage.....	72
2.6	Number of datasets.....	72
2.7	Regular updating of datasets.....	73
2.8	Evaluation of the database .....	73
2.9	Networking and Live Activities .....	73
2.10	Personal information saved in a database.....	73
<b>3</b>	<b>Technical Quality Criteria .....</b>	<b>77</b>
3.1	Multilinguality .....	77
3.2	Usability.....	78
3.3	Search.....	82
3.4	How to register in the database.....	82

**Chapter 5 – Future Perspectives for Linking Women  
Scientists Databases – A Feasibility Study ..... 89**

<b>1</b>	<b>Studying future perspectives .....</b>	<b>89</b>
1.1	Introduction.....	89
1.2	The process and the partners .....	89
1.3	The organization of the document .....	90
<b>2</b>	<b>The decision for a meta-database .....</b>	<b>92</b>
<b>3</b>	<b>Information Retrieval through a Meta-Database.....</b>	<b>93</b>
3.1	Understanding and Evaluating the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH).....	94
3.2	Relevance of the OAI-PMH for the DATAWOMSCI Project.....	96
3.3	Final Criteria for the selection of the OAI-PMH .....	97
<b>4</b>	<b>Describing the application domain for the feasibility study .....</b>	<b>98</b>
4.1	Picking the databases for the feasibility study .....	98

4.2	Adapting the Vocabulary of the OAI-PMH and Identifying the Actors.....	98
<b>5</b>	<b>Analysis and Description of the Feasibility Study Data Providers.....</b>	<b>100</b>
5.1	Objectives of the Databases .....	101
5.2	Existing Services and User Support Relevant to the DATAWOMSCI Service Provider.....	101
<b>6</b>	<b>The feasibility of the DATAWOMSCI Service Provider .....</b>	<b>105</b>
6.1	The content of the DATAWOMSCI Service Provider.....	105
6.2	Conclusions on the feasibility of the DATAWOMSCI Service Provider .....	112
<b>7</b>	<b>Analysis and the Evaluation of Cross-Lingual Information Retrieval for the DATAWOMSCI Service Provider.....</b>	<b>113</b>
7.1	Introduction to cross-lingual information retrieval.....	113
7.2	Cross-lingual information retrieval methods.....	114
<b>8</b>	<b>Introduction to International Web Sites .....</b>	<b>120</b>
8.1	Monolingual international site .....	120
8.2	Multilingual international site.....	122
<b>9</b>	<b>Three DATAWOMSCI Provider Solutions .....</b>	<b>123</b>
9.1	The multilingual federated solution: Multilingual DATAWOMSCI Service Provider (MultiDSP) .....	124
9.2	The monolingual federated solution: Monolingual DATAWOMSCI Service Provider (MonoDSP) .....	129
9.3	The distributed solution: the DATAWOMSCI Women Scientists Data Collections Information Provider (DIP).....	131
9.4	Developing a DATAWOMSCI Platform: A Service Provider, an Information Provider, or both.....	139
<b>10</b>	<b>Metadata for the future DATAWOMSCI Service Providers and Data Providers .....</b>	<b>141</b>
10.1	CERIF.....	141
10.2	The CERIF Data Model .....	144
<b>11</b>	<b>Concepts for Search Interfaces on a Meta-database .....</b>	<b>152</b>
11.1	Search and Navigation .....	153
11.2	Display of Search Results .....	158
<b>12</b>	<b>Conclusion .....</b>	<b>159</b>
	<b>References.....</b>	<b>161</b>
	<b>Chapter 6 – Follow up of the Project .....</b>	<b>165</b>



## Foreword

Databases of women scientists are an important strategy to promote women in science and to increase the percentage of women in top-positions in the scientific field. The disproportionate representation of women in science that we find in all countries of the EU makes it necessary to enhance gender equality. In order to get women more involved in the European Research Area databases of women scientists are effective instruments. Yet, the women scientists' database landscape has never been examined.

This gap is closed with this publication by offering an overview of existing databases of women scientists and a catalogue of high quality criteria. With a view to the European Research Area we need a closer cooperation between national databases. A feasibility study presented in this publication offers a possibility to link databases of women scientists in order to create a common basis and to make them more accessible nationally and internationally.

The publication owes its origin to the project "A study on **databases of women scientists**" (shortly named DATAWOMSCI Project) funded by the European Commission within the Sixth Framework Programme. Thanks to the results of this project today we know more about the status quo and the basic issues concerning databases of women scientists and their future perspectives in Europe. Such work is essential for governments, database maintainers and specialists, as well as for researchers and research organisations. This publication may be used as a guiding document for future policy development, funding schemes and other supporting activities related to data resources on women scientists in Europe.

One of the results of the project is the confirmation of the need to establish a common standard of quality for databases of women scientists. The project results published here provide an encouraging starting point for the adaptation of these quality standards.

In order to realise the meaning, potential and perspectives of databases of women scientists for the scientific area this publication can be a useful manual.

I am grateful to all persons who participated in this project, especially to all the work done by the Consortium Partners. Special thanks to the European Commission who made this project possible.

Bonn, February 2005

*Dr. Brigitte Mühlenbruch*

Managing Director  
Center of Excellence Women and Science CEWS



## Chapter 1 – Introduction

Networks and databases are an important strategy to promote women in science and to increase the percentage of women in top-positions in the scientific field. The significance of networks has been highlighted in a Communication by the EU-Commission "Women and Science: mobilising women to enrich European Research" (COM (1999) 76). In a first move, the Commission brought together some 150 representatives of networks, publications and website of women scientists in 1999, at a conference entitled "Networking the networks". In the concluding address the conference demanded that:

*"EU and Member States should support the development of expert databases of women scientists and commission research on the databases held by networks on women in science."*

While there are projects covering networks, an international cooperation and linking of mostly national databases is still missing. The promotion of women scientists is urgently needed and databases are an effective means to achieve this on a national and international level.

The Science and Society Unit of the European Commission wants to promote the effective use of scientific advice and seeks to encourage interaction between experts, civil society groups and policy-makers. For the best experts to be involved in these discussions, full knowledge of as many experts as possible is needed. At present we still find that many experts, especially women experts, are likely to be excluded from discussions due to missing women networks and missing information about women experts. Databases of women experts can help to close that information gap and as a result women can get more involved in European projects like the European Research Area (ERA).

Science and Society is an essential element of the ERA. The objectives of the ERA are to create an equivalent of the "common market" for goods and services for research and innovation. The mobility within the EU and an exchange of scientists (female and male) between member states is essential part of the ERA. Linked databases can offer support for these objectives. They are a means to insure trans-national access to research infrastructure that will help foster women's participation in the ERA.

Within the Women and Science Unit one of the main objectives of Science and Society is to raise women's participation in science and technological development. The disproportionate representation of women in science that we find in all countries of the EU hinders the full development of the potential of the European Research Area. It is not only a waste of human resources but also a distortion with regard to the relationship between science and society. This can be seen as one crucial point for a low trust of society in science. Building up gender equality can assure the value of science for a European society.

This is an ambitious political aim that relies on human resources both male and female. In order to achieve better representation of women in science and research greater mobilisation of women scientists is needed. Mobilisation requires knowledge

of vacancies (jobs, committees, panels) for women scientists and knowledge of women scientists ready to fill vacancies for those having to decide about staffing or compositions of panels and committees.

This knowledge can be collected in and be spread by databases of women scientists and those maintaining them. But not all databases are the same. They differ with respect to who can be listed (e.g. men and women; women only; scientists with a special degree or all experts, academics and non-academics), how much information on a scientist is available or who can attain the information listed, to name just a few possible differences. Without some standard of good quality databases cannot be used to full advantage because one might not find what one expects under a certain "name". Also, if databases exist on an national level but there is no knowledge of these databases internationally (e.g. because of language barriers) they are of reduced value for the EU. To build up the European Research Area (ERA) language barriers have to be overcome.

The motivation of the project "A study on **databases of women scientists**" (shortly named DATAWOMSCI Project) conforms with the goals mentioned above and embraces them through its project objectives. The objectives are:

- to analyse and evaluate the status quo on existing databases and other resources of women scientists in Europe
- to develop criteria of quality and formulate recommendations on developing and revising women scientists databases
- to undertake a feasibility study on future perspectives for linking women scientists databases

Six well-established partners from five countries who stand for profound gender knowledge with particular respect to women and science have come together in the DATAWOMSCI Project:

1. Center of Excellence Women and Science CEWS, University of Bonn, Germany  
(<http://www.cews.org>)
2. DIMEB, University of Bremen, Germany  
(<http://www.dimeb.de>)
3. KILDEN Information and Documentation Centre for Women's Studies and Gender Research in Norway, Oslo, Norway  
(<http://kilden.forskningradet.no>)
4. National Contact Centre - Women and Science (NCC-WS), Institute of Sociology, Czech Academy of Sciences, Prague, Czech Republic  
(<http://www.zenyaveda.cz>)
5. Observatoire des Sciences et des Techniques (OST), Paris, France  
(<http://www.obs-ost.fr>)
6. School of Business and Social Science, Roehampton University, London, United Kingdom  
(<http://www.roehampton.ac.uk/staff/JudithGlover/>)

The DATAWOMSCI Project is a Specific Support Action funded by the European Commission within the Sixth Framework Programme. The project started on the 1st April 2004 and ended on January 31, 2005.

In Chapter two of this publication the methodology of the DATAWOMSCI Project is described. Basic information on how existing databases of women scientists and other resources in Europe were taken stock of, the way the information was gathered, analysed, evaluated and presented in this publication is clarified.

The third Chapter is named "Women Scientists in Europe – Databases and other Resources" and contains an overview and classification of resources on European women scientists and an analysis of the status quo. With this Chapter we ensure that existing databases and knowledge about those databases are made accessible for the use of everyone involved in the European Research Area. It presents an instrument to show the potential of women scientists in Europe and can also be used to promote existing databases so that more women will be part of such databases. The internationalisation and collaboration of women scientists is urgently needed. Hence, this Chapter presents the basis to build up European-wide co-operations.

The fourth Chapter named "Best Practice Guideline for Developing and Revising Women Scientists Databases" is based on the results of Chapter three as well as some of the technical analyses which took place during the feasibility study (Chapter 5). Chapter four includes best practice criteria and recommendations for developing new databases as well as for maintaining and revising existing databases. It presents standards of quality databases have to meet. This will help speeding up the process of building up new databases and improving existing ones at a high standard of quality.

The guideline is also relevant for decision-makers: before using a database the consultation of the best practice guideline could make sure that the database about to be used for a search for highly qualified women scientists is likely to present good results because it does meet relevant criteria.

Chapter five is named "Future Perspectives for Linking Women Scientists Databases – A Feasibility Study" and builds upon the results of Chapter three (overview and classification of databases and other resources) and is closely connected to Chapter four (best practice guideline). It is a feasibility study on linking five selected best practice women scientists databases. The Chapter also includes evaluations of technologies for linking databases, methods for dealing with the complexities of multilingual international web sites, and three alternative models for linking women scientists databases.

Databases are an effective way of collecting information and make information accessible across Europe. This is especially true on the basis that the feasibility study shows that it is possible to link existing databases and also sets requirements for new databases to be linked in years to come. This create synergies between national and regional actions and policies.

Chapter six focuses on the "Follow up of the Project" and makes further recommendations for the enhancement of the European women scientists' database landscape on a European and national level.

This publication may on the one hand be used as a source of information on the status quo of databases and resources on women scientists. On the other hand the Chapters three through six may be used as guiding documents for future policy, funding and supporting activities related to data resources on women scientists in Europe. Finally, an overview of the fundamental issues concerning women scientists databases and their future perspectives in Europe is given.

## Chapter 2 – Methodology

For the first objective of analysing and evaluating the status quo on existing databases and other resources of women scientists in Europe the following steps were taken: preparation of a short manual for the identification of the databases; the identification, listing, definition and selection of existing data collections; and the development of a questionnaire sent out to a subset of the institutions and organisations maintaining data collections.

Concerning the identification of databases and other resources the following persons, groups or objects are of core interest to our analysis:

- women scientists and researchers,
- women experts and
- databases.

Before concentrating on these elements it was necessary to clarify the landscape in which the project is embedded. This requires an understanding of the different concepts of existing databases and data resources to be found in the EU countries, i.e.

- databases and data collections of women scientists with a university degree,
- databases and data collections of women experts with work experience and specialist knowledge,
- databases and data collections of scientists for both sexes,
- network databases, membership-databases and data collections of (women) scientists in specific fields,
- databases and data collections of scientific projects and/or research projects.

Bibliographical data collections or databases for literature mentioning authors or other women scientists or women experts were excluded.

The databases and data collections as listed above are overlapping as shown in the following diagram:

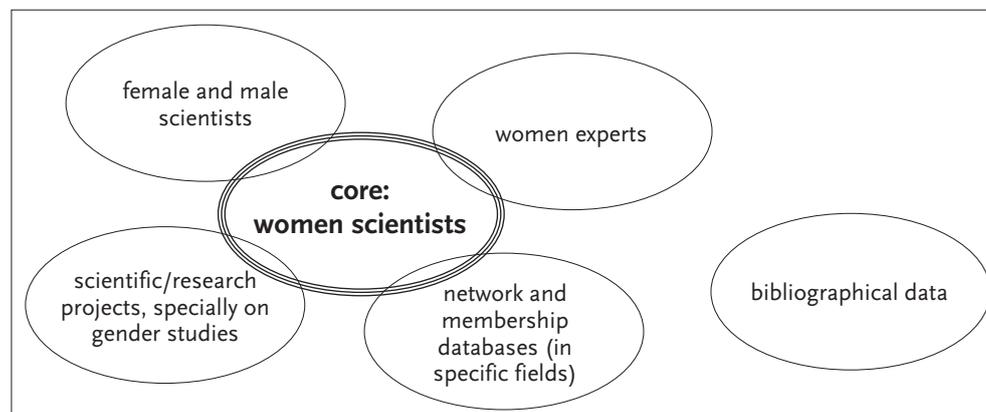


Diagram showing the intersection of different kinds of databases and data collections

The DATAWOMSCI Project consisted of first taking stock of the situation in each European country according to a large definition of databases/data collections/lists on women scientists.

The databases were identified via internet and personal contacts. Existing documents, reports or directories on organisations related to women scientists and gender issues were also used for this purpose, such as the "Directory of Networks" of the European Commission, ETAN national reports, ENWISE report.

The first step of the identification process led to a huge number of databases/data collections, and lists mentioning women scientists/researchers or experts. Hence, the consortium partners took stock of the situation in the EU countries and decided on a restricted definition of databases to be selected for further in-depth analysis and evaluation.

For this, the notion of databases, the term "expert" and the objectives of the databases had to be clarified. Basically, database means structured data, use of software, functionalities and a formal structure. Data lists may be a simple piece of paper. Data collection is the general term including databases and data lists. Data resources are a synonym for data collections.

Experts may be defined by qualification only, by experience only, or by qualification plus experience. Concerning databases it is important to know who the experts are, how they are registered as such (self designation), who the end users are, what expertise consists of, and what the links are between academic degree, qualification and expertise.

In addition, the objectives of the databases was a major criteria for selection, e.g. promoting women in science, networking, role models etc.

The consortium partners came up with the following definition of databases for the in-depth analysis:

*"Data resources promoting women in science (broadly defined) or helping to place women in key positions at universities, in research centres, in committees, in industrial research, in commissions and on monitoring panels. These data resources should offer contact data of women scientists/experts (online and/or offline) and information concerning their scientific field and work experience. Information should be organised into a database."*

According to the above definition, the following data collections have been excluded from the in-depth analysis:

- listings of scientific personnel in academic institutions,
- membership-data collections,
- mailing-lists,
- project and research report data collections.

This exclusion was well-founded on the basis that every academic institution and research organisation would have data/database on scientific personnel for administration purposes mainly.

We also excluded:

- data collections on scientists and experts which did not display information on field/working area, qualification or did not allow search by sex whenever scientists of both sexes were registered,
- network-databases or membership-databases simply consisting of a mailing-list without further information (name, scientific field, interest, work experience etc.).

A questionnaire for an in-depth analysis was elaborated and sent to the selected database maintainers to gather information on the structure of the database, datasets, technical software aspects, maintenance, updating, registration conditions and users facilities. The results of the questionnaire and the definition of the databases were used for a further in-depth analysis of data collections around Europe.

The evaluation of the returned questionnaires allowed for a description of relevant databases of women scientists in Europe (see Chapter 3, Section 1). A characterisation of the selected databases presented as synthetic grids of analysis are placed before a listing of the selected databases by countries in alphabetical order (see Chapter 3, Section 2) and may be used for handling the list of databases/data collections more comfortable.

Last but not least a list of organisations, institutions and networks, which may serve as contact point to women scientists is displayed in Chapter 3, Section 3. This list applies to countries where it was not possible to find a database.

**For the second objective** of listing criteria of quality and formulating recommendations on developing and revising women scientists databases the steps were: an evaluation of the questionnaires sent out to organisations and institutions maintaining data collections, preparation of usability reports on selected best practice databases, selection of best practice criteria from these reports and analyses. Finally, each of the consortium partners with experience in women scientists data collections brought in their experiences and thus revised and enriched the catalogue of quality criteria.

Based on the analysis and evaluation of women scientists databases we were able to write a „Best Practice Guideline for Developing and Revising Women Scientists Databases“ (Chapter four). This guideline completes the overview of the databases by offering criteria of quality and recommendations for developing new databases or revising already existing ones.

**For the third objective** of undertaking a feasibility study on future perspectives for linking women scientists databases the steps were: the selection of best practice databases as objects of the feasibility study, the development and communication of a feasibility questionnaire, and research on current technologies on meta-databases. The information communicated through the feasibility questionnaire and the research done provided the basis for the evaluation of alternative models for a future meta-database of women scientists.

In this fifth Chapter named "Future Perspectives for Linking Women Scientists Databases – A Feasibility Study" the results of Chapter three and Chapter four are utilized in order to improve information retrieval, strengthen cooperation and increase the visibility of women scientists databases and the institutions maintaining these.

Finally, Chapter 6 focuses on the "Follow up of the Project" and makes further recommendations for the enhancement of the European women scientists' database landscape on a European and national level. These recommendations are based on the results of the previous Chapters.

All Chapters together give an overview of the fundamental issues concerning women scientists databases and their future perspectives in Europe.

The consortium partner DIMEB, University of Bremen was in charge of Chapter 5. All other Chapters were produced by all consortium partners under the coordination of the Center of Excellence Women and Science CEWS. Hence, the content, suggestions and recommendations of the Chapters are results of close exchange and intensive discussion between the project partners and reflect the profound knowledge and experiences of all partners.

## Chapter 3 – Women Scientists in Europe – Databases and other Resources

The following overview “Women Scientists in Europe – Databases and other Resources” closes a gap by collecting information on existing databases of women scientists, clarifying what kinds of databases already exist and how they can be distinguished from other data collections. The study aims to draw a line between networks and network databases and actual databases of women scientists.

Objectives of the databases in question in this project should be to support universities, women’s representatives and equal opportunities officers as well as scientific and political institutions in their search for qualified women scientists and women experts. Such support may be utilized to increase the proportion of women in leadership positions at universities and research institutions, on panels and committees.

All objectives are based on the different functions databases have, e.g. making visible the existing potential of women scientists, offering a targeted search for new members of boards and monitoring panels by the EU, scientific organisations, ministries, universities or research centres or supporting the building of networks by enabling women scientists to find project partners, especially at an international level.

The following results of the DATAWOMSCI Project are presented:

1. an evaluation of the collected questionnaires from selected databases including general information about the specific situation in different parts of Europe and other clusters and database categories,
2. a listing of existing databases relevant to this investigation presented by countries in alphabetical order and their name (in English) and
3. a list of organisations, institutions and networks from those European countries which do not have a database or data collection in accordance with the definition used in this project. This list may be used to make out contact with women scientists and experts.

With this overview existing databases are taken stock of. The information is meant for members of the European Research Area, namely users of databases, women scientists, gender experts, National ministries, the European Commission, the future European Platform of Women Scientists, national and international networks and the maintainers of databases.

### 1 Databases in Europe

The analysis and evaluation of 31 questionnaires showed interesting details about existing databases/data collections in Europe. As a result we present different clusters, e.g. different kind of databases and different categories.

## 1.1 Specific situation in the EU countries

According to our definition of databases, the situation in the European countries is very contrasted: in some countries there exists at least one relevant database, in some other countries databases have been very recently set up while in other countries no database could be found at all.

A short description of the situation in the European countries concerning the existence of relevant data collections is presented below.

### 1.1.1 *Situation in German speaking countries: Germany, Austria, Switzerland*

Most databases/data collections relevant to this investigation were found in German speaking countries. The situation in Germany is exceptional due to the great number of existing databases/data collections. One third of all relevant databases/data collections are located there.

Altogether more than 50 % of all databases/data collections listed in this investigation are located in Germany, Austria or Switzerland. In each one of these countries there is at least one major and interdisciplinary database with online access and similar functions (FemConsult, Germany; FEMtech, Austria; femdat, Switzerland). FEMtech is a newly founded database fulfilling all criteria to become the major interdisciplinary online database for women scientists/experts in Austria in the future. Apart from the above mentioned three databases all other data collections are much smaller and generally not interdisciplinary. In the case of some tiny databases/data collections it seems that they are set up only for legitimating purpose of an organisation or institution.

### 1.1.2 *Situation in the Northern European Countries, the United Kingdom and Ireland: Denmark, Sweden, Norway, Iceland and United Kingdom, Ireland*

In each of these countries could be found at least one database/data collection (altogether 10 databases were identified). With one exception (Women's Engineering Society Membership Database, UK) these databases are national databases. In addition, it is remarkable that 2 of 5 offline data resources in whole Europe are located in the United Kingdom and Ireland.

### 1.1.3 *Situation in Eastern European countries/CEE countries: Baltic countries, Bulgaria, Czech Republic, Hungary, Romania, Slovakia, Slovenia*

Three databases have been identified as relevant to this study. One is located in Hungary ('Tárki Hungarian Gender Databank') and two in the Czech Republic ('Database of women scientists' and the database of the Central European Centre for Women and Youth in Science (CEC-WYS), both situated at the National Contact Centre – Women and Science in Prague). The database of CEC-WYS is a very recent (October 2004) database covering the Czech Republic, Hungary, Slovakia and Slovenia and replaced the Czech database ('Database of women scientists').

In these countries there are many organisations and institutions which serve as a first contact point for women scientists/experts such as:

- organisations and information centres for women scientists,
- institutions located at universities covering gender studies/gender research offering further information according to their scientific field,
- non-governmental, non-profit organisations dealing with gender equality, human rights, scientific issues.

#### *1.1.4 Situation in continental Western European countries: France, Luxemburg, Belgium, the Netherlands*

One database was located in the Netherlands (Experts for women studies and gender research).

No databases fitting our definition could be found in France, Luxemburg (where a university was established recently) and Belgium. Nevertheless, in these countries there are many women associations and networks, very active in promoting women scientists on the regional or national level, but the situation is similar to the Southern European countries: no database.

#### *1.1.5 Situation in Southern European countries and Israel: Cyprus, Greece, Italy, Malta, Spain, Portugal and Israel*

In these countries, the identification of relevant databases was very difficult due to the fact that national language websites were provided (no English versions). In many cases existing websites were out dated, and contact address did not answer.

Therefore, due to a lack of information, it was not possible to identify existing databases/data collections from countries like Israel, Cyprus, Greece, Italy, Spain and Portugal. Some women associations and networks are active in these countries but do not have a database (apart from membership lists or mailing lists).

In Malta, one database is used at a governmental level only (Directory of Maltese Women, National Commission for Gender Equality).

This short presentation of the situation in the different parts of Europe shows that we could not identify relevant databases in countries such as Belgium, Cyprus, France, Greece, Israel, Italy, Luxemburg, Portugal and Spain.

The "Best Practice Guideline for Developing and Revising Women Scientists Databases" which is to be found in Chapter 4 of this publication will be very helpful for institutions in these countries if they decide to develop or revise databases in the future.

## **1.2 Other clusters and database categories**

The following description gives an overview of the most relevant characteristics collected in the questionnaires.

### *1.2.1 Objectives of the databases/data collections*

The main objectives of most selected databases are 'promoting women in science and research' and 'networking in science and research' (24 of 31 databases/data collections). 19 databases/data collections mention 'consulting and expertise' and 'staffing of key-positions in science and research' as a further objective. Therefore, there seems to be a difference between 'promoting women in science and research' in general and 'staffing of key positions in science and research' as a concrete measure to promote women in science and research.

In addition, only 5 databases mentioned 'collaboration in R&D projects' as an objective. The reason for this may lie in the structure most of these databases which are not suitable for searching after collaborations; moreover, such contacts are usually made in a more informal way by the scientific community. Another reason may be a lack of recognition of European and international dimension in this sector. If the databases/data collections were linked with each other and better known, they would be useful tools to favour 'collaboration in R&D projects'.

### *1.2.2 Registered persons*

In all databases/data collections scientists and researchers (e.g. scientists/researchers with a university degree; from all disciplines and not just natural sciences) are registered and 55 % of all databases/data collections also contain experts. This means that academics are definitely registered in all selected data resources but less than half of them contain data of ONLY academics (see above in Chapter 2 the difference between scientists/researchers and experts according to the definition made in this study).

8 data collections contain also other groups of persons besides scientists/researchers and experts.

One exceptional data collection (Professorships in women and gender studies (ProfFG)) lists professorships for women and gender research instead of persons but a link leads to the website of the professorship and to the name and contact address of the professor.

### *1.2.3 Target groups*

In accordance with the persons being mostly registered (scientists/researchers) and the main objectives (networking and promoting women in science and research) more than 80% of the data collections are intended for 'universities and higher education institutions' and 'individuals in universities, research institutions, business enterprises or political organisations'.

'Research institutions', 'business enterprises', 'political institutions', 'women's representatives and equal opportunity officers' and 'women scientists and researchers' are also mentioned as target groups but in a smaller proportion.

#### 1.2.4 *Fields of sciences*

There are 13 interdisciplinary databases/data collections and 18 specialist data resources. The 18 specialist data resources can be divided into two specific groups. The first group consists of 10 databases/data collections specialised in women and gender studies/gender research. The second group consists of 8 databases: one database is limited to the thematic area of migration but covering all fields of sciences. The other databases/data collections cover specific professional groups: 4 are for scientists/experts in the sector of science, engineering and technology (SET) and 3 are for professional groups like historians, political scientists and artists.

#### 1.2.5 *Personal information about the registered scientists/experts*

Concerning the quality of the personal data on registered persons, nearly all databases/data collections (97 %) display information on name, address, level of education (highest academic degree), academic position, competence/specialisation/expertise and field of activity (e.g. university, research centre, business enterprise, politics etc.). 70 % of all databases/data collections save also information about projects, publications and work experience. Grants and fellowships are rarely asked for. Usually enough information is saved to present good profiles about the registered persons in general.

A special problem is the possibility to search by sex. There are two kinds of data collections to handle the 'search by sex-problem'. The first category consists of databases/data collections being restricted to women only (altogether 19 data resources).

The second category consists of 12 databases being open to both sexes. 8 of these 12 databases have more than 90 % women and one database has only 50 % women. But nevertheless in only 2 cases a search by sex is possible!

This fact leads to the conclusion that the smaller the percentage of females in a database/data collection without the possibility to search by sex the smaller the possibility that this database/data collection may be used as an instrument towards gender equality. Because in order to use databases/data collections as an instrument towards gender equality search by sex must be possible.

#### 1.2.6 *Geographical limitation*

Another important aspect of the evaluation of the questionnaires is the geographical limitation of the data resources. In fact, 23 of 31 data resources are geographically limited mainly to the country of the database maintainer.

8 data resources are without geographical limitations and 5 databases are *de facto* geographically limited by offering the data only in a national language. Some databases/data collections offer information in English, and are more open geographically and therefore more internationally available. From this point of view only 3 of the 8 data resources are geographically open (Women's Engineering Society Membership Database, UK; femdok, Sweden; FEMtech, Austria; femdat, Switzerland) but are not really international. They usually provide information on scientists/experts working in the country the database is located in. This is the case of the two databases from the German speaking countries called FEMtech (Austria) and femdat (Switzerland). Despite

the fact that they are the major interdisciplinary online databases in the country they are located in, the percentage of scientists/experts from other German speaking countries registered in these databases is rather low. This underlines the impression that the geographical borders between countries are also borderlines for data resources. This situation becomes more obvious if the data resources of the German speaking countries are compared between each other (e.g. FemConsult (Germany) vs. FEMtech (Austria) vs. femdat (Switzerland) or Database for Gender Experts (GISA; Germany) vs. Database of scientists in women and gender studies (Austria)).

#### *1.2.7 Technical background: databases/data collections and online access*

Regarding the technical background 27 of 31 data resources are using database software. The online display of the data collections is very satisfactory but searching and registration functionalities are usually limited: 25 data collections are available online for searching, but only 18 data collections offer online registration.

For some data collections this is due to the fact that the data are collected and updated independently from the scientists/experts who may not even be aware of being mentioned in the database (e.g. ProFFG or Women research professors/Women's Research Network NRW in Germany).

#### *1.2.8 Direct contact with the registered persons*

Online searching does not necessarily imply direct contact with the registered scientists/ experts. 25 data resources offer direct contact to the scientists/experts, 6 databases keep the personal information of registered individuals anonymous (and therefore the organisations maintaining the database act as go-between for institutions/organisations/individuals searching for women scientists/experts), 2 databases/data collections offer both possibilities and one offers no contact data at all (only names; Icelandic women with doctoral degree). But it is necessary to note that handling of the contact data is not depending on the database/data collection being an actual database (using database software) or if being online available.

### **1.3 Short summary**

The analysis and evaluation of the questionnaires showed that the main characteristics of the 31 selected databases/data collections are as follows:

- objectives:
  - promoting women in science and research
  - networking in science and research
- registered persons:
  - scientists and researchers, from all disciplines (not just natural sciences)
- target groups:
  - universities and higher education institutions
  - research institutions

- individuals in universities, research institutions, business enterprises or political organisations
- interdisciplinarity
- personal information of the registered persons saved in the data collection:
  - name
  - address
  - level of education (highest academic degree)
  - academic position
  - competence/specialisation/expertise
  - field of activity (e.g. university, research centre, business enterprise, politics etc.)
- for women only
- geographically limited
- only a national language version
- real databases
- online for searching, but only slightly over half of the data resources offer online registration
- offer direct contact to the scientists/experts

In summary, this analysis and evaluation of the database questionnaires has two benefits. On the one hand, the analysis shows the status quo of the existing databases and gives an impression of the existing database landscape. A comparison between the databases is possible so that the complexity and the advantages and disadvantages of certain database functions (e.g. direct or indirect contact to the registered persons) also become obvious.

On the other hand, the evaluation shows that none of the existing database is the perfect one. Some databases listed in this Chapter and in the feasibility study (Chapter three) combine important criteria of quality of women scientists databases. Yet, all existing databases show that there is space for improvement.

Based on this study we were able to present a „Best Practice Guideline for Developing and Revising Women Scientists Databases“ (see Chapter 4). This guideline completes the analysis of Chapter three by offering criteria of quality and recommendations for developing new databases or revising already existing databases.

## 2 Listing of selected databases

Databases listed in this Section by countries in alphabetical order and their name (in English) are those which fit our definition of databases as mentioned in Chapter two. Basically, this list asserts no claim on completeness. It is based on information available via internet, personal contacts and existing documents, reports or directories on organisations related to women scientists and gender issues at the point of time the data were taken stock of (2004).

## 2.1 Grids of analysis of the selected databases

The following two grids of analysis summarise the main characteristics of the 31 selected databases. In addition, it may be used an index-list of the separate database listing by referring to the page number on which the database is listed in this Section.

Name of database	Country	Interdisciplinary	Specialised in women and gender studies	covering specific professional groups	Women only	Search by sex possible	National	Online	With database software	Page No.
Database of scientists in women and gender studies	Austria		x			x		x	x	26
Database of women with a habilitation in Austria	Austria	x			x		x		x	27
Expert Database of Lower Austria	Austria	x			x		x	x	x	28
FEMtech Female Expert Database	Austria	x			x			x	x	29
fiftitu-database - female artists	Austria			x	x		x	x	x	30
Database of the Central European Centre for Women and Youth in Science (CEC-WYS)	Czech Republic	x			x		x	x	x	31
Researcher database, Gender in science-gender research	Denmark						x	x	x	32
Women online	Denmark	x	x		x		x	x	x	33
Minna - the database for women's studies and equality experts	Finland		x				x	x	x	34
Database - More female professors in universities of applied sciences	Germany	x			x			x	x	35
Database for Gender Experts	Germany		x				x	x	x	36
Database of female academics	Germany	x			x				x	37
Database of scientists of the working group research in women's history	Germany			x			x	x	x	38
Expert database Migration	Germany			x			x	x	x	39
FemConsult - Database of Women Scientists	Germany	x			x			x	x	40
femina politica - database of women experts	Germany			x	x			x	x	41
Kompetenzz-database	Germany			x			x	x	x	42
Non-professorial academic staff in the Women's Research Network NRW	Germany		x		x		x	x		43
Professorships in women and gender research (ProffG)	Germany		x				x	x		44
Women research professors, Women's Research Network NRW	Germany		x		x		x	x		45
Tárki Hungarian Gender Databank, Hungary	Hungary		x				x	x	x	46
Icelandic women with doctoral degree	Iceland	x			x		x	x		47
The Icelandic Women Database	Iceland	x			x		x	x	x	48
WITS Talent Bank	Ireland			x	x		x		x	49
Directory of Maltese Women, National Commission for Gender Equality	Malta	x			x		x		x	50
Experts for women studies and gender research (DVG)	Netherlands		x				x	x	x	51
KILDEN's researcher database - sub base for women researchers	Norway	x			x		x	x	x	52
femdok - database for women's and gender studies and equality research	Sweden		x				x	x	x	53
femdat - Swiss Database of Women Experts	Switzerland	x			x			x	x	54
European Database of Women Experts in SET	United Kingdom			x	x		x	x	x	55
Women's Engineering Society Membership Database	United Kingdom			x		x			x	56

Name of database	Country	Registration in database			Objectives of database				
		Academics only	SEtR with university degree	Experts	Promoting Women	Networking	Consulting & expertise	Staffing of key positions	Collaboration in R&D projects
Database of scientists in women and gender studies	Austria	x	x		x	x	x		
Database of women with a habilitation in Austria	Austria	x	x				x	x	
Expert Database of Lower Austria	Austria		x	x	x		x	x	
FEMtech Female Expert Database	Austria		x	x	x	x	x	x	
fiftitu-database - female artists	Austria		x	x	x	x		x	
Database of the Central European Centre for Women and Youth in Science (CEC-WYS)	Czech Republic	x	x		x	x	x	x	x
Researcher database, Gender in science-gender research	Denmark	x	x			x	x	x	
Women online	Denmark		x	x	x		x	x	
Minna - the database for women's studies and equality experts	Finland		x	x	x		x		
Database - More female professors in universities of applied sciences	Germany	x	x		x	x		x	
Database for Gender Experts	Germany		x	x	x	x	x		
Database of female academics	Germany	x	x		x			x	
Database of scientists of the working group research in women's history	Germany	x	x		x	x			
Expert database Migration	Germany		x	x					
FemConsult - Database of Women Scientists	Germany	x	x		x	x	x	x	x
femina politica - database of women experts	Germany		x	x	x	x	x	x	x
Kompetenzz-database	Germany		x	x	x	x			
Non-professorial academic staff in the Women's Research Network NRW	Germany	x	x			x			
Professorships in women and gender research (ProffG)	Germany	x	x		x				
Women research professors, Women's Research Network NRW	Germany	x	x			x			
Tárki Hungarian Gender Databank	Hungary		x	x	x	x			
Icelandic women with doctoral degree	Iceland	x	x		x				
The Icelandic Women Database	Iceland		x	x	x	x	x	x	x
WITS Talent Bank	Ireland		x	x	x	x	x	x	
Directory of Maltese Women, National Commission for Gender Equality	Malta		x	x		x	x	x	
Experts for women studies and gender research (DVG)	Netherlands		x	x	x	x	x		
KILDEN's researcher database - sub base for women researchers	Norway	x	x		x	x	x	x	
femdok - database for women's and gender studies and equality research	Sweden	x	x			x	x		
femdat - Swiss Database of Women Experts	Switzerland		x	x	x	x		x	
European Database of Women Experts in SET	United Kingdom		x	x	x	x	x	x	x
Women's Engineering Society Membership Database	United Kingdom		x	x	x	x	x	x	

## 2.2 Austria

<i>Name of database:</i>	Database of scientists in women and gender studies WissenschaftlerInnen-Datenbank Frauen- und Geschlechterforschung
<i>Address:</i>	Koordinationsstelle für Geschlechterstudien, Frauenforschung und Frauenförderung Universität Graz Beethovenstr. 19 A-8010 Graz
<i>Country:</i>	Austria
<i>Contact person:</i>	Dr. Barbara Hey
<i>Phone:</i>	++43-(0)316-380-5722
<i>E-Mail:</i>	koordff@uni-graz.at
<i>Website of maintainer of database:</i>	<a href="http://www.kfunigraz.ac.at/kffwww/">http://www.kfunigraz.ac.at/kffwww/</a>
<i>Website of database:</i>	<a href="http://www.mavas.at/members/genderstudies/">http://www.mavas.at/members/genderstudies/</a>
<i>Foundation:</i>	1995
<i>Number of datasets:</i>	721
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	scientists/researchers
<i>Sex of registered persons:</i>	both sexes (95% women)
<i>Main objectives of the database:</i>	<ul style="list-style-type: none"> <li>- promoting women in science and research</li> <li>- staffing of boards and committees</li> <li>- networking in science and research</li> <li>- consulting/expertise</li> </ul>
<i>Target groups of the database:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- political institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this database:</i>	gender research
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	weekly

<i>Name of database:</i>	Database of women with a habilitation in Austria Datenbank habilitierter Frauen in Österreich
<i>Address:</i>	Arbeitskreis für Gleichbehandlungsfragen der Karl-Franzens-Universität Graz Harrachgasse 34 A-8010 Graz
<i>Country:</i>	Austria
<i>Contact person:</i>	Mag. Ulrike Schustaczek; Karin Schranz
<i>Phone:</i>	++43-(0)316-380-1025
<i>E-Mail:</i>	akgl@uni-graz.at
<i>Website of maintainer of database:</i>	<a href="http://www.kfunigraz.ac.at/akglwww/">http://www.kfunigraz.ac.at/akglwww/</a>
<i>Website of database:</i>	-
<i>Foundation:</i>	1996
<i>Number of datasets:</i>	450
<i>Geographical coverage of database:</i>	Austria
<i>Content/navigation language:</i>	German; English
<i>Who is registered in the database:</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- staffing of key - positions in science and research - staffing of boards and committees - consulting/expertise
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	offline
<i>Regular updating of database:</i>	continuously

<i>Name of database:</i>	Expert Database of Lower Austria Niederösterreichische Expertinnen-Datenbank NED II – NÖ Expertinnenpool
<i>Address:</i>	NÖ Landesakademie Neue Herrengasse 17A A-3109 St.Pölten
<i>Country:</i>	Austria
<i>Contact person:</i>	Michael Urban
<i>Phone:</i>	++43-(0)2742-294-17405
<i>E-Mail:</i>	michael.urban@noe-lak.at
<i>Website of maintainer of database:</i>	www.ned.noe-lak.at
<i>Website of database:</i>	www.ned.noe-lak.at -> NED II – NÖ Expertinnenpool
<i>Foundation:</i>	1998
<i>Number of datasets:</i>	143
<i>Geographical coverage of database:</i>	Austria
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - role models - consulting/expertise
<i>Target groups of the database:</i>	- business enterprises - individuals in universities, research institutions, business enterprises or political organisation - other sectors: health; education-development-youth; communication
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	yes

<i>Name of database:</i>	FEMtech Female Expert Database FEMtech Expertinnendatenbank
<i>Address:</i>	ÖGUT – Austrian Society for Environment and Technology Hollandstr. 10/46 A-1020 Vienna  Koordinierungsstelle FEMtech TIG Technologie Impulse Gesellschaft Grillparzer Straße 7 A-1010 Vienna
<i>Country:</i>	Austria
<i>Contact person:</i>	Inge Schrattenecker (ÖGUT) or Dorothea Sturn (FEMtech)
<i>Phone:</i>	++43-1-3156393-12 (ÖGUT) or ++43-1-5132627 (FEMtech)
<i>E-Mail:</i>	inge-schrattenecker@oegut.at or office@femtech.at
<i>Website of maintainer of database:</i>	www.oegut.at
<i>Website of database:</i>	www.femtech.at -> Expertinnendatenbank
<i>Foundation:</i>	May 2004
<i>Number of datasets:</i>	150
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German; English
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - networking in science and research - consulting/expertise - raising the proportion of women in jury meetings and research projects
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women scientists and researchers
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	continuously

<i>Name of database:</i>	fiftitu-database – female artists fiftitu – Künstlerinnen-Datenbank
<i>Address:</i>	FIFTITU – Vernetzungsstelle für Frauen in Kunst und Kultur in OÖ Kapuzinerstr. 36/1 A-4020 Linz
<i>Country:</i>	Austria
<i>Contact person:</i>	Manuela Mittermayer
<i>Phone:</i>	++43-(0)732-770353
<i>E-Mail:</i>	fiftitu@servus.at or database@fiftitu.at
<i>Website of maintainer of database:</i>	www.fiftitu.at
<i>Website of database:</i>	www.fiftitu.at/kunstdb/
<i>Foundation:</i>	October 2003
<i>Number of datasets:</i>	40
<i>Geographical coverage of database:</i>	Austria; Germany (so far)
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - experts - female artists or female cultural workers with a special knowledge
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - networking in science and research - promoting female artists and females who work in the area of cultural organisation
<i>Target groups of the database:</i>	- universities/higher education institutions - individuals in universities, research institutions, business enterprises or political organisations - women scientists and researchers - cultural organisations (managers etc.)
<i>Fields of science covered by this database:</i>	- humanities, arts and services - gender research
<i>Specification of database/particular field covered by this database:</i>	artists, cultural workers, architects, speakers
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	no

## 2.3 Czech Republic

<i>Name of database:</i>	Database of the Central European Centre for Women and Youth in Science (CEC-WYS)
<i>Address:</i>	Institute of Sociology, Academy of Sciences CR National Contact Centre - Women and Science Jilska 1 CZ-110 00, Prague 1
<i>Country:</i>	Czech Republic
<i>Contact person:</i>	Marcela Linkova
<i>Phone:</i>	+420-222-222-322
<i>E-Mail:</i>	marcela@zenyaveda.cz
<i>Website of maintainer of database:</i>	<a href="http://www.cec-wys.org/html/">http://www.cec-wys.org/html/</a>
<i>Website of database:</i>	<a href="http://www.cec-wys.org/html/index.php?s1=1&amp;t2=7&amp;t3=2&amp;lng=13">http://www.cec-wys.org/html/index.php?s1=1&amp;t2=7&amp;t3=2&amp;lng=13</a>
<i>Foundation:</i>	September 2004
<i>Number of datasets:</i>	101
<i>Geographical coverage of database:</i>	Czech Republic, Hungary, Slovakia, Slovenia
<i>Content/navigation language:</i>	English
<i>Who is registered in the database:</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	<ul style="list-style-type: none"> <li>- promoting women in science and research</li> <li>- staffing of key-positions in science and research</li> <li>- staffing of boards and committees</li> <li>- networking in science and research</li> <li>- consulting/expertise</li> <li>- collaboration in EU R&amp;D-projects</li> </ul>
<i>Target groups of the database:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- business enterprises political institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women's representatives and equal opportunity officers</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	every six months

## 2.4 Denmark

<i>Name of data collection:</i>	Researcher database, Gender in science–gender research Forskerdatabase, Køn i forskning – forskning i køn
<i>Address:</i>	Co-ordination for Gender Studies Department of Sociology, University of Copenhagen DK- 1361 Copenhagen N
<i>Country:</i>	Denmark
<i>Contact person:</i>	-
<i>Phone:</i>	+45-3532-3501
<i>E-Mail:</i>	kkf@sociology.ku.dk
<i>Website of maintainer of data collection:</i>	<a href="http://www.sociology.ku.dk/sochrc/index.htm">http://www.sociology.ku.dk/sochrc/index.htm</a>
<i>Website of data collection:</i>	<a href="http://www.sociology.ku.dk/sochrc/Database/gmd.asp">http://www.sociology.ku.dk/sochrc/Database/gmd.asp</a>
<i>Foundation:</i>	2000
<i>Number of datasets:</i>	280
<i>Geographical coverage of data collection:</i>	Denmark
<i>Content/navigation language:</i>	Danish
<i>Who is registered in the data collection:</i>	scientists/researchers
<i>Sex of registered persons:</i>	both sexes (95% women)
<i>Main objectives of the data collection:</i>	- staffing of key-positions in science and research - networking in science and research - consulting/expertise
<i>Target groups of the data collection:</i>	- universities/higher education institutions - research institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - university students
<i>Fields of science covered by this data collection:</i>	- natural sciences - life sciences - health and care - education and training - humanities, arts and services - social science, business and law - gender research
<i>Specification of database/particular field covered by this data collection:</i>	gender in science, gender research
<i>Access to database for registration in the data collection:</i>	-
<i>Access to database for searching in the data collection:</i>	-
<i>Regular updating of data collection:</i>	once a year

<i>Name of database:</i>	Women online Kvinder på Linjen
<i>Address:</i>	KVINFO Christians Brygge 3 DK-1219 Copenhagen K
<i>Country:</i>	Denmark
<i>Contact person:</i>	Line Holst, editor
<i>Phone:</i>	+45-33 13 50 88
<i>E-Mail:</i>	Ekspert@kvinfo.dk
<i>Website of maintainer of database:</i>	<a href="http://www.kvinfo.dk">http://www.kvinfo.dk</a>
<i>Website of database:</i>	<a href="http://www.kvinfo.dk/side/383">http://www.kvinfo.dk/side/383</a>
<i>Foundation:</i>	1995
<i>Number of datasets:</i>	1.400
<i>Geographical coverage of database:</i>	Denmark
<i>Content/navigation language:</i>	Danish
<i>Who is registered in the database:</i>	- scientists/researchers - experts - artists
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - consulting/expertise
<i>Target groups of the database:</i>	- business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - the media - the general public
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	from 2004 the aim is every 6 months

## 2.5 Finland

<i>Name of database:</i>	Minna – the database for women's studies and equality experts
<i>Address:</i>	Christina Institute for Women's Studies P.O.Box 59 (Unioninkatu 38 E) FIN-00014 University of Helsinki, Finland
<i>Country:</i>	Finland
<i>Contact person:</i>	Hanna Lehto
<i>Phone:</i>	+358-9-191-24324
<i>E-Mail:</i>	hanna.x.lehto@helsinki.fi
<i>Website of maintainer of database:</i>	
<i>Website of database:</i>	<a href="http://www.minna.fi/manna/Tietokanta/index.htm">http://www.minna.fi/manna/Tietokanta/index.htm</a>
<i>Foundation:</i>	December 2004
<i>Number of datasets:</i>	-
<i>Geographical coverage of database:</i>	Finland
<i>Content/navigation language:</i>	Finnish
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	both sexes (80% women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - consulting/expertise
<i>Target groups of the database:</i>	- individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	women's studies and gender equality
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	-

## 2.6 Germany

<i>Name of database:</i>	Database – More female professors in universities of applied sciences Datenbank FH-Professorin
<i>Address:</i>	Landeskonzferenz der Frauenbeauftragten an Fachhochschulen in Baden-Württemberg c/o Fachhochschule Nürtingen / Standort Geislingen Postfach 1251 D-73302 Geislingen/Steige
<i>Country:</i>	Germany
<i>Contact person:</i>	Prof. Dr. Margot Körber-Weik
<i>Phone:</i>	++49-(0)7331-22485
<i>E-Mail:</i>	lakoffhbw@asg.fh-nuertingen.de
<i>Website of maintainer of database:</i>	www.gleichstellung-fh-bw.de
<i>Website of database:</i>	www.gleichstellung-fh-bw.de -> Datenbank
<i>Foundation:</i>	2002
<i>Number of datasets:</i>	1.000
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - female professionals interested in professorships, working outside universities; universities as employers; employers outside universities interested in female professionals
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women towards a professorship at universities of applied sciences, especially in Germany, Austria and Switzerland - promoting women in science and research - staffing of key-positions in science and research - mentoring - networking in science and research
<i>Target groups of the database:</i>	- universities/higher education institutions - business enterprises - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	approx. annually

<i>Name of database:</i>	Database for Gender Experts Gender Mainstreaming-Information-System ExpertInnen-Datenbank des Gender-Institut Sachsen-Anhalt (GISA)
<i>Address:</i>	Gender-Institut Sachsen-Anhalt (GISA) Ebendorfer Str. 3 D-39108 Magdeburg
<i>Country:</i>	Germany
<i>Contact person:</i>	Ute Wanzek or Thomas Claus
<i>Phone:</i>	++49-(0)391-5066577 or ++49-(0)391-5066566
<i>E-Mail:</i>	Ute.Wanzek@g-i-s-a.de or Thomas.Claus@g-i-s-a.de
<i>Website of maintainer of database:</i>	<a href="http://www.g-i-s-a.de">http://www.g-i-s-a.de</a>
<i>Website of database:</i>	<a href="http://www.g-i-s-a.de/content/rubrik/153">http://www.g-i-s-a.de/content/rubrik/153</a> -> ExpertInnen
<i>Foundation:</i>	March 2004
<i>Number of datasets:</i>	40
<i>Geographical coverage of database:</i>	Germany
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - experts - consultants, trainers, coaches for Gender Mainstreaming
<i>Sex of registered persons:</i>	both sexes (79% women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - mentoring - networking in science and research - consulting/expertise - Gender and Gender Mainstreaming consultants, training, research
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers - all institutions, which need expert knowledge
<i>Fields of science covered by this database:</i>	- all disciplines - Economics, Employment, Youth/Sport/Culture, Budgeting
<i>Specification of database/particular field covered by this database:</i>	- gender (gender equality issues and gender mainstreaming implementation)
<i>Access to database for registration in the database:</i>	online; offline
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	quarterly

<i>Name of database:</i>	Database of female academics Datenbank für Akademikerinnen, Frauenpunkt COURAGE e.V.
<i>Address:</i>	Frauenpunkt COURAGE e.V. Gehrenseestr. 4 D-13053 Berlin
<i>Country:</i>	Germany
<i>Contact person:</i>	Jutta Cujass or Dr. Beate Eibl
<i>Phone:</i>	++49-(0)30-97896935 or ++49-(0)30-98315513
<i>E-Mail:</i>	fbs.courage@t-online.de
<i>Website of maintainer of database:</i>	-
<i>Website of database:</i>	-
<i>Foundation:</i>	1994
<i>Number of datasets:</i>	120-200
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - women's representatives and equal opportunity officers
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	offline
<i>Regular updating of database:</i>	sequentially

<i>Name of database:</i>	Database of scientists of the working group research in women's history Datenbank WissenschaftlerInnen des Arbeitskreises Historische Frauen- und Geschlechterforschung (AKHFG)
<i>Address:</i>	Universität Flensburg Institut für Geschichte und ihre Didaktik Auf dem Campus 1 D-24943 Flensburg
<i>Country:</i>	Germany
<i>Contact person:</i>	Prof. Dr. Bea Lundt
<i>Phone:</i>	++49-(0)461-8052802
<i>E-Mail:</i>	lundt@uni-flensburg.de
<i>Website of maintainer of database:</i>	<a href="http://www.uni-flensburg.de/akhfg">http://www.uni-flensburg.de/akhfg</a>
<i>Website of database:</i>	<a href="http://www.uni-flensburg.de/akhfg">http://www.uni-flensburg.de/akhfg</a> -> Zum Inhalt -> Datenbank WissenschaftlerInnen
<i>Foundation:</i>	since 2002 online
<i>Number of datasets:</i>	260
<i>Geographical coverage of database:</i>	Germany (national part of the International Federation for Research in Women's History-IFRWH)
<i>Content/navigation language:</i>	German
<i>Who is registered in the database :</i>	scientists/researchers
<i>Sex of registered persons:</i>	both sexes (90% women)
<i>Main objectives of the database:</i>	<ul style="list-style-type: none"> <li>- promoting women in science and research</li> <li>- networking in science and research</li> <li>- national group in the International Federation for Research in Women's History (IFRWH)</li> </ul>
<i>Target groups of the database:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this database:</i>	<ul style="list-style-type: none"> <li>- humanities, arts and services</li> <li>- gender research</li> <li>- research in women's history</li> </ul>
<i>Specification of database/particular field covered by this database:</i>	gender history
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	continuously

<i>Name of database:</i>	Expert database Migration Expertendatenbank Migration
<i>Address:</i>	Bundeszentrale für politische Bildung Adenauerallee 86 D-53113 Bonn
<i>Country:</i>	Germany
<i>Contact person:</i>	-
<i>Phone:</i>	++49-(0)1888 -515-0
<i>E-Mail:</i>	info@bpb.de
<i>Website of maintainer of database:</i>	<a href="http://www.bpb.de">http://www.bpb.de</a>
<i>Website of database:</i>	<a href="http://www.bpb.de/wissen/E8LOR9,0,0,Migration.html">http://www.bpb.de/wissen/E8LOR9,0,0,Migration.html</a>
<i>Foundation:</i>	October 2003
<i>Number of datasets:</i>	250
<i>Geographical coverage of database:</i>	Germany, Austria, Switzerland
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	both sexes (50% women)
<i>Main objectives of the database:</i>	- information about experts in migration
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - individuals in universities, research institutions, business enterprises or political organisations
<i>Fields of science covered by this database:</i>	- Life sciences - Education and training - Humanities, arts and services - Social science, business and law - Gender research
<i>Specification of database/particular field covered by this database:</i>	Migration; Germany as immigration society; living together cross-cultural
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	every six months

<i>Name of database:</i>	FemConsult – Database of Women Scientists Wissenschaftlerinnen-Datenbank FemConsult
<i>Address:</i>	Center of Excellence Women and Science (CEWS) University of Bonn Poppelsdorfer Allee 15 D-53115 Bonn
<i>Country:</i>	Germany
<i>Contact person:</i>	Almuthe Schlueter
<i>Phone:</i>	++49-(0)228-73-4838
<i>E-Mail:</i>	femconsult@cews.uni-bonn.de
<i>Website of maintainer of database:</i>	<a href="http://www.cews.org">http://www.cews.org</a>
<i>Website of database:</i>	<a href="http://www.femconsult.de">http://www.femconsult.de</a>
<i>Foundation:</i>	1994
<i>Number of datasets:</i>	6.000
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German
<i>Who is registered in the database :</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	<ul style="list-style-type: none"> <li>- promoting women in science and research</li> <li>- staffing of key-positions in science and research</li> <li>- staffing of boards and committees</li> <li>- networking in science and research</li> <li>- networking in science and research</li> <li>- consulting/expertise</li> <li>- Gender and Gender Mainstreaming consultants, training, research</li> </ul>
<i>Target groups of the database:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- business enterprises</li> <li>- political institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women's representatives and equal opportunity officers</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	weekly

<b>Name of database:</b>	<b>femina politica – database of women experts Politologinnen-Datenbank</b>
<b>Address:</b>	femina politica Zeitschrift für feministische Politikwissenschaft c/o FU-Berlin Otto-Suhr-Institut Innestr. 22 D-14165 Berlin
<b>Country:</b>	Germany
<b>Contact person:</b>	-
<b>Phone:</b>	++49-(0)30-838-53829
<b>E-Mail:</b>	poldat@gmx.de
<b>Website of maintainer of database:</b>	<a href="http://www.femina-politica.de">http://www.femina-politica.de</a>
<b>Website of database:</b>	<a href="http://www.politologin.de">http://www.politologin.de</a>
<b>Foundation:</b>	2001
<b>Number of datasets:</b>	94
<b>Geographical coverage of database:</b>	international
<b>Content/navigation language:</b>	German
<b>Who is registered in the database:</b>	- scientists/researchers - experts
<b>Sex of registered persons:</b>	women only
<b>Main objectives of the database:</b>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - networking in science and research - consulting/expertise - collaboration in EU R&D-projects
<b>Target groups of the database:</b>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers
<b>Fields of science covered by this database:</b>	- Social science, business and law - Gender research
<b>Specification of database/particular field covered by this database:</b>	Political sciences and related issues
<b>Access to database for registration in the database:</b>	online
<b>Access to database for searching in the database:</b>	online
<b>Regular updating of database:</b>	no

<i>Name of database:</i>	<b>Kompetenzz-database Kompetenzz-Datenbank</b>
<i>Address:</i>	Frauen geben Technik neue Impulse e.V. Fachhochschule Bielefeld Wilhelm-Bertelsmann-Straße 10 D-33602 Bielefeld
<i>Country:</i>	Germany
<i>Contact person:</i>	Araya Körner
<i>Phone:</i>	++49-(0)521-106-7323
<i>E-Mail:</i>	info@kompetenzz.de
<i>Website of maintainer of database:</i>	http://www.kompetenzz.de or http://www.frauen-technik-impulse.de
<i>Website of database:</i>	http://www.kompetenzz.de/netzwerke/kompetenzz_datenbank
<i>Foundation:</i>	1998
<i>Number of datasets:</i>	1.514
<i>Geographical coverage of database:</i>	Germany
<i>Content/navigation language:</i>	German
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	both sexes (unknown, but most of the registered persons are women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - mentoring - role models - networking in science and research
<i>Target groups of the database:</i>	- individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers
<i>Fields of science covered by this database:</i>	- natural sciences - engineering, manufacturing and construction - education and training - gender research
<i>Specification of database/particular field covered by this database:</i>	women in information society and technology
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	no

<i>Name of data collection:</i>	Non-professional academic staff in the Women's Research Network Northrhine-Westfalia Wissenschaftliche Mitarbeiterinnen der Frauen- und Geschlechterforschung des Netzwerkes Frauenforschung NRW
<i>Address:</i>	Netzwerk Frauenforschung NRW Universität Dortmund D-44221 Dortmund
<i>Country:</i>	Germany
<i>Contact person:</i>	Prof. Dr. Ruth Becker; Dr. Beate Kortendiek
<i>Phone:</i>	++49-(0)231-755-5142
<i>E-Mail:</i>	kortendiek@netzwerk-frauenforschung.de
<i>Website of maintainer of data collection:</i>	www.netzwerk-frauenforschung.de
<i>Website of data collection:</i>	www.netzwerk-frauenforschung.de -> Mittelbau
<i>Foundation:</i>	1999
<i>Number of datasets:</i>	75-100
<i>Geographical coverage of data collection:</i>	Northrhine-Westfalia (NRW), Germany
<i>Content/navigation language:</i>	German; English; French
<i>Who is registered in the data collection:</i>	scientists/researchers
<i>Sex of registered persons:</i>	female only
<i>Main objectives of the data collection:</i>	networking in science and research
<i>Target groups of the data collection:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women's representatives and equal opportunity officers</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this data collection:</i>	all disciplines
<i>Specification of database/particular field covered by this data collection:</i>	women and gender studies
<i>Access to database for registration in the data collection:</i>	offline
<i>Access to database for searching in the data collection:</i>	online
<i>Regular updating of data collection:</i>	twice per year

<i>Name of data collection:</i>	Professorships in women and gender research Professuren für Frauen- und Geschlechterforschung (ProfFG)
<i>Address:</i>	Freie Universität Berlin Zentraleinrichtung zur Förderung von Frauen- und Geschlechterforschung Habelschwerdter Allee 45 D-14195 Berlin
<i>Country:</i>	Germany
<i>Contact person:</i>	Dr. Ulla Bock
<i>Phone:</i>	++49-(0)30/838-53378 or ++49-(0)30/838-56254
<i>E-Mail:</i>	bocku@zedat.fu-berlin.de
<i>Website of maintainer of data collection:</i>	<a href="http://www.fu-berlin.de/zefrauen/ze/ze_wir.html">http://www.fu-berlin.de/zefrauen/ze/ze_wir.html</a>
<i>Website of data collection:</i>	<a href="http://www.fu-berlin.de/zefrauen/doku/doku_prof.html">www.fu-berlin.de/zefrauen/doku/doku_prof.html</a>
<i>Foundation:</i>	1992
<i>Number of datasets:</i>	approx. 120
<i>Geographical coverage of data collection:</i>	Germany
<i>Content/navigation language:</i>	German
<i>Who is registered in the data collection:</i>	scientists/researchers
<i>Sex of registered persons:</i>	both sexes (98% women)
<i>Main objectives of the data collection:</i>	- promoting women - promoting women in science and research
<i>Target groups of the database:</i>	- universities/higher education institutions - women scientists and researchers
<i>Fields of science covered by this data collection:</i>	all disciplines
<i>Specification of database/particular field covered by this data collection:</i>	professorships of women and gender research
<i>Access to database for registration in the data collection:</i>	-
<i>Access to database for searching in the data collection:</i>	online
<i>Regular updating of data collection:</i>	minimum 4 times a year

<i>Name of data collection:</i>	Women research professors, Women's Research Network Northrhine–Westfalia Frauenforschungsprofessuren des Netzwerkes Frauenforschung NRW
<i>Address:</i>	Netzwerk Frauenforschung NRW Universität Dortmund D-44221 Dortmund
<i>Country:</i>	Germany
<i>Contact person:</i>	Prof. Dr. Ruth Becker; Dr. Beate Kortendiek
<i>Phone:</i>	++49-(0)231-755-5142
<i>E-Mail:</i>	kortendiek@netzwerk-frauenforschung.de
<i>Website of maintainer of data collection:</i>	www.netzwerk-frauenforschung.de
<i>Website of data collection:</i>	www.netzwerk-frauenforschung.de -> Professorinnen
<i>Foundation:</i>	1999
<i>Number of datasets:</i>	60
<i>Geographical coverage of data collection:</i>	Northrhine–Westfalia (NRW), Germany
<i>Content/navigation language:</i>	German; English; French
<i>Who is registered in the data collection:</i>	women scientists and researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the data collection:</i>	- networking in science and research - female professors (women and gender studies) either full or associated who work at NRW universities
<i>Target groups of the data collection:</i>	- universities/higher education institutions - research institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers
<i>Fields of science covered by this data collection:</i>	all disciplines
<i>Specification of database/particular field covered by this data collection:</i>	women and gender studies
<i>Access to database for registration in the data collection:</i>	offline
<i>Access to database for searching in the data collection:</i>	online
<i>Regular updating of data collection:</i>	twice per year

## 2.7 Hungary

<i>Name of database:</i>	Tárki Hungarian Gender Databank
<i>Address:</i>	Budaörsi út 45H-1112 Budapest
<i>Country:</i>	Hungary
<i>Contact person:</i>	Nagy Ildiko
<i>Phone:</i>	++36-1-3097676
<i>E-Mail:</i>	nagyildi@tarki.hu
<i>Website of maintainer of database:</i>	<a href="http://www.tarki.hu">http://www.tarki.hu</a>
<i>Website of database:</i>	<a href="http://www.tarki.hu/adatbank-h/nok/kutatasok/">http://www.tarki.hu/adatbank-h/nok/kutatasok/</a>
<i>Foundation:</i>	1999
<i>Number of datasets:</i>	57
<i>Geographical coverage of database:</i>	Hungary
<i>Content/navigation language:</i>	Hungarian; English
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	both sexes (88-90% women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - role models - networking in science and research
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - individuals in universities, research institutions, business enterprises or political organisations - women scientists and researchers
<i>Fields of science covered by this database:</i>	- life sciences - health and care - education and training - social science, business and law - gender research
<i>Specification of database/particular field covered by this database:</i>	sociology
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	online; offline
<i>Regular updating of database:</i>	yearly (also depending on the budget)

## 2.8 Iceland

<i>Name of data collection:</i>	Icelandic women with doctoral degree Íslenskir kvendoktorar
<i>Address:</i>	Women's History Archives (Kvennasögusafn Íslands) Arngrímogata 3 IS-107 Reykjavík
<i>Country:</i>	Iceland
<i>Contact person:</i>	Auður Styrkársdóttir
<i>Phone:</i>	+354-525-5779
<i>E-Mail:</i>	kona@bok.hi.is
<i>Website of maintainer of data collection:</i>	www.kona.bok.hi.is
<i>Website of data collection:</i>	www.kona.bok.hi.is/Kvendoktorar/kvendoktorar.html
<i>Foundation:</i>	2002
<i>Number of datasets:</i>	344
<i>Geographical coverage of data collection:</i>	Iceland
<i>Content/navigation language:</i>	Icelandic
<i>Who is registered in the data collection:</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the data collection:</i>	<ul style="list-style-type: none"> <li>- promoting women</li> <li>- promoting women in science and research</li> <li>- role models</li> </ul>
<i>Target groups of the data collection:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institution</li> <li>- research institutions</li> <li>- business enterprises</li> <li>- political institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women's representatives and equal opportunity officers</li> <li>- women scientists and researchers</li> </ul>
<i>Fields of science covered by this data collection:</i>	all disciplines
<i>Specification of database/particular field covered by this data collection:</i>	no
<i>Access to database for registration in the data collection:</i>	offline
<i>Access to database for searching in the data collection:</i>	online
<i>Regular updating of data collection:</i>	once per month

<i>Name of database:</i>	<b>The Icelandic Women Database</b> <b>Kvennaslóðir</b>
<i>Address:</i>	Centre for Women's and Gender Studies University of Iceland SudurgatalS-101 Reykjavik
<i>Country:</i>	Iceland
<i>Contact person:</i>	Erla Hulda Halldorsdottir
<i>Phone:</i>	+354-525-4634
<i>E-Mail:</i>	rikk@hi.is
<i>Website of database:</i>	<a href="http://www.rikk.hi.is">http://www.rikk.hi.is</a>
<i>Website of maintainer of database:</i>	<a href="http://www.kvennaslodir.is">http://www.kvennaslodir.is</a>
<i>Foundation:</i>	October 2003
<i>Number of datasets:</i>	550
<i>Geographical coverage of database:</i>	Iceland (Icelandic experts in other countries can register, also experts of foreign origins living in Iceland).
<i>Content/navigation language:</i>	Icelandic is the navigation language; Keywords are also in English
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - role models - networking in science and research - consulting/expertise - membership records - collaboration in EU R&D-projects
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers - Media seeking experts for interviews
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	regularly

## 2.9 Ireland

<i>Name of database:</i>	<b>WITS Talent Bank</b>
<i>Address:</i>	WITS (Women in Technology and Science) PO Box 3783 Dublin 4, Ireland
<i>Country:</i>	Ireland
<i>Contact person:</i>	Elisabeth Creed
<i>Phone:</i>	+353-(0)21-4706012
<i>E-Mail:</i>	wits@iol.ie
<i>Website of maintainer of database:</i>	www.witsireland.com
<i>Website of database:</i>	www.witsireland.com
<i>Foundation:</i>	1998 (previous database); new directory will be available August 2004
<i>Number of datasets:</i>	152
<i>Geographical coverage of database:</i>	Ireland
<i>Content/navigation language:</i>	English
<i>Who is registered in the database:</i>	- scientists/researchers - experts - women who do not have specific SET background but have expertise in the sector e.g. policy, business development
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - role models - networking in science and research - consulting/expertise
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers
<i>Fields of science covered by this database:</i>	- natural sciences - life sciences - engineering, manufacturing and construction - health and care - education and training - agriculture and veterinary - social science, business and law - gender research
<i>Specification of database/particular field covered by this database:</i>	criteria for inclusion is to have expertise in the SET sector
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	offline
<i>Regular updating of database:</i>	no

## 2.10 Malta

<i>Name of database:</i>	Directory of Maltese Women, National Commission for Gender Equality
<i>Address:</i>	Executive Director National Commission for Promotion of Equality 2 Cavalier Street Valletta, Malta
<i>Country:</i>	Malta
<i>Contact person:</i>	Ms Sina Bugeja
<i>Phone:</i>	+356-(0)25903851
<i>E-Mail:</i>	genderequality@gov.mt
<i>Website of maintainer of database:</i>	<a href="http://mfss.gov.mt/services/subpages/content.asp?id=1595">http://mfss.gov.mt/services/subpages/content.asp?id=1595</a>
<i>Website of database:</i>	<a href="http://mfss.gov.mt/services/subpages/content.asp?id=371">http://mfss.gov.mt/services/subpages/content.asp?id=371</a>
<i>Foundation:</i>	1999
<i>Number of datasets:</i>	400
<i>Geographical coverage of database:</i>	Malta
<i>Content/navigation language:</i>	English
<i>Who is registered in the database:</i>	- scientists/researchers - experts - those with any area of expertise
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women - staffing of key-positions in science and research - staffing of boards and committees - networking in science and research - consulting/expertise
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	offline
<i>Regular updating of database:</i>	no

## 2.11 Netherlands

<i>Name of database:</i>	Experts for women studies and gender research Deskundigen Vrouwenstudies en Genderonderzoek (DVG)
<i>Address:</i>	Obiplein 4 NL-1094 RB Amsterdam
<i>Country:</i>	Netherlands
<i>Contact person:</i>	Evelien Rijsbosch
<i>Phone:</i>	++31-(0)20-6651318
<i>E-Mail:</i>	secr@iiav.nl
<i>Website of maintainer of database:</i>	www.iiav.nl
<i>Website of database:</i>	<a href="http://www.iiav.nl/nl/databases/deskundigen/index.html">http://www.iiav.nl/nl/databases/deskundigen/index.html</a>
<i>Foundation:</i>	2000
<i>Number of datasets:</i>	350
<i>Geographical coverage of database:</i>	The Netherlands
<i>Content/navigation language:</i>	Dutch
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	both sexes (95% women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - role models - networking in science and research - consulting/expertise - listing of gender experts and their research areas
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers - media
<i>Fields of science covered by this database:</i>	Gender research
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	every other year

## 2.12 Norway

<i>Name of database:</i>	KILDEN's researcher database, sub base for women researchers Kildens forskerbase, delbase for kvinnelige forskere
<i>Address:</i>	KILDEN - Norwegian Information and Documentation Centre for Women's Studies and Gender Research Grensen 5 NO-0159 Oslo
<i>Country:</i>	Norway
<i>Contact person:</i>	Hanne Størset
<i>Phone:</i>	++47-24 05 59 94
<i>E-Mail:</i>	post@forskerbasen.no
<i>Website of maintainer of database:</i>	<a href="http://kilden.forskningsradet.no">http://kilden.forskningsradet.no</a>
<i>Website of database:</i>	<a href="http://forskerbasen.no">http://forskerbasen.no</a>
<i>Foundation:</i>	1998
<i>Number of datasets:</i>	1.861
<i>Geographical coverage of database:</i>	Norway
<i>Content/navigation language:</i>	Norwegian
<i>Who is registered in the database:</i>	scientists/researchers
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	<ul style="list-style-type: none"> <li>- promoting women in science and research</li> <li>- staffing of key-positions in science and research</li> <li>- staffing of boards and committees</li> <li>- networking in science and research</li> <li>- consulting/expertise</li> <li>- Media seeking researchers for interviews</li> </ul>
<i>Target groups of the database:</i>	<ul style="list-style-type: none"> <li>- universities/higher education institutions</li> <li>- research institutions</li> <li>- individuals in universities, research institutions, business enterprises or political organisations</li> <li>- women's representatives and equal opportunity officers</li> <li>- women scientists and researchers</li> <li>- media, students and the general public</li> </ul>
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	depends on the individual researcher

## 2.13 Sweden

<i>Name of data collection:</i>	femdok – database for women's and gender studies and equality research femdok – databas för kvinno- genus- och jämställdhetsforskning
<i>Address:</i>	Centre for Gender Studies Lund University P. O. Box 117 SE-221 00 Lund
<i>Country:</i>	Sweden
<i>Contact person:</i>	Catarina Carlsson
<i>Phone:</i>	++46-46-222-9778
<i>E-Mail:</i>	genus@genus.lu.se
<i>Website of maintainer of data collection:</i>	<a href="http://www.genus.lu.se/genus.en.htm">http://www.genus.lu.se/genus.en.htm</a>
<i>Website of data collection:</i>	<a href="http://www.genus.lu.se/femdok">http://www.genus.lu.se/femdok</a>
<i>Foundation:</i>	1990
<i>Number of datasets:</i>	500
<i>Geographical coverage of data collection:</i>	Sweden
<i>Content/navigation language:</i>	Swedish; English; French
<i>Who is registered in the data collection:</i>	scientists/researchers
<i>Sex of registered persons:</i>	both sexes (90% women)
<i>Main objectives of the data collection:</i>	- networking in science and research - consulting/expertise
<i>Target groups of the data collection:</i>	- universities/higher education institutions - research institutions - individuals in universities, research institutions, business enterprises or political organisations
<i>Fields of science covered by this data collection:</i>	Gender research
<i>Specification of database/particular field covered by this data collection:</i>	no
<i>Access to database for registration in the data collection:</i>	online
<i>Access to database for searching in the data collection:</i>	online
<i>Regular updating of data collection:</i>	-

## 2.14 Switzerland

<i>Name of database:</i>	femdat – Swiss Database of Women Experts
<i>Address:</i>	femdat Office Gesellschaftsstrasse 25 CH-3012 Berne
<i>Country:</i>	Switzerland
<i>Contact person:</i>	Maria Iannino Gerber
<i>Phone:</i>	0041-31-6313701
<i>E-Mail:</i>	femdat@afg.unibe.ch
<i>Website of maintainer of database:</i>	<a href="http://www.femdat.ch">http://www.femdat.ch</a>
<i>Website of database:</i>	<a href="http://www.femdat.ch">http://www.femdat.ch</a>
<i>Foundation:</i>	October 2001
<i>Number of datasets:</i>	1.500
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	German; English; French; Italian
<i>Who is registered in the database:</i>	- scientists/researchers - experts
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - networking in science and research
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers - media
<i>Fields of science covered by this database:</i>	all disciplines
<i>Specification of database/particular field covered by this database:</i>	no
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	twice a year

## 2.15 United Kingdom

<i>Name of database:</i>	European Database of Women Experts in Science, Engineering and Technology (SET)
<i>Address:</i>	UK Resource Centre for Women in SET (UKRC) Faculty of Development and Society Sheffield Hallam University Sheffield, UK
<i>Country:</i>	United Kingdom
<i>Contact person:</i>	Pat Turrell
<i>Phone:</i>	++44-(0)114-225-4695
<i>E-Mail:</i>	p.m.turrell@shu.ac.uk
<i>Website of maintainer of database:</i>	www.oakland.co.uk
<i>Website of database:</i>	www.setwomenexperts.org.uk
<i>Foundation:</i>	1997
<i>Number of datasets:</i>	598
<i>Geographical coverage of database:</i>	Sweden, UK, Netherlands, Spain, Germany
<i>Content/navigation language:</i>	English
<i>Who is registered in the database:</i>	- scientists/researchers - experts - public speakers in SET
<i>Sex of registered persons:</i>	women only
<i>Main objectives of the database:</i>	- promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - role models - networking in science and research - consulting/expertise - collaboration in EU R&D-projects
<i>Target groups of the database:</i>	- universities/higher education institutions - research institutions - business enterprises - political institutions - individuals in universities, research institutions, business enterprises or political organisations - women's representatives and equal opportunity officers - women scientists and researchers - European Commission
<i>Fields of science covered by this database:</i>	- natural sciences - life sciences - engineering, manufacturing and construction - health and care - agriculture and veterinary - gender research
<i>Specification of database/particular field covered by this database:</i>	science, engineering and technology (SET)
<i>Access to database for registration in the database:</i>	online
<i>Access to database for searching in the database:</i>	online
<i>Regular updating of database:</i>	Last funding ended in 2002 but new government funding starting in 2004 will ensure the updating of the database.

<b>Name of database:</b>	<b>Women's Engineering Society Membership Database</b>
<i>Address:</i>	Women's Engineering Society 22 Old Queen Street London SW1H 9HP 0207 233 1974, UK
<i>Country:</i>	United Kingdom
<i>Contact person:</i>	Pam Wain
<i>Phone:</i>	++44-(0)207-233-1974
<i>E-Mail:</i>	info@wes.org.uk
<i>Website of maintainer of database:</i>	www.wes.org.uk
<i>Website of database:</i>	www.wes.org.uk
<i>Foundation:</i>	1919
<i>Number of datasets:</i>	650
<i>Geographical coverage of database:</i>	international
<i>Content/navigation language:</i>	English
<i>Who is registered in the database:</i>	- scientists/researchers - experts - students
<i>Sex of registered persons:</i>	both sexes (99% women)
<i>Main objectives of the database:</i>	- promoting women - promoting women in science and research - staffing of key-positions in science and research - staffing of boards and committees - mentoring - role models - networking in science and research - consulting/expertise - membership records
<i>Target groups of the database:</i>	- women scientists and researchers - people interested in women and engineering
<i>Fields of science covered by this database:</i>	- engineering, manufacturing and construction
<i>Specification of database/particular field covered by this database:</i>	engineering
<i>Access to database for registration in the database:</i>	offline
<i>Access to database for searching in the database:</i>	offline
<i>Regular updating of database:</i>	weekly

### 3 List of selected organisations, institutions and networks

The organisations, institutions and networks which are listed here do not have databases/data collections in accordance with the definition used in this project. This list may be used for finding out contact with women scientists/experts. This applies to countries where it was not possible to find a database, which does not necessarily mean that there exists none. Basically, this list asserts no claim on completeness. Additional information may be found in other documents, reports or directories on organisations related to women scientists and gender issues, such as the "Directory of Networks" of the European Commission, ETAN national reports or ENWISE report.

#### Czech Republic:

- National Contact Center – Women and Science (NCC-WS),  
Institute of Sociology, Czech Academy of Sciences

#### Denmark:

- Database for women in physics

#### Estonia:

- Estonian Women's Studies and Recourse Center (ENUT)

#### European Commission:

- fp6 experts database

#### European wide:

- WITEC – European Association for Women in Science, Engineering and Technology
- IFUW – International Federation of University Women

#### France:

- Association femmes et mathématiques
- Association Femmes et Sciences
- Association française des Femmes Ingénieurs
- Association of female academics (AFFDU)

#### Italy:

- Associazione Donne e Scienza

#### Lithuania:

- LAUW – Lithuanian Association of University Women

#### Portugal:

- Association of women in science

#### Romania:

- National Association of Women Physicians in Romania and Romanian Association of University Women
- National Committee for Promoting Women in Science in Romania
- Women in Science and Technology Association

#### Slovakia:

- The Professional Women's Association

#### Spain:

- AMIT – Asociación de Mujeres Investigadoras y Tecnólogas

### 3.1 Czech Republic

<i>Name of institution/organisation/network:</i>	National Contact Center – Women and Science, Institute of Sociology, Czech Academy of Sciences Národní kontaktní centrum – Ženy a věda, Sociologický ústav, AV ČR
<i>Address:</i>	Husova 7 CZ-110 00 Prague 1
<i>Country:</i>	Czech Republic
<i>Contact person:</i>	Marcela Linkova
<i>Phone:</i>	++420-222 222 322
<i>E-Mail:</i>	info@zenyaveda.cz
<i>Website of the institution/organisation/network:</i>	www.zenyaveda.cz
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Czech Republic
<i>Content/navigation language:</i>	Czech; English
<i>Specific remarks :</i>	The centre is focusing on promoting and supporting women in science.

### 3.2 Denmark

<i>Name of institution/organisation/network:</i>	Database for women in physics Faglig database over kvinder i fysikk
<i>Address:</i>	Netværk for Kvinder i Fysik
<i>Country:</i>	Denmark
<i>Contact person:</i>	Tina Christensen
<i>Phone:</i>	-
<i>E-Mail:</i>	tic@dmu.dk
<i>Website of the institution/organisation/network:</i>	<a href="http://www.nbi.dk/kif/">http://www.nbi.dk/kif/</a>
<i>Website of the data collection:</i>	<a href="http://www.nbi.dk/kif/database.html">http://www.nbi.dk/kif/database.html</a>
<i>Geographical coverage:</i>	Denmark
<i>Content/navigation language:</i>	Dansk
<i>Specific remarks :</i>	Network having a tiny database of women physicists in order to promote women in physics.

### 3.3 Estonia

<i>Name of institution/organisation/network:</i>	Estonian Women's Studies and Recource Center (ENUT) Eesti Naisuurimus- ja Teabekeskus
<i>Address:</i>	Estonian Women's Studies and Recource Center (ENUT) Tallinn Pedagogical University Narva mnt. 25-410 10120 Tallinn, Estonia
<i>Country:</i>	Estonia
<i>Contact person:</i>	Marju Järve
<i>Phone:</i>	++372-6409-173
<i>E-Mail:</i>	marju.jarve@enut.ee, enut@enut.ee
<i>Website of the institution/organisation/network:</i>	<a href="http://www.enut.ee/enut.php?keel=ENG&amp;tid=16">http://www.enut.ee/enut.php?keel=ENG&amp;tid=16</a>
<i>Website of the data collection:</i>	<a href="http://www.enut.ee/enut.php?keel=ENG&amp;tid=7">http://www.enut.ee/enut.php?keel=ENG&amp;tid=7</a>
<i>Geographical coverage:</i>	Estonia
<i>Content/navigation language:</i>	Estonian; English
<i>Specific remarks :</i>	-

### 3.4 European Commission

<i>Name of institution/organisation/network:</i>	fp6 experts database
<i>Address:</i>	CORDIS Help Desk B.P. 2373 L-1023 Luxemburg
<i>Country:</i>	Luxemburg
<i>Contact person:</i>	-
<i>Phone:</i>	++352-26 64 801
<i>E-Mail:</i>	helpdesk@cordis.lu
<i>Website of the institution/organisation/network:</i>	<a href="http://www.cordis.lu/experts/fp6_candidature.htm">http://www.cordis.lu/experts/fp6_candidature.htm</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	European wide
<i>Content/navigation language:</i>	English
<i>Specific remarks :</i>	EU framework programme experts. Independent experts for evaluation of proposals and monitoring and reviewing of projects. This database is for the sole use of the EC does not allow any contact, just an online registration is possible.

### 3.5 European wide

<i>Name of institution/organisation/network:</i>	<b>WITEC – European Association for Women in Science, Engineering and Technology</b>
<i>Address:</i>	WITEC – European Association for Women in Science, Engineering and Technology c/o G/I/S/A Gender-Institut Sachsen-Anhalt GbR Ebendorfer Straße 3 D-39108 Magdeburg
<i>Country:</i>	Germany
<i>Contact person:</i>	Ute Wanzek
<i>Phone:</i>	++49-(0)391-50 665 77
<i>E-Mail:</i>	info@witec.net
<i>Website of the institution/organisation/network:</i>	<a href="http://www.g-i-s-a.de/intwitec/network.htm">http://www.g-i-s-a.de/intwitec/network.htm</a>
<i>Website of the data collection:</i>	
<i>Geographical coverage:</i>	international
<i>Content/navigation language:</i>	English
<i>Specific remarks :</i>	<p>WiTEC has the following aims at European level:</p> <ul style="list-style-type: none"> <li>- to increase the number of girls and women studying SET subjects and to help them progress to related careers,</li> <li>- to develop women's technical and entrepreneurial skills through training initiatives and projects,</li> <li>- to create information exchanges and networking opportunities for women in SET,</li> <li>- to support initiatives to promote the Gender Mainstreaming Policy,</li> <li>- to promote regional, national and international awareness and interest in this field</li> </ul>

<i>Name of institution/organisation/network:</i>	<b>IFUW – International Federation of University Women</b>
<i>Address:</i>	IFUW Headquarters 8 rue de l'Ancien-Port CH-1202 Geneva
<i>Country:</i>	Switzerland
<i>Contact person:</i>	-
<i>Phone:</i>	++41-22-731-2380
<i>E-Mail:</i>	info@ifuw.org; enquiries@ifuw.org
<i>Website of the institution/organisation/network:</i>	<a href="http://www.ifuw.org">http://www.ifuw.org</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	international
<i>Content/navigation language:</i>	English
<i>Specific remarks :</i>	There are several national branches of this federation.

## 3.6 France

<i>Name of institution/organisation/network:</i>	Association femmes et mathématiques
<i>Address:</i>	Institut Henri Poincaré Rue P et M Curie F-7523 Paris cedex 05
<i>Country:</i>	France
<i>Contact person:</i>	Véronique Slovacek-Chauveau (presidente)
<i>Phone:</i>	++33-(0)1-44 27 64 20
<i>E-Mail:</i>	fetm@ihp.jussieu.fr
<i>Website of the institution/organisation/network:</i>	www.femmes-et-maths.fr.fm/
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	France
<i>Content/navigation language:</i>	French
<i>Specific remarks :</i>	Membership-list for internal use

<i>Name of institution/organisation/network:</i>	Association Femmes et Sciences
<i>Address:</i>	Association Femmes et Sciences 93, Rue de Vaugirard F-75006 Paris
<i>Country:</i>	France
<i>Contact person:</i>	Colette Guillopé (Presidente)
<i>Phone:</i>	++33-(0)1-47 70 85 35
<i>E-Mail:</i>	femmes.sciences@wanadoo.fr
<i>Website of the institution/organisation/network:</i>	www.int-evry.fr/femmes_et_sciences
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	France
<i>Content/navigation language:</i>	French
<i>Specific remarks :</i>	Membership-list for internal use

<b>Name of institution/organisation/network:</b>	<b>Association française des Femmes Ingénieurs</b>
<i>Address:</i>	Association française des Femmes Ingénieurs c/à CNISF 7 rue Lammenais F-75008 Paris
<i>Country:</i>	France
<i>Contact person:</i>	Aline Aubertin
<i>Phone:</i>	++33-(0)6 71 69 39 36
<i>E-Mail:</i>	femmes_ingenieurs@yahoo.com
<i>Website of the institution/organisation/network:</i>	www.femmes-ingenieurs.org
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	France
<i>Content/navigation language:</i>	French; English
<i>Specific remarks :</i>	Collect statistics on French women engineers (regular study organised by the National Council of French Engineers and Scientists)

<b>Name of data collection:</b>	<b>Association of female academics Association française des femmes diplômées des universités (AFFDU)</b>
<i>Address:</i>	AFFDU 4 rue de Chevreuse F-75006 Paris
<i>Country:</i>	France
<i>Contact person:</i>	Danielle Gondard
<i>Phone:</i>	++33-(0)1-43 20 01 32
<i>E-Mail:</i>	affdu@club-internet.fr
<i>Website of maintainer of data collection:</i>	<a href="http://www.int-evry.fr/affdu">http://www.int-evry.fr/affdu</a>
<i>Website of data collection:</i>	-
<i>Geographical coverage of data collection:</i>	international
<i>Content/navigation language:</i>	French
<i>Specific remarks:</i>	Membership list for internal use

### 3.7 Italy

<i>Name of data collection:</i>	Associazione Donne e Scienza
<i>Address:</i>	Istituto di Neurobiologia e Medicina Molecolare CNR Viale Marx 15-43 I-00137 Roma
<i>Country:</i>	Italy
<i>Contact person:</i>	Flavia Zucco
<i>Phone:</i>	+39-(0)6-86090245
<i>E-Mail:</i>	f.zucco@inemm.cnr.it
<i>Website of maintainer of data collection:</i>	-
<i>Website of data collection:</i>	-
<i>Geographical coverage of data collection:</i>	Italy
<i>Content/navigation language:</i>	Italian
<i>Specific remarks:</i>	Maintain a membership-list

### 3.8 Lithuania

<i>Name of institution/organisation/network:</i>	LAUW – Lithuanian Association of University Women Lietuvos universitetu moteru asociacija
<i>Address:</i>	P.Visinskio 25, 5419 Siauliai
<i>Country:</i>	Lithuania
<i>Contact person:</i>	Eitutyte Daina
<i>Phone:</i>	++ 370-1-595757
<i>E-Mail:</i>	mnc@cr.su.lt
<i>Website of the institution/organisation/network:</i>	<a href="http://www.su.lt/en">http://www.su.lt/en</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Lithuania
<i>Content/navigation language:</i>	Lithuanian; English
<i>Specific remarks :</i>	LAUW is a non- governmental organisation and member of the International Federation of University Women (IFUW). National organisation (non-governmental organisation) working in the area of improving the status of women in Lithuania with the regards to women employment issues.

### 3.9 Portugal

<i>Name of institution/organisation/network:</i>	Association of women in science (under construction)
<i>Address:</i>	Universidade Nova De Lisboa Faculdade de Ciencias e Tecnologia Quinta da Torre P-2829 Monte da Caparica
<i>Country:</i>	Portugal
<i>Contact person:</i>	Ana Maria Lobo
<i>Phone:</i>	++351-212948387
<i>E-Mail:</i>	aml@fct.unl.pt
<i>Website of the institution/organisation/network:</i>	under construction
<i>Website of the data collection:</i>	
<i>Geographical coverage:</i>	Portugal
<i>Content/navigation language:</i>	Portuguese
<i>Specific remarks :</i>	Are building up an Association for Women in Science at the moment.

### 3.10 Romania

<i>Name of institution/organisation/network:</i>	National Association of Women Physicians in Romania and Romanian Association of University Women
<i>Address:</i>	8 Gheorghe Bratianu Street (Daniceni) RO-Bucharest 1
<i>Country:</i>	Romania
<i>Contact person:</i>	Dr. Maria A. Ciochirc̃a
<i>Phone:</i>	++40-1-222-7126
<i>E-Mail:</i>	mcio22@yahoo.com
<i>Website of the institution/organisation/network:</i>	<a href="http://www.anasaf.ro/english/you/ngo/n/physicianwomen.html">http://www.anasaf.ro/english/you/ngo/n/physicianwomen.html</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Romania
<i>Content/navigation language:</i>	English
<i>Specific remarks :</i>	

<i>Name of institution/organisation/network:</i>	<b>National Committee for Promoting Women in Science in Romania</b>
<i>Address:</i>	19 Barajul Arges RO-sector 1 Bucharest 615161
<i>Country:</i>	Romania
<i>Contact person:</i>	Prof. Dr. Mioara Florica Tripsa (President)
<i>Phone:</i>	++40-21-2331655
<i>E-Mail:</i>	mtripsa@excite.com
<i>Website of the institution/organisation/network:</i>	under construction
<i>Website of the data collection:</i>	under construction
<i>Geographical coverage:</i>	Romania
<i>Content/navigation language:</i>	Romanian
<i>Specific remarks :</i>	The main aims are: <ul style="list-style-type: none"> <li>- to promote and sustain women scientists in their work</li> <li>- to provide management of the institutes and scientific international projects</li> <li>- to identify their needs</li> <li>- to focus on gender studies and equal opportunities,</li> <li>- to promote quality</li> <li>- to support young scientists</li> <li>- to analyse brain drain</li> </ul>

<i>Name of institution/organisation/network:</i>	<b>Women in Science and Technology Association</b> <b>Asociatia pentru Promovarea Femeilor in Stiinta si Tehnologie</b>
<i>Address:</i>	PO Box MG-22 409 Atomistilor St. RO-76900 Magurele
<i>Country:</i>	Romania
<i>Contact person:</i>	Alexandra Caramazoiu
<i>Phone:</i>	++40-4574498 or ++40-(0)721837142
<i>E-Mail:</i>	lexandracaramazoiu@yahoo.com
<i>Website of the institution/organisation/network:</i>	-
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Romania
<i>Content/navigation language:</i>	-
<i>Specific remarks :</i>	A non-profit organisation focusing on promoting women in science and lobbying for women in science.

### 3.11 Slovakia

<i>Name of institution/organisation/network:</i>	<b>The Professional Women's Association</b>
<i>Address:</i>	Professional Women's Association Prešovská 39 SK-82108 Bratislava
<i>Country:</i>	Slovak Republic
<i>Contact person:</i>	Dagmar Simnkova
<i>Phone:</i>	++421-2-55 56 85 57
<i>E-Mail:</i>	profiwom@zutom.sk
<i>Website of the institution/organisation/network:</i>	<a href="http://www.zutom.sk/profwomen">http://www.zutom.sk/profwomen</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Slovakia
<i>Content/navigation language:</i>	Slovakian; English
<i>Specific remarks :</i>	- non-governmental, non-profit organisation supporting education, creativity, interest and involvement of women in social, economic, scientific and political issues - provide research, scientific and publishing activities

### 3.12 Spain

<i>Name of institution/organisation/network:</i>	<b>AMIT – Asociación de Mujeres Investigadoras y Tecnólogas</b>
<i>Address:</i>	Calle Pinar, 25 E-28006 Madrid
<i>Country:</i>	Spain
<i>Contact person:</i>	Profa. Flora de Pablo (Presidenta)
<i>Phone:</i>	++34-91-585-5164
<i>E-Mail:</i>	AMIT@ifs.csic.es
<i>Website of the institution/organisation/network:</i>	<a href="http://www.amit-es.org">www.amit-es.org</a>
<i>Website of the data collection:</i>	-
<i>Geographical coverage:</i>	Spain
<i>Content/navigation language:</i>	-
<i>Specific remarks :</i>	-

## Chapter 4 – Best Practice Guideline for Developing and Revising Women Scientists Databases

This “Best Practice Guideline for Developing and Revising Women Scientists Databases” is based on the results of the overview (Chapter 3) as well as on some of the technical analyses which took place during the feasibility study (Chapter 5).

Fulfilling criteria of quality may lead to a more effective use of databases by their target groups. This Chapter includes best practice criteria and recommendations for developing new databases as well as for maintaining and revising existing databases.

The guideline may be useful for:

- institutions and organisations wanting to develop databases,
- institutions and organisations maintaining and revising existing databases,
- institutions and organisations wanting to harmonise their databases with European standards,
- administrators of women scientists databases,
- persons interested in evaluating and enhancing women scientists databases.

In addition, the document may also be useful for institutions or organisations which maintain databases with both male and female scientists to improve their database with regard to gender equality.

In addition to recommendations listed in this Chapter, we suggest the following two documents for further information on European standards for developing databases:

**CERIF 2000 Toolkit:** the objective of the CERIF 2000 Toolkit is to define a full Current Research Information (CRIS) data model with flexibility to allow the majority of existing CRIS to accommodate their own database structures. And, to define a base framework for data exchange among current Research Information Systems. Also included are recommendations for thesauri and classifications to use for research subject, economic activity, products etc. (<http://www.cordis.lu/cerif/src/toolkit.htm>)

**Code of Good Practice for Current Research Information Systems:** has been created to establish a framework for encouraging interoperability and harmonisation between European (and other) research institutions. It should be regarded as good practice for existing research institutions to review how they operate from time to time as well as making available their experience to new and emerging institutions. Special issues relating to women scientists databases are not covered, since this is a general document for Current Research Information Systems. Still some of the issues mentioned in the Best Practice Guideline are attended to in more detail (i.e. organisational responsibilities) and although a document prepared in 1998, the Code of Good Practice holds information still relevant today.

([http://www.eurocris.org/en/code\\_of\\_good\\_practice/codegpr.doc](http://www.eurocris.org/en/code_of_good_practice/codegpr.doc))

In this Best Practice Guideline there is a distinction between basic and optional criteria. Basic criteria make up the minimum conditions for a good women scientists database – indicated as “highly recommended”. Optional criteria enhance these basic criteria and are the recommendations for a best-practice database – indicated as “recommended”.

Examples of databases which fulfil the criteria of quality may be found in Chapter 3.

## 1 Pre-Development or Pre-Revision Quality Criteria

It is highly recommended that prior to developing the database the following criteria are fulfilled. The procedures described in the following criteria can save time and resources and may enhance the quality of the database project immensely.

### 1.1 Institutional or organisational qualifications

Based on the information which we collected in the DATAWOMSCI Project it is highly recommended that a new database is targeted and specifically designed for women in science and as a political action towards the promotion of women in science. Hence, it is desirable that the institution which want to develop a database have one of the following qualifications represented:

- some members of the organisation or the institution are informed on the structures of and the prevailing conditions in science institutions or organisations, especially on gender equality and women in science.
- the institutions or organisations function as national and/or international contact points, information centres or service points in this issue.

The fulfilment of the first qualification is highly recommended. The institutions or organisations which encompasses the second qualification are likely to spread information in manifold ways and occasions. For this reason databases maintained by such institutions may be used more effectively.

### 1.2 Analysis of the context

#### 1.2.1 Requirements analysis

It is highly recommended that the institution or organisation interested in building a database formulate the answers to the following four questions:

- why do you want to develop this database
- what kind of database do you want to develop
- who are the stakeholders
- when and in which time frame should the database be build up

### 1.2.2 Context analysis

Once a rough definition of the database exists a context analysis is highly recommended in which the following questions can be considered:

- which databases already exist
- who are their target groups, stakeholders and their objectives
- who are their user groups
- what type of database is it (online-offline, national-international, interdisciplinary-specialist etc.)
- what is the difference, added-value and the innovation of the planned database in contrast to the other already existing databases
- are there any (national or international) co-operations possible

### 1.3 Financial resources

A well-planned long term funding of a database is necessary to provide a consistent and reliable database to the scientific community. Therefore, to maintain a new database the calculation of the following durable budgets to the substantial budget needed to develop and maintain a database is highly recommended:

- keeping the database up-to-date (updating of datasets/checking of updated datasets; reminders to update the dataset)
- regular statistics on the database (statistics on data requests, datasets, new registrations and updates of datasets)
- marketing (Flyer, promotion, advertising, publicity)
- user support and networking
- physical face-to-face meetings and activities

Additional financial resources must be calculated for:

- adaptations of the database to newly developed standards according to the content of the database and the software (i.e. software updating and improving the design and structure of the database)
- regular qualitative/quantitative evaluations (see Section 2.8) as well as usability studies.

Depending on the availability of the financial resources it may be necessary to review the requirements analysis of the database.

## 2 Data Collection Quality Criteria

In this guideline there is a distinction between basic and optional criteria. Basic criteria make up the minimum conditions for a good women scientists database – indicated as “highly recommended”. Optional criteria enhance these basic criteria and are the recommendations for a best-practice database – indicated as “recommended”.

The entries 2.1 (Objectives of a database) through 2.5 (Geographical and language coverage) make up the basic decision criteria for any database. They are all highly recommended as best practice criteria for developing databases.

The entries 2.6 (Number of datasets) through 2.10 (Personal information saved in a database) are recommendations yet contain basic as well as optional criteria. In many cases the fulfilment of these latter criteria depend on the decisions taken while fulfilling the criteria 2.1 through 2.5.

### 2.1 Objectives of a database

It is highly recommended that a definition of the objectives and the target groups of a database are given prior to developing the database. Supporting the recruitment of personal staff and/or the recruitment of experts for boards, panels or committees may be recommended as minimal objectives. The database may also be utilised as documentation on the collected information. The relevant information which has to be collected in the database depends upon the interest of the stakeholders and users. That implies that the number of fields may differ according to the objectives of a database. A field is a specific category of information in a database. Each dataset in a database is made up of a combination of fields (e.g. First Name, Last Name, Telephone Number, Address etc.).

It is highly recommended that the objectives of the database are published on the site in order to make them transparent to the users of the database.

### 2.2 Target groups

It is highly recommended that institutions or organisations developing databases define their target groups. This can be done during requirements analysis which may include surveys. The definition of target groups should distinguish between registering women scientists, users searching women scientists and others who have vested interests in the database. The following target groups may be considered:

- universities/higher education institutions
- research institutions
- business enterprises
- political institutions
- individuals in universities, research institutions, business enterprises or political organisations
- women's representatives and equal opportunity officers
- women scientists and researchers

The interests and the technical competencies of the target groups should be taken into consideration while developing a database.

### 2.3 Distinguishing women scientists databases from other data resources or collections

In order to distinguish women scientists databases from other collections we highly recommend the use of the following criteria:

- databases should consist of structured data embedded in a formal structure which offer certain functionalities (e.g. searching, online and/or offline registration etc.).
- they contain a collection of personal data with information concerning scientific field, working area, qualification etc.
- they are not a simple mailing list, membership-database, listing of scientific personnel in academic institutions, project or research report.
- a best practice database should use an up-to-date database software or similar structured documents.

### 2.4 Online database, search and registration

It is highly recommended that the database is available online. Online availability includes:

- search access
- a separate registration service for the women scientists to enter their personal data and a password protected update service

Women scientists should register themselves. Their data should not be entered or included in the database without their consent. Further best practice criteria on registration are listed in Section 3.4.

Finally, users wanting to search the database should be able to do so without any enrolment.

### 2.5 Geographical and language coverage

Prior to developing the database it should be clear if the registration of the women scientists shall be restricted per geography or language. A database may be national, regional or international. All further geographical boundaries should be made explicit. For example:

- nationality
- language
- current working place
- geographical coverage of the searching users (if possible) etc.

Language may also determine who can register and search in the database. Therefore, multilinguality is recommended due to the fact that this may contribute to making the database internationally more accessible. Such databases are meaningful as information sources for the international scientific community as well as for enhancing the visibility of the registered scientists internationally.

Yet, multilinguality is expensive. Hence, bilingual databases in which the datasets are available in the national language and, if different, in the English language are highly recommended.

Further details on multilinguality are listed in Section 3.1.

## 2.6 Number of datasets

The number of datasets may vary according to the disciplines and the particular fields of science which are covered by a database, e.g.:

- an interdisciplinary database covering all disciplines and all fields of science,
- a specialist database covering several or all disciplines but in one particular field of science only (e.g. Gender studies in all disciplines) or
- a specialist database covering one particular discipline (e.g. database for artists only).

In order to avoid the mistreatment of the term "database" for small collections it is highly recommended that the maintainers of the database are informed of the following:

- the total number of women scientists who are working in the scientific field/s and in the country/countries covered by that particular database.
- the percentage of women scientists registered in the database relative to the total number of all existing scientists in the particular field/s and country/countries covered by the database.

## 2.7 Regular updating of datasets

In order to guarantee the quality of datasets the following updating criteria are highly recommended:

- the datasets should be updated at regular intervals
- registered scientists should be able to update their dataset online at any time by using a personal password (see additional best practice criteria in Section 3.4.5)
- automatically generated reminders for update should be sent to the registered scientists in regular intervals (minimum once a year and not too often)
- the updates should be validated
- the last update of each dataset should be displayed online.

It should be noted that the regular updating of the datasets of any database is one of the most important and most time consuming aspects of building up and maintaining

a database - without this, a database for (women) scientists is obsolete within months. For this reason, it is recommended that the scientists are informed prior to registration of the necessity to update. In addition, the users should be surveyed for motivating factors for updating their data.

## 2.8 Evaluation of the database

In order to make the effectiveness and profile of a database transparent and to investigate if the database is living up to its original objectives the following activities are recommended:

- a regular analysis of data requests, datasets, new registrations and of updates of datasets,
- a detailed evaluation of the database carried out regularly by questioning
  - the benefit of the database for the registered persons
  - the importance of the database for institutions/organisations/persons searching for scientists/experts
  - the popularity of the database within its target group

The results of these activities should be publicly available.

## 2.9 Networking and Live Activities

In order to strengthen the communities around the database, networking activities can be organised. Online activities may include:

- invitations to use or register in the database
- newsletters
- mailing lists
- online discussions, meetings or conferences etc.

Such communication should be limited in order to avoid spamming the interested users.

Networking activities can be strengthened through live meetings. Conferences, workshops, or training can be used to:

- introduce the database to further communities
- to encourage scientists to register in the database
- to promote the use of the database by institutions and organisations

## 2.10 Personal information saved in a database

The collection of personal data is sensitive. In deciding which data to collect the maintainers of the databases should be aware of the conflict between the right to self determination of information of the women scientists registered in the database

and the quality of the service for the users. It is therefore recommended to reduce the compulsory information to a minimum without decreasing the quality of service.

In addition, it is recommended that even if the datasets do not have to be as detailed as a curriculum vitae, they should be specific enough to enable searching users to get a well-informed impression of the registered scientists.

The following criteria on personal information to be collected about registered persons are recommendations yet contain basic as well as optional criteria. Those fields which are highly recommended may be used as compulsory fields. Further fields are recommended as optional.

#### *2.10.1 Data protection*

It is highly recommended that the data protection act valid for the database is published on the site, so that all users know what is allowed to happen with the data. When women scientists register in the database they should confirm that they agree to the use of their data.

If any further use of the data is planned (i.e. through data exchange) the data protection act should include a clause which asks the women scientists for permission. In some cases further use of data will not have been planned or no explicit permission from the scientists exists. Then, according to the data protection act of the database, these women will have to be contacted for an extension of their permission to further use their data.

#### *2.10.2 Security*

It is highly recommended that integrity and confidentiality of private data are guaranteed.

#### *2.10.3 Name and address/contact information*

An online database that displays the name and address/contact information of the registered person is highly recommended. In some cases maintainers or users of the database may prefer to hide some of the information for privacy reasons. Technologies can be implemented to guarantee such privacy measures and are in such cases recommended. Yet, for reasons of usability, networking between searchers and registered scientists and potential data exchange among European databases, the anonymity of data is not recommended.

It is highly recommended that the first names and the family names of the registered users are identified in two separate fields in the database.

#### 2.10.4 Sex

It is highly recommended to allow search by sex especially if the database is for scientists of both sexes. Additionally, in the case of databases for scientists of both sexes the percentage of women scientists must be mentioned. These measures may be utilized as an instrument towards gender equality.

For the sake of future data exchange the database should contain a field "sex". This guarantees the following:

- if the data exchange is with a database that contains scientists of both sexes, one can still distinguish the women scientists
- search by sex is further possible.

#### 2.10.5 Level of education

It is highly recommended that the highest academic degrees of the scientists are visible. In addition, a classification of status groups should be available (e.g. PhD's and/or scientists with a doctorate or a university degree etc.). This classification may then be used to search for scientists on the basis of a particular degree.

Further information around the academic degree is also recommended. This information could include theme/title of the dissertation or the *habilitation* thesis, subject, year and location/university at which this degree was received. Since many scientists have several degrees (i.e. PhD M.Sc.) the database should offer multiple fields for each degree (dissertation and habilitation theses).

#### 2.10.6 Academic position

The display of the position of the women scientists is highly recommended (e.g. full professor, associate professor, assistant professor, lecturer etc.). This would allow searching users to identify scientists with a particular standing which may be required for certain positions like distinguished committee or board memberships etc.

#### 2.10.7 Competence/specialism/expertise

It is highly recommended to mention the teaching and researching areas and the subjects or priorities of the scientist. The use of index-lists or thesauri are also recommended. In any case, there should also be an additional field in which the scientists can add detailed information not reflected in the indexes or thesauri. The structure may therefore be an index list for main subjects (e.g. medicine or art history etc.), a more detailed index-list (e.g. oncology or impressionism) and a free-text field to fill in independently (e.g. breast cancer or Claude Monet).

#### *2.10.8 Additional personal information*

Following personal information is recommended to complete the profile of any scientist registered in a database. These are optional criteria of quality:

- language skills (including self-assessment, e.g. basic, fluently etc.),
- projects (title, running time and location),
- membership in scientific organisations (up to five listings),
- grants or fellowships (year and name/title of the grant/fellowship),
- experience in committees and on boards,
- publications (can be a listing, an archive or a separate document)
- a description of the work experience (year/running time, position, organisation)
- a field for free text entry (limited in space to avoid an overfilled database),
- experience in third-party funds and
- the year of birth

#### *2.10.9 Labels and semantics of fields*

In the future it is recommended to standardize the semantics of the fields in women scientists databases or to be aware of the semantics of field names in other databases. Clearing the semantics and using field names consistently will enhance the quality of the data and the user experiences with women scientists databases.

#### *2.10.10 Basic criteria on quality of classifications*

Since the nomenclatures of academic positions are different in most European countries, no standardised specification which fits all European countries exists. Yet, classifications are important for data-exchange or international searches for women scientists. Therefore, the following two measures are highly recommended:

1. There should be efforts at the European level to make these different positions internationally comparable and adaptable for national as well as international databases. This may be done by an expert group consisting of database maintainers, software programmers and academic experts internationally aware of the specific criteria of each academic position also with regard to academic status, degree, denomination etc.
2. The institutions or organisations maintaining databases can contribute towards the standardisation of classifications by:
  - avoiding very detailed or very general classifications especially in the case of national databases. The maintaining institutions are recommended to favour classifications which are easier to map to future standardised European classification systems.
  - using data fields which are flexible and break the information down such that each field only contains one specific piece of information (i.e. separate fields for academic degree and the title of thesis).

### 3 Technical Quality Criteria

The technical criteria include recommendations for enhancing the usability and technical quality of the databases. The recommendations include quality criteria on multilinguality, usability, search and the online registration form.

#### 3.1 Multilinguality

##### 3.1.1 *Awareness of cultural differences*

It is highly recommended that when a site is realised for people from various subcultures that have differences in interpretation of symbols, hierarchies, and aesthetics that these differences are attended to. In addition, scientific, academic or technological cultures can be so different, that according to the context certain terms can be interpreted divergently, e.g. the English term "science" refers to the technical and natural science disciplines in some contexts, and in others it refers to all disciplines, as in its German translation "Wissenschaft".

##### 3.1.2 *Multilingual information retrieval*

In order to support internationalisation of databases and to support the existence of languages other than the dominant scientific languages (i.e. English) it is highly recommended to consider the implementation of multilingual services.

One form of multilingual service is the development of multilingual interfaces. Multilingual interfaces allow users to navigate the database or use the search mask in multiple languages. Especially in the case of databases with multilingual data such interfaces are highly recommended. If the cost of applying multilingual interfaces are too high, then it is recommended to translate key documents, statistics documents, and/or help documents into various languages used by the community of users.

Another multilingual service is the possibility to do cross-lingual search. Cross lingual search allows the users to formulate queries in one language and get search results in another. Cross lingual search may be granted through free text fields or through multilingual classifications.

It may be even more comfortable for the users to search, if they can use a search form in their preferred language, while they are searching for documents in another language. This differentiation is realized in celex, where you can choose the language of the menus and the documents independently (see Fig.1).

In databases with multilingual content and services it is recommended to distinguish between the language of the search form, the language in which the search terms have to be filled in and the language of the results.

### 3.1.3 Switching between different language interfaces

If a database has a multilingual interface it is important that the content is the same when the user changes the language and this must be visible even if the user is not fluent in that other language. If after the change of the language a page appears that looks completely different and seems to have a different content, users may be confused.

**Choose another language ...**

<p>... for the menus</p> <ul style="list-style-type: none"> <li><input type="radio"/> Spanish</li> <li><input type="radio"/> Danish</li> <li><input type="radio"/> German</li> <li><input type="radio"/> Greek</li> <li><input checked="" type="radio"/> English</li> <li><input type="radio"/> French</li> <li><input type="radio"/> Italian</li> <li><input type="radio"/> Dutch</li> <li><input type="radio"/> Portuguese</li> <li><input type="radio"/> Finnish</li> <li><input type="radio"/> Swedish</li> </ul>	<p>... for the documents</p> <ul style="list-style-type: none"> <li><input type="radio"/> CELEX Spanish database</li> <li><input type="radio"/> CELEX Czech database</li> <li><input type="radio"/> CELEX Danish database</li> <li><input type="radio"/> CELEX German database</li> <li><input type="radio"/> CELEX Estonian database</li> <li><input checked="" type="radio"/> CELEX Greek database</li> <li><input type="radio"/> CELEX English database</li> <li><input type="radio"/> CELEX French database</li> <li><input type="radio"/> CELEX Italian database</li> <li><input type="radio"/> CELEX Latvian database</li> <li><input type="radio"/> CELEX Lithuanian database</li> <li><input type="radio"/> CELEX Hungarian database</li> <li><input type="radio"/> CELEX Maltese database</li> <li><input type="radio"/> CELEX Dutch database</li> <li><input type="radio"/> CELEX Polish database</li> <li><input type="radio"/> CELEX Portuguese database</li> <li><input type="radio"/> CELEX Slovak database</li> <li><input type="radio"/> CELEX Slovenian database</li> <li><input type="radio"/> CELEX Finnish database</li> <li><input type="radio"/> CELEX Swedish database</li> </ul>
---	--

Figure 1: Language choice in celex  
[http://europa.eu.int/celex/cgi/sga\\_rqst?SESS=5865!CTXT=3!UNIQ=2!APPLIC=celexext!FILE=mots!DGP=0!LANG](http://europa.eu.int/celex/cgi/sga_rqst?SESS=5865!CTXT=3!UNIQ=2!APPLIC=celexext!FILE=mots!DGP=0!LANG)

Multilinguality	Essential	Important	Nice to have
Awareness of cultural differences		✓	
Multilingual information retrieval			✓

## 3.2 Usability

### 3.2.1 Usability surveys

To be able to realize a system with good usability, it is highly recommended to run surveys with the target groups to find out on:

- how potential users expect the database to work
- how users evaluate the interface

It is also useful to observe and to investigate the needs of the user as they test the functionalities of the database. This could be done through participatory software development methods and prototyping.

### 3.2.2 *Documentation on site*

It is recommended to provide a documentation of the database. This documentation has to be accessible for all users and should contain

- information about the institution maintaining the service
- the purpose of the database
- the possibility to contact the institution via email
- a search tutorial and a global help function in addition to the context sensitive help on each page
- a list of FAQs (Frequently Asked Questions)

The documentation should preferably be provided in different languages and not only in the language of the database, if the database is monolingual.

### 3.2.3 *Description of structure, content and navigation*

If a site has a complex structure and a specific arrangement of its content it is highly recommended to provide a page, where the purpose and the structure of the content are described. This gives an overview to the user, so the orientation and the interpretation of information is easier.

It is also recommended to explain the navigation and its principles as for example done on the site of EUR-Lex (<http://europa.eu.int/eur-lex/en/>), a site where the European union law is published. Without this description it would not be possible to use the site in a reasonable way for the lay public, as this subject is very complex and difficult to understand.

The different navigation needs of the various target groups may also be considered in designing and documenting the navigation.

### 3.2.4 *Visibility of navigation*

To be able to navigate from any page to all other places of the site, it is important for the user to see the navigation bar on every page. It would be very uncomfortable if the user was forced to go from a page with search results several steps back until the navigation appears through which s/he can navigate to another part of the site.

### 3.2.5 *Navigation path and sitemaps*

To simplify the orientation and the navigation it is very useful to display the path the user took to the page s/he is now on (see Fig 2). On a lot of sites this is done by a horizontal list. If the horizontal list gets too long or the number of subjects displayed on top of the page (like navigation, list, search terms,...) increases to a confusing degree and there is not enough space for the content another solution should be considered.



It could be helpful to use more than one level in one classification scheme as in Figure 3 where first a coarse classification of the disciplines are available, through which the sub-listings are provided (see Fig.4). If hierarchical display of classifications are used, these should be easy to navigate and organized carefully, so that the user does not have to go through many branches in order to find the relevant category.

### 3.2.7 *Returning to the search form*

There are two situations in which users may want to return to a search form. One is if they want to refine the search and the other is if they want to start a new search.

In the case of search refinement it is recommended that when navigating back to the search form with the back button of the browser, the entries in the form remain visible. When a search doesn't return satisfying results, it is important for the usability that it is possible to go back to the search form and modify one or more search terms. When refining search it is inconvenient to fill in the complete search form each time.

By the same token, when a user starts a new search it is recommended to provide the searching user with empty search fields. It is also recommended to introduce a "new search"-button so users can easily start a completely new search where the fields of the form are cleared, so that the users don't have to delete the entries per hand.

### 3.2.8 *Browser Compatibility*

It is highly recommended that the display of all the web documents are checked for browser compatibility. Any specific suggestions on which browser to use should be placed visibly on the home page of the database. It is also recommended to organize the web pages so that they are compatible with most browsers, especially with older versions, and are easily accessible even through a slow modem based connection.

### 3.2.9 *Monitor size*

It is recommended that the site is adaptable to a monitor size of 600 x 800 pixels. Monitors of this size are still in use and the site should be accessible to users with such monitors. This means that particularly the navigation is visible and that there is no vertical scrolling. Both the partial display of the navigation as well as vertical scrolling decrease the usability of the site immensely.

### 3.2.10 *Descriptive icon alt-text*

The use of alt-text tags are highly recommended. Alt-text tags appear when the user stops with the mouse pointer over a button, icon or link. These inform the user on the functionality or the content of the button, icon or link. Especially when icons are ambiguous alt-text may enhance the usability of the site immensely.

Usability	Essential	Important	Nice to have
Requirements and usability survey	✓		
Documentation on site	✓		
Description of structure, content and navigation		✓	
Visibility of navigation	✓		
Navigation path and sitemaps	✓		
Classification of data provided in the databases	✓		
Returning to the search form		✓	
Browser compatibility	✓		
Monitor size		✓	
Descriptive icon alt-text	✓		

### 3.3 Search

#### 3.3.1 Multiple Search methods

It is highly recommended that the user has more than one way of searching the database. Some of the alternatives are: free text search fields, alphabetical index, subject listings, thesauri, filters etc.

#### 3.3.2 Variable number of search results displayed

It is recommended that it is possible for the user to change the number of search results displayed on one page. This would be in accordance with the usability principle of adaptability. The optimal number of search results displayed depends on the size of the display and the number of lines each search result consists of.

It is not recommended to have only one long list with all the results to search through by scrolling vertically. This is the case because, firstly it is not very easy to remember the position of certain search results in the list. It is easier to remember the position of a result for example "on the second page on the top". Last, the less scrolling is necessary on a page the better is the usability of the page.

#### 3.3.3 Use of "and", "or" and "not" in search (Boolean operators)

When there are several search fields it is very useful to be able to combine them by Boolean operators to do a more precise search. Many databases allow the users

to return all results which contain all query items, the "and" operator, or at least one of the fields, the "or" operator. Few databases allowed the users to search negations, i.e. "not computing". The logical operator "not" is also recommended so that users can then exclude results with the given keywords and refine their search results.

In cross lingual search is implemented logical operators may deliver worse results, because the translation of terms is often not very precise and if two of these translated terms are combined by an operator the results can easily get even more unprecise. Therefore the use of the logical operators has to be re-evaluated in the case of cross lingual free-text search.

#### 3.3.4 *Visibility of search terms*

If it is possible to combine several search terms, then it is highly recommended to display them on top of the page with the results as a memory hook for the user. When the results are saved or printed it is also very important to be able to remember the search criteria later on, otherwise the search results may lose meaning once the context is lost.

#### 3.3.5 *Labelling of search fields*

There will be no adequate search results if the fields are not well labelled. If the semantics of the user and the database owner are different it may be difficult to get satisfactory results that make sense. Therefore, it is highly recommended that clear labels are selected for the search fields and any ambiguities are taken care of through context based help or extra documentation.

#### 3.3.6 *Description of terms to be filled in*

Short descriptions of the content of the search fields are highly recommended in order to further support the searching user. The descriptions could be summed up as a help page with a list of search terms. If this possibility is chosen, the list has to appear in a new browser window so that it is possible to view both the search form and the help page at the same time.

Another possibility is to have context sensitive pop-up windows that display the help for one field only. The help and the form can easily be displayed at the same time and the user doesn't have to find the term in a long list of explanations. Usually pop-up blockers only prevent unrequested pop-ups, so that there would be no problems opening the helper pop-ups.

In Figure 5, the pop-up has been opened by clicking on the question mark and it is even indicated on top of the pop-up, to which subject it refers.

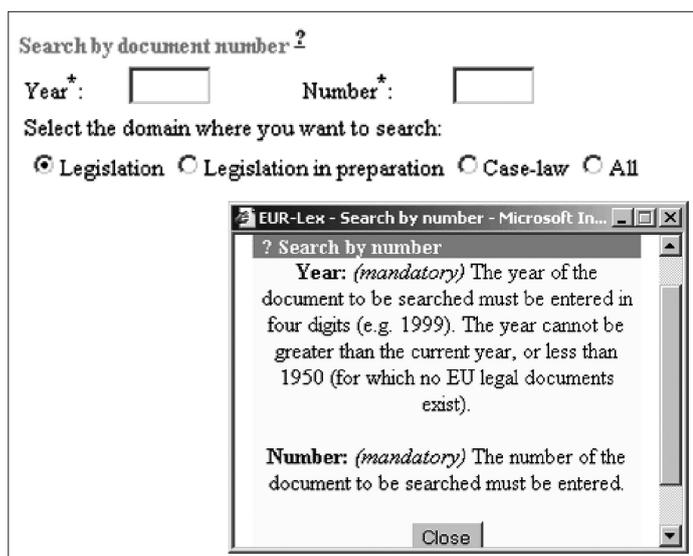


Figure 5: Pop-up window with explanation for search

### 3.3.7 Create a link for each search result

It is useful to create a link for a dataset the user found, so s/he can save it and can later on go back to the results without starting the search again.

Search	Essential	Important	Nice to have
Multiple Search Methods	✓		
Variable number of search results displayed		✓	
Use of Boolean operators	✓		
Visibility of search terms	✓		
Labelling of search fields	✓		
Description of terms to be filled in	✓		
Create a link for each search result			✓

### 3.4 How to register in the database

#### 3.4.1 *Online registration form*

For an online database it is highly recommended to provide the women scientists with an online registration form, through which they can fill in their data and then submit it to the database maintainer. The recommended format and function of the form are described below.

#### 3.4.2 *Preview of registration form*

A registration form can consist of a lot of fields to be filled in and the result is that it won't fit on one page. It is therefore recommended to break down the form into multiple pages. If the form is broken down to many pages it is recommended to have a preview of the registration form. The preview allows users to see the questions they will have to answer during the registration process before starting to fill in the form. The registering scientists may then decide if they want to enter their information into the database.

Preview of the registration form also allows users who only want to have a look at the database and want to find out which data are collected in the database. Otherwise, users may fill in fantasy data, that may cause needless work for the database administrator or affect the quality of the datasets.

#### 3.4.3 *Correction of data before final registration*

Before submitting the data filled in, it is highly recommended that the user who filled out the form is able to review her data. This would allow users to correct potential mistakes that they have perhaps overlooked when filling in the form.

#### 3.4.4 *Email confirmation of registration*

To prevent somebody giving the data of another person unjustifiably to the database, it is highly recommended that it is guaranteed that the data may only be completed by the person herself. It is recommended that when women scientists apply to a database, they get a second look at their data and then either can discard them or validate them for display in the database.

In most databases women scientists are sent a password to the email address they filled in the registration form. This password allows them to login to update or delete their data whenever they want to. This is no absolute protection against misuse but it increases the effort people have to make if they want to publish data of another person in a database. Another recommendation is the assessment of how likely it is that somebody is interested in causing harm to somebody by publishing her name with the wrong data in a database.

#### *3.4.5 Password generation and distribution*

It is highly recommended that the users can select their own passwords. Initially the user may be sent an automatically generated password. An email based confirmation of the password is highly recommended. At the initial login it is recommended that the users are prompted to enter a password of their choice. At this login also recommendations should be made on the selection of passwords.

It is highly recommended that in case the user forgets her password, a new one is generated and sent to her email address.

#### *3.4.6 Plausibility check of inputs in the fields of the registration form*

For the fields of the registration form a plausibility check is recommended to prevent senseless inputs or mistakes that may cause problems afterwards. This means that the format of the inputs is checked e.g. is the email address correct, are dates filled in correctly or are all the compulsory fields filled out. When a mistake is detected an error message can be generated by the system. This explains the mistake and instructs the users on how to overcome the problem.

#### *3.4.7 Confirmation and timescale for the publication of data*

When women scientists have applied to the database, it is important, that the institution maintaining the data collection confirms the registration and informs the expert woman what will then happen with their data and how long it will take to publish the data on the internet.

#### *3.4.8 Spam Protection*

It is recommended that the displayed address/contact information of the registered person should consist of the e-mail address and a link to the personal website of the scientist if it exists. A publicly published e-mail address may be an additional source for spam. Therefore, technologies for spam-resistant publication of emails are highly recommended. In addition, an email service should make it possible to communicate with the scientists without publishing their emails.

The main advantage of this solution is that direct contact with the scientists is possible and therefore, there is the possibility of networking between any database user. In addition, if direct contact to the registered persons is possible, the database usage and its procedure is more transparent, very user friendly and not time consuming. At least, this handling eliminates all possible objections or doubts concerning any control or influence on the search result through the organisation which maintains the database. In addition, the organisation maintaining the database needs no additional staff to answer the search inquiries and therefore is able to save time and money.

How to register in the database	Essential	Important	Nice to have
Online registration form	✓		
Preview of registration form		✓	
Correction of data before final registration	✓		
Email confirmation of registration	✓		
Plausibility check of inputs in the fields of the registration form		✓	
Compulsory and optional fields	✓		
Confirmation and timescale for the publication of data			✓
Spam Protection	✓		



## Chapter 5 – Future Perspectives for Linking Women Scientists Databases – A Feasibility Study

### 1 Studying future perspectives

#### 1.1 Introduction

This Chapter of the DATAWOMSCI Project publication documents a study on the feasibility of linking existing women scientists' databases in Europe. Linking the existing women scientists' databases could enhance information retrieval on women scientists, strengthen cooperation among European institutions and organizations maintaining women scientists data collections. Further it will increase the visibility of the women scientists as well as the institutions and organizations maintaining databases.

Technically as well as otherwise, linking databases, or in the technical language "building a **meta-database**", is not an easy task. Actually, during our study, we found out that many technologies exist to link existing databases. Yet, it is the content, which has a strong interplay with the technology, that requires special attention and demands cooperation among the future partners of a meta-database project.

Four issues are central to the feasibility of a European women scientists meta-database. The first is the negotiation of the content of a future meta-database. Is there a common denominator of the local contents of the existing women scientists' databases which also makes sense at the European level? Second, is there a way to use European standards in order to organize, classify and link this content in a meaningful way and to enhance the comparability of content between the countries? Next, how could issues of multilinguality be dealt with within the context of women scientists' databases? Last, which technologies would accommodate the most data collections of women scientists' databases, given that the technological infrastructure among the countries vary immensely?

#### 1.2 The process and the partners

We ran the feasibility study in the second half of the year 2004. Five databases were selected for the feasibility study. The maintainers and administrators of these databases showed great support and interest in the feasibility study. They answered our questions and generously provided us with the necessary information to complete the study. The databases and the maintaining organizations which cooperated with us are:

- CEC-WYS of the Central European Centre for Women and Youth in Science, Czech Republic
- FemConsult, Database of Women Scientists of the Center of Excellence Women and Science CEWS, Germany
- femdat - Swiss Database of Women Experts of the femdat Office, Switzerland

- Kilden's researcher database, sub base for women researchers of KILDEN (Norwegian Information and Documentation Centre for Women's Studies and Gender Research), Norway
- European Database of Women Experts in Science Engineering and Technology (SET) of WITEC UK, United Kingdom

In order to complete the feasibility study, the maintainers and the administrators of these databases were sent a feasibility questionnaire. The questionnaires and the rest of the analyses in this document reflect the state of the art in the Summer of 2004. All databases have undergone changes since then. Yet, the analysis of the feasibility is still relevant for a future DATAWOMSCI meta-database since it highlights the important issues regardless of whether the examples they are based on are obsolete.

### 1.3 The organization of the document

This document deals with all four issues listed in the Introduction. The rest of the document is organized as a catalogue of the issues to be dealt with when building a meta-database. Readers interested in finding out about the course of the feasibility study and the process of taking decisions or making recommendations may prefer to read the document from the beginning to the end. Those readers who are interested in specific issues may simply skip to the relevant sections.

Each section is self-contained, yet makes references to other sections when necessary. Furthermore, each section identifies the criteria for making one decision against another, describes the relevant processes within the DATAWOMSCI Project and identifies open questions which are addressed to the future partners of a meta-database project. Further, some of the recommendations are elicited and listed in the "Best Practice Guideline for Developing and Revising Women Scientists' Databases".

We now describe each of the sections:

#### **Section 2 – The decision to set up a meta-database:**

describes the reasons for linking existing data resources in a meta-database.

#### **Section 3 – Information retrieval through a Meta-database:**

gives a technical explanation of the problems involved in building a meta-database. In addition, we list the technologies that we evaluated for building a meta-database and present the results of the evaluation. Last, we list the final criteria for selecting the Open Archives Initiative Protocol for Harvesting Metadata (OAI-PMH) as the technology for a meta-database of women scientists.

#### **Section 4 – Describing the application domain for the feasibility study:**

lists the criteria for the selection of the best practice databases which were evaluated for the feasibility of a meta-database. Next, the domain of the study, which includes the databases and the stake-holding actors, is described using the vocabulary of the OAI-PMH.

**Section 5 – Analysis and description of the databases:**

describes the characteristics of the selected databases relevant for a meta-database. These include understanding the objectives and functionalities of existing women scientists' databases in order to reflect these in a meta-database. Additional topics are registration, conditions for searching, passwords, finances and languages.

**Section 6 – The feasibility of the DATAWOMSCI Service Provider:**

analyzes the possibilities and obstacles for linking (harvesting) the content of the databases in a central meta-database. Other than the common denominator in the content, redundancy of data, privacy, security, character encoding and data protection are considered. The work to be done in order to implement such a database is outlined.

**Section 7 – Analysis and the evaluation of cross-lingual information retrieval for the DATAWOMSCI Service Provider:**

The DATAWOMSCI Service Provider raises two important questions. The first question relates to the challenge of searching and using multilingual content. The second question is about developing a provider for an international multilingual community. In both cases an evaluation of the technologies for multilingual content is necessary. In Section 7 we introduce cross-lingual technologies, which allow users to search in one language content in many languages. Current state of the research is summarised and recommendations are made for the implementation of exciting but costly cross-lingual search technologies.

**Section 8 – Introduction to international web sites:**

Developing a project for an international community and with multilingual content raises further questions and produces different sets of requirements. The section on international web sites sketches the issues at hand and makes recommendations on dealing with the complexities of international web sites.

**Section 9 – Three DATAWOMSCI Provider Solutions:**

In this section we fulfil the main goal of the feasibility study: we sketch three alternative meta-database models for the future. The three models relate to the different problems and possibilities listed in the previous sections and offer feasible solutions. The solutions are called: the multilingual DATAWOMSCI Service Provider, the monolingual DATAWOMSCI Service Provider and the DATAWOMSCI Information Provider. Finally, it is shown that combinations of the three alternatives are also possible.

**Section 10 – Metadata for the DATAWOMSCI Providers:**

Exchange of data is easily feasible through the use of metadata which describes the content. Yet, if the metadata are not consistent or standardized data exchange may be difficult. The CERIF 2000 Toolkit offers standardized metadata models for Current Research Information Systems. In this section we analyze the adaptability of the CERIF 2000 metadata models to future women scientists' databases as well as to a future DATAWOMSCI Provider. Also an analysis of intelligent providers of the future as sug-

gested by the CERIF initiative is introduced as long term technological considerations. This section is the most technical and is recommended reading before preparing or revising data models of women scientists' databases.

**Section 11 – Concepts for Search Interfaces on a Meta-Database:**

deals with issues of usability and user friendliness in searching and the displaying of search results. Tips are also included on the organization of cross-lingual search.

**Section 12 – Conclusion:**

In the conclusion we list the recommendations of the consortium partners of the DATAWOMSCI Project following up the feasibility study.

We hope that this document will be useful for institutions and organizations which are currently interested in developing women scientists' databases, as well as institutions which want to revise their existing women scientists' databases for a European meta-database. The appropriateness of our recommendations for each maintaining institution or organisation will depend on their resources as well as the needs and requirements of their users and stake holders.

The "Best Practice Guideline for Developing and Revising Women Scientists' Databases" is partially derived from the recommendations elicited in this document. Further, the Best Practice Guideline complements this document by identifying the cycles of database development and making further related recommendations.

Last but not least, we thank all the partners of the DATAWOMSCI Project as well as the institutions and organisations maintaining the databases which we studied for their support, ideas, and feedback. We also thank them sincerely for giving permission to publish the data collected through the feasibility questionnaires.

## **2 The decision for a meta-database**

Users searching for specific resources on the Internet often encounter frustration in their efforts. Even though powerful search engines on the Internet provide users with qualitatively better search results than ever before, they do not accommodate the needs of the kind of users who would be interested in searching in specialized data resources like databases of women scientists.

Various difficulties may arise when searching such specialized databases. First of all, many users interested in these resources do not know of the existence of these resources. Next, since search engines do not index the entries of databases, the contents of specialized databases will not be returned in the search results. Once the users find out about the resources, in order to find the information they are looking for they may decide to go through the resources sequentially. This can be a tedious task.

Our goal in the feasibility study of the DATAWOMSCI Project is to attend to both problems in the field of women scientists' databases through services offered by meta-databases. **Meta-databases** link existing databases and make them accessible through a central interface. In our study of meta-databases, we attend to the two problems described above. We first look at the possibility of searching numerous databases through a meta-search. Further, we attend to the objective of finding and accessing resources on women scientists through a meta-database.

### 3 Information Retrieval through a Meta-Database

Once a user searching for women scientists has found the databases in which s/he can search, s/he will most likely try out each of these databases sequentially. Sequentially running through databases or other electronic resources has some disadvantages due to the **heterogeneity** of these sources. The user will have to acquaint herself/himself with resource-specific interfaces. These interfaces may not always be in the language of the user. In some cases s/he may have to use a variety of inconsistent query languages and semantic conventions.

For example, the "name" field in one resource may mean the complete name of the person, and in another resource only the surname, varying search results. After lengthy searches of numerous resources, the user may find the relevant information in the last resource down the list. It may be difficult for the user to decipher the search results which are in different formats. These and others are problems typical of what in computing is called **information retrieval in heterogeneous distributed information systems**. By **information retrieval** we mean the process of determining the relevant documents from a collection of documents, based on a query presented by the user.

At the beginning of the feasibility study, we looked at possible technical solutions to dealing with the problem of information retrieval in heterogeneous distributed information systems. Specifically, we studied protocol based distributed information retrieval. **Information retrieval protocols** regulate the communication between a central system and the distributed databases and allow users to simultaneously search the distributed databases.

In total, we evaluated two information retrieval protocols, the Z39.50 Information Retrieval Protocol and the Open Archives Initiative Protocol for Metadata Harvesting. In our evaluation we analyzed the advantages and disadvantages of each protocol. We also considered the relevance of the protocols to the general goals of the DATAWOMSCI Project to make distributed resources on women scientists more accessible.

At the end of the evaluation we decided to select the OAI-PMH. The selection criteria are listed in Section 3.3 We do not include a study of the Z39.50 Information Retrieval Protocol in this document, but refer to the advantages of the OAI-PMH over Z39.50 in the case of the DATAWOMSCI meta-database.

### 3.1 Understanding and Evaluating the Open Archives Initiative Protocol for Meta-data Harvesting (OAI-PMH)

#### 3.1.1 *Origins of the Open Archives Initiative Protocol for Metadata Harvesting*

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), which has been developed since 1999, attends to the problems of heterogeneous distributed information systems, just like the Z39.50 Information Retrieval Protocol. The main idea is to collect metadata in central servers, since the cross searching possible with protocols like Z39.50 does not scale well to large numbers of resources. Further shortcomings of the Z39.50 protocol are attended to. These include issues concerning mapping to varying query languages, ranking of results integrated from multiple resources, and browsing of the resources.

Initially, the Open Archives Community was mainly interested in developing technological tools in order to bring changes to the scholarly publishing paradigm. The technologies to be developed were expected to accommodate the rapid pace of dissemination of scholarly results and form an alternative to the economical problems of scholarly publishing, which has resulted in rising subscription prices and relatively stagnant research library budgets.

Following from these goals, at a meeting in Santa Fe, the participants decided that a **low-barrier solution** was critical towards widespread adoption among communities publishing scholarly work. Hence, they decided to adopt an interoperability solution known as **metadata harvesting**. In the following years, communities other than those initially interested in scholarly publishing also showed interest in the harvesting protocol. Hence, the OAI community decided to generalize the applicability of the protocol to a broad range of domains.

The term **archive** in the title of the Open Archives Initiative is intended as a repository for stored information. The term **open** is defined from the architectural perspective to mean "defining and promoting machine interfaces that facilitate the availability of content from a variety of providers." Openness, hence, does not mean "free" or "unlimited" access to the information repositories.

The adoption of technical standards depends on the stability of the organizations responsible for the promotion and maintenance of the standards. In August 2000, the Digital Library Federation and the Coalition of Networked Information supported the building of an OAI steering committee with membership from a cross-section of communities and a level of international participation. These develop and promote the OAI-PMH related specifications and standards, so far, very successfully. This stable support through the OAI community also plays a role in selecting the OAI technology.

#### 3.1.2 *Description of the OAI-PMH*

The Open Archives Initiative Protocol for Metadata Harvesting provides an interoperability framework based on the harvesting or retrieval of metadata from any num-

ber of widely distributed databases. Two main technical participants are identified within the protocol: the **data providers** and the service providers. The data providers publish metadata about their content using the OAI technical framework. The **service providers** collect the metadata using the OAI protocol on regular intervals and offer a search interface and other value-added services to users or further service providers as shown in Figure 1.

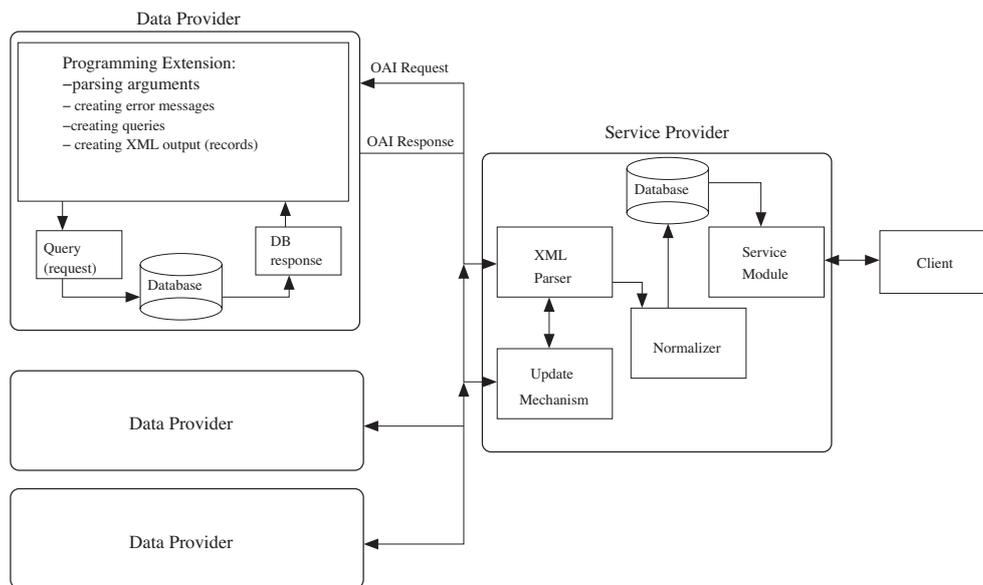


Figure 1: The OAI model for a service provider harvesting from multiple data providers.

In the OAI Harvesting protocol data providers maintain the data which is regularly harvested by the service providers. Often a summary of each record is passed onto the service provider. An OAI data provider must run software or programming extensions that allows its metadata to be requested in bulk by service providers participating in the Open Archives Initiative. Specifically, the programming extensions recognize the requests made to the database, run queries to collect the requested data, prepare the requested data by creating an XML output, and send the XML output to the requesting service provider.

The service provider offers value-added services. The service providers collect all the data in the data providers in its database. It does this by requesting the relevant data from the data providers, updating the data at regular intervals through an update service, and organizing the data.

The service provider may then offer additional services like search, navigation, ranking etc. For each summarized record that a user searches and finds in the service provider, the originating data provider is listed and a link is set to the complete record in the originating database. If the user wants access to the complete record, then s/he may use this link, as shown in the user screens in Figure 2, and is forwarded to the display of the complete record presented by the originating data provider.

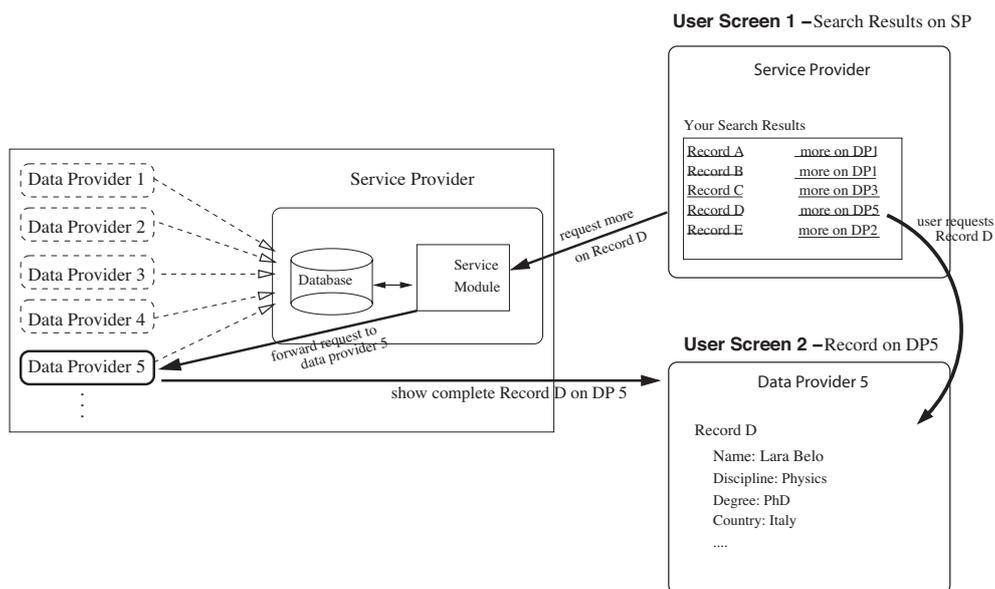


Figure 2: The display of the search results and records in OAI. User makes a request for more information on one of his/her search results and receives the complete record displayed by the originating data provider.

### 3.2 Relevance of the OAI-PMH for the DATAWOMSCI Project

The OAI-PMH Service Provider indexes all the distributed information on a central server. This brings about some efficiencies in comparison to the Z39.50 Information Retrieval Protocol. For Z39.50 based search to be successful all the searched resources need to be available at the time of the search. If not, the search is incomplete. If all the resources are running, the protocol runs only as fast as the slowest resource. Further, the Z39.50 does not scale well to a large number of resources.

Having a common index, as in the OAI-PMH, is a more powerful way to search than via distributed searching. Searching through a common index is normally much faster and is not affected if one of the distributed databases is not functioning at the time of the search. Since efficiency and flexibility in terms of running resources are desirable in the context of women scientists' databases, the efficiency of the OAI-PMH in the recounted issues turns out to be very important.

The OAI-PMH is a low-barrier approach, and hence has functional limitations. There are interoperability issues which are covered more completely by the Z39.50 protocol. But, it is notable that the interoperability work increases in cost with the increase in functionality. This is where the low-barrier approach of the OAI-PMH is valuable.

The most obvious disadvantage of the OAI protocol is that the index might not be up to date if the metadata are not harvested regularly or in a timely manner. There is generally no way for the user to know when the metadata were last harvested. The protocol hence needs to be fine-tuned to the update needs of its users.

The simplicity of the OAI-PMH makes it fairly straightforward to implement. Yet, the lack of standardized metadata formats may require extra work to be done in order to

normalize metadata before the integration or presentation of results. The normalizing process could ultimately lead to writing wrappers, which is a tedious task. A possible solution may be the development of middleware tools that service providers could use for metadata normalization. A possible better suggestion is the definition of community based standard metadata formats (or using existing ones) and convincing data providers to export them.

Another interesting option using the OAI-PMH is the possibility to directly navigate through content provided by a data provider. Especially in the case of European women scientists' databases, this is an important alternative that can be used to offer the users multiple ways of searching the content and multilingual search through navigation.

In Section 4 we introduce the databases selected as potential data providers. Next, in Sections 5 and 6 we analyze the feasibility of applying OAI-PMH and developing a service provider in the context of these databases.

### 3.3 Final Criteria for the selection of the OAI-PMH

After an analysis of the two protocols in the context of the DATAWOMSCI Project, it becomes clear that the Open Archives Initiative Protocol for Metadata Harvesting is more appropriate than the Z39.50 Information Retrieval Protocol for numerous reasons. These can be summed up as follows:

- although a great community of supporters for digital library implementations of Z39.50 exist, applying Z39.50 to databases of scientists, a new application domain, would require a complex and costly implementation process.
- the time and financial cost of implementing the Z39.50 would lie more on the existing distributed databases than on the developers of the meta-database (i.e. query mapping etc.). Especially in the case of women scientists' databases the institutions or organisations maintaining the databases may not have the resources for such an implementation.
- the OAI-PMH is easier and less costly to implement and is more flexible with respect to new application domains. In addition, the OAI-PMH is supported by a growing number of enthusiastic supporters.
- there are inefficiencies to the Z39.50 protocol which are solved within the OAI-PMH protocol.
- the OAI-PMH allows users to navigate through content, which may be an important feature for multilingual solutions.

Hence, we selected the OAI-PMH protocol for evaluation of the feasibility of a meta-database. In Section 4, we evaluate the feasibility of a meta-database on women scientists implemented using the OAI-PMH.

## 4 Describing the application domain for the feasibility study

### 4.1 Picking the databases for the feasibility study

At the beginning of the DATAWOMSCI Project we picked 5 best practice databases as objects of the feasibility study. The following criteria were used to pick the databases:

- the data collection is an online database and online registration for data contributors is available.
- the data collection covers a specific geographic region.
- the data collection is established in its region.
- the owners of the data collection are willing to take part in the feasibility study. This means two things:
  - the owner of the data collection is willing to answer a feasibility questionnaire.
  - the owner of the data collection can imagine contributing to a future meta-database of women scientists' databases.

Last but not least, the representation of as many geographic areas as possible within the feasibility study was taken as a selection criterion.

The databases selected for the feasibility study, with the consent of the institutions and organisations maintaining the databases, were the following:

- CEC-WYS of the Central European Centre for Women and Youth in Science, Czech Republic
- FemConsult, Database of Women Scientists of the Center of Excellence Women and Science CEWS, Germany
- femdat - Swiss Database of Women Experts of the femdat Office, Switzerland
- Kilden's researcher database, sub base for women researchers of KILDEN (Norwegian Information and Documentation Centre for Women's Studies and Gender Research), Norway
- European Database of Women Experts in Science Engineering and Technology (SET) of WITEC UK, United Kingdom

In order to complete the feasibility study, the maintainers and the administrators of these databases were sent a feasibility questionnaire. The questionnaires and the rest of the analyses in this document reflect the state of the art in the Summer of 2004. All databases have undergone changes since then. Yet, the analysis of the feasibility is still relevant for a future DATAWOMSCI meta-database.

### 4.2 Adapting the Vocabulary of the OAI-PMH and Identifying the Actors

According to the terminology of the OAI-PMH the five databases CEC-WYS, FemConsult, femdat, Kilden and SET are called **data providers**. The actors responsible for the data

providers are the institutions or organisations maintaining a database, as well as companies or further persons administrating the database, or shortly **administrators**.

The portal on which the harvesting meta-database is located and services (i.e. search) are available for end users is called a **service provider**. We shall call this portal the **DATAWOMSCI Service Provider** as portrayed in Figure 3. The DATAWOMSCI Service Provider will also most likely have a maintaining institution or organisation as well as administrating persons.

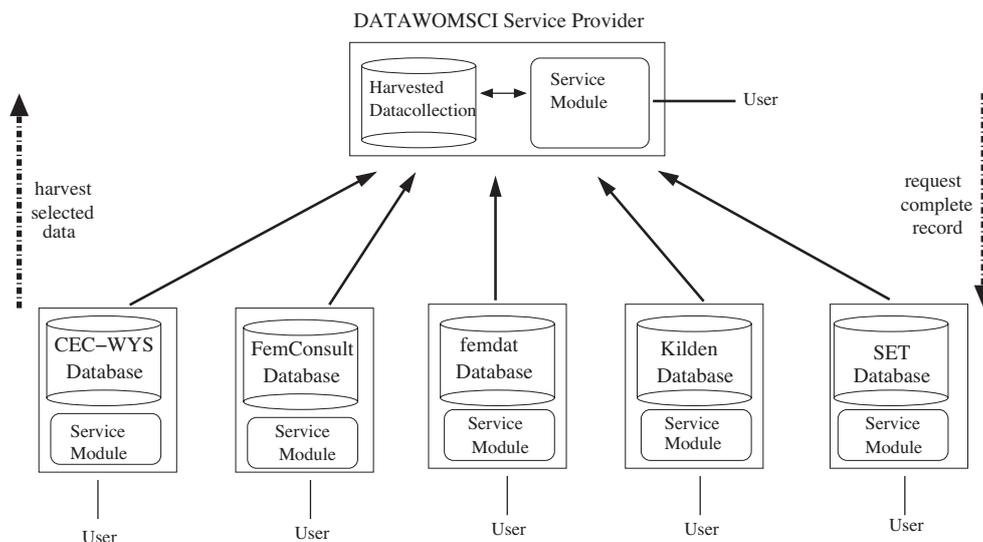


Figure 3: An overview of the DATAWOMSCI Service Provider

According to this OAI-PMH meta-database model, the DATAWOMSCI Service Provider allows users to search for datasets on women scientists from all five databases at once. These are the **searching users**. A searching user of the DATAWOMSCI Service Provider is only able to search the fields which are common to all the databases.

Once the user wants more information on a woman scientist s/he can get the complete record of this woman scientist from the originating database. This record would then include all the information about the woman scientist available on the originating database.

The DATAWOMSCI Service Provider hence functions as a gateway. It allows for global search on limited fields of the data providers and makes the women scientists in local databases visible at a European level. In addition, the DATAWOMSCI Service Provider contributes to the visibility of the data providers.

The women who want to register in one of the data providers are called **data contributors**. Data contributors can register in their local database of choice. The DATAWOMSCI Service Provider will not offer registration services. Instead, the DATAWOMSCI Service Provider will encourage women to register with one of the data providers. Please refer to Section 5.2.1 for more information on registration and updates in the DATAWOMSCI Service Provider model.

There may be further actors or stakeholders (i.e. universities or organisations interested in the existence of women scientists' databases, funding organisations, Research Information Systems providers, researchers etc.). We only consider the actors listed above: the maintaining institutions and organisations of the data and service providers, administrators, searching users, and data contributors. The distinction of the actors affects the analysis of the system and hence the feasibility of a meta-database.

Section 6 is a description of the relevant characteristics of the selected databases for the DATAWOMSCI Service Provider. In Section 6.1 we describe the fields which can be harvested by the DATAWOMSCI Service Provider from the five databases and address problematic issues. In Section 7 we address the alternative solutions to multilingual information retrieval.

## 5 Analysis and Description of the Feasibility Study Data Providers

In order to do a study on the feasibility of the DATAWOMSCI Service Provider, the objectives and the emphases of the DATAWOMSCI Service Provider (DSP) need to be defined. The question "which information should be at the center of the DATAWOMSCI Service Provider?" needs to be answered.

There are at least two approaches which are useful in determining the emphases of the DATAWOMSCI Service Provider. The first is an agreement among all the institutions taking part in the DATAWOMSCI Service Provider on a common set of emphases which would then be implemented in cooperation with the existing databases. The second is finding the common denominator in the emphases of the existing women scientists' databases and developing a meta-database with these emphases.

In this study the latter approach is taken. We determine the technical limitations or possibilities given the definitions, data models and content of the existing databases. We try to define the greatest common denominator among the existing content which can be displayed on the DATAWOMSCI Service Provider. The user searching on the DATAWOMSCI Service Provider will then be able to search only this content. This is a top-down database integration which does not require the completeness of the data in the DATAWOMSCI Service Provider. If the searching user is interested in finding out the complete information about a women in his/her search results, then s/he gets forwarded to a page on the originating database with the complete record.

In the upcoming sections we analyze some of the characteristics of the potential data providers, the five selected databases, in order to understand their effect on the feasibility of the DATAWOMSCI Service Provider. This analysis is also used to make recommendations for a future DATAWOMSCI Service Provider, which not only harvests datasets but enhances the visibility of potential data providers in Europe.

## 5.1 Objectives of the Databases

The objectives of the five databases are overwhelmingly the same. A list of the results is presented in Table 1. Briefly the common goals of the databases are:

the promotion, staffing, and networking of women in research and science, including boards and committees. In addition, encouraging the collaboration of women in EU Research and Development projects.

For the technical feasibility study these will constitute the guiding objectives. There was one exceptional goal of the Kilden database: "to find relevant literature written by the [women] researchers." This is an objective that requires a different emphasis on the listing of publications. A similar interest in listing "projects" in which the scientists participate in exists, although not explicitly formulated as a goal. We shall take these as desirable objectives of the DATAWOMSCI Service Provider in fulfilling the aims of the DATAWOMSCI Project.

Table 1: Objectives of the databases

	CEC-WYS	FemConsult	femdat	Kilden	SET
promoting women					
promoting women in science and research	x	x	x	x	x
staffing of key position in science and research	x	x	x	x	x
staffing of boards and committees	x	x	x	x	x
mentoring			x		x
role models					x
networking in science and research	x	x	x	x	x
consulting and expertise	x	x		x	x
membership records					
collaboration in EU R&D projects	x	x			x
other: public speakers					x
other: media interviews				x	

## 5.2 Existing Services and User Support Relevant to the DATAWOMSCI Service Provider

### 5.2.1 Registration and Updates for Data-Contributors

The DATAWOMSCI Service Provider will be solely for searching and displaying search results and not for registering user data. Therefore, no data registration service will be offered by the DATAWOMSCI Service Provider. All of the selected databases allow

users to register online and to update their data online. See Table 2 for an overview of registration and data-update services.

Still, the DATAWOMSCI Service Provider can offer the interested users with information about registering themselves in one of the five databases. For each of the databases a short description of the aims and conditions of the databases can be listed and a link to the registration pages and home pages of the database can be listed.

Calls for updates are explicitly mentioned in only two of the five databases, femdat and CEC-WYS. Regular calls for updates is a quality measure which we recommend as a best practice criteria. Whether a call for update should be combined with stipulations is left open. The only database that has stipulations is femdat, where, if for a year no updates have been made, the data gets deleted. It would be interesting to find out through a user survey which methods best motivate users to update their data.

Table 2: Registration and Data-update Services

	CEC-WYS	FemConsult	femdat	Kilden	SET
online registration	x	x	x	x	x
offline registration					x
online update	x	x	x	x	x
call for updates	x		x		

The following is an overview of the update services for the data collections:

<b>CEC-WYS:</b>	users may update their information online. The CEC-WYS database will be accompanied by a number of activities for dissemination. These activities will also include calls for updates.
<b>FemConsult:</b>	users may update their information online. To our knowledge calls for updates are made, but it is not clear how often. Any updates made by the contributors are checked for consistency by the CEWS.
<b>femdat:</b>	users may update their information online. Every six months a call is sent out to all registered users to update their information. If after two calls no updates have been made, the user-data are deleted.
<b>Kilden</b>	users may update their information online. After the funding ended in 2003 no further calls for update were sent out to the users. Currently a self-updating system has been implemented and will be available for the data contributors soon. Once this takes place, contributors will start receiving update calls regularly.
<b>SET</b>	The SET database has not had funding since the year 2002. Hence, the SET states that the database is out of date. There is also no mention of "calls for updates" prior to that date. It is also not known how in the future these services will be implemented. In the past it was also possible to nominate a person to the database, which could be done on paper.

### 5.2.2 Registration for searching users

None of the existing databases require users to register before searching the database. If contact information about contributors is hidden for privacy reasons the searching users may have to contact the institution and provide his/her identity, but this is not a formal registration. For more information on direct contacts to institutions see Section 6.1.4

The DATAWOMSCI Service Provider is recommended to keep the search facility open to all users. Registration for other services to be offered by the DATAWOMSCI Service Provider may be considered for each specific service (i.e. mailing lists, newsletter, personalized access to the service provider etc.)

### 5.2.3 Passwords for Data-Contributors

All databases have a password solution for their contributors to update their data. The solutions include generated passwords, self-selected passwords, and email address of the contributor as password.

The security of personal data is valuable. Therefore we recommend the use of the latest security tools for providing the contributors with good passwords through secure channels. The security of the data is a precondition for the quality of the data. Furthermore, the quality of the data is a precondition for the quality of the DATAWOMSCI Service Provider.

If passwords are to be distributed for any kind of registration to the DATAWOMSCI Service Provider the following best practice criteria should be taken into consideration:

user determines password:	if users are allowed to determine their passwords there should be an introductory text listing criteria for good passwords. A password quality checker should be implemented to check if the user has adhered to the minimum criteria.
user password reminder:	users often forget their passwords. There should be a reminder mechanism or a new password should be generated and sent to the user with a request to change the password upon first use.

### 5.2.4 Financial Resources

All of the databases depend either on funds internal to the organization, from the government or from the EU. None of the databases depend on registration fees or advertisement for their funding. If this were the case, the DATAWOMSCI Service Provider could have been detrimental to their further funding.

It is not known if the database maintainers need to report to their funders the statistics on visits and searches for further legitimating and funding of the database. If this is the case, it is recommended that these databases are provided with DATAWOMSCI Service Provider visitor statistics. The databases should also be asked if they need any further documentation from the DATAWOMSCI Service Provider for similar purposes.

### 5.2.5 Languages

The languages used within the databases can be analyzed at three levels. The first is the navigation language. The second is the language of the data entered by the contributors, and last is the language of the search. Table 3 provides an overview of the languages used within the selected databases.

Table 3: Languages of the Databases, (N=Navigation, D = Data collection, S = Search)

	CEC-WYS			FemConsult			femdat			Kilden			SET		
	N	D	S	N	D	S	N	D	S	N	D	S	N	D	S
Czech		x													
English	x	x	x				x	x	x		x		x	x	x
French							x	x	x						
German				x	x	x	x	x	x						
Italian							x	x	x						
Norwegian										x	x	x			

The only database that has multilingual navigation are the femdat database: in English, French, Italian and in German. Furthermore, the data collection exists in four languages: English, French, German and Italian. Partially it is possible to do cross-lingual search by navigating through a classification of scientific fields (See Section 7.2.5 for further details on this method). The free text fields for search are not language aware and only search for patterns.

Kilden and CEC-WYS currently have bilingual data collections. Kilden requires the data contributors to enter some of the data in English. The Kilden search engine is not language aware and is based on pattern matching, meaning that if there are matching English words, then the results are displayed.

CEC-WYS currently contains data in Czech as well as in English. The contributors of Czech data shall be requested to translate their content into English in the near future. Until then a set of keywords is available on the search mask in Czech, German, and English. In the future, the administrators of the database are planning to restrict the data collection language to English.

## 6 The feasibility of the DATAWOMSCI Service Provider

### 6.1 The content of the DATAWOMSCI Service Provider

#### 6.1.1 The common fields

In each of the databases studied in this feasibility report the data are structured differently. The end user receives the complete data displayed. In the database this data are broken down into what in database language are called fields. A field is a specific category of information in a database. A field for example can be the first or last name, age or telephone number. Sets of fields define entities, which can be for example a person, the complete address of a person or a project. The complete set of fields which make up the entities, and all the relationships between these entities are called the data model of the database. No two data models of the selected databases are the same.

In addition, even if the content were the same the name of the fields differs between the databases. Some commonalities exist, like "name" is the name of the person in every database. But, in some cases it is the field holding the family name and in others it is the field holding both the first name and the family name. Figure 4 shows a user searching the field "Name" in database A which refers only to the family name. The same search in database B returns different results since "Name" in database B refers to the complete name of the person. Such differences need to be considered when harvesting data from the selected databases.

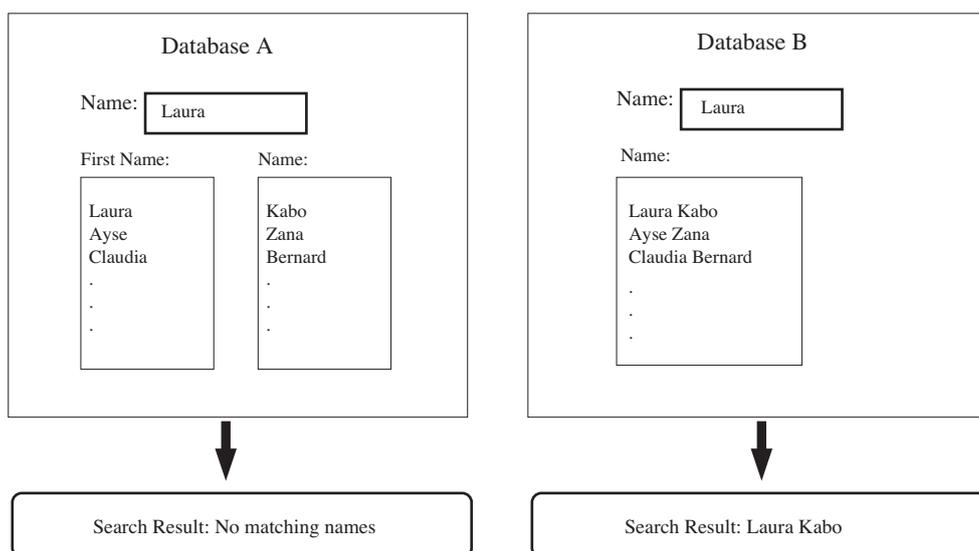


Figure 4: Searching the field "Name" in database A returns no results since only field "Name" and not "First Name" is searched. Searching the field "Name" in database B returns Laura Kabo

The OAI-PMH solves this problem in that it defines common fields and receives the defined content from each database. So if we were to define the field "CompleteName" as a field in the DATAWOMSCI Service Provider which includes first and family name,

then data would be exported in XML from the originating database to fit this criterion. Problems could arise if we defined two fields "First Name" and "Family Name" for the DATAWOMSCI Service Provider and if one or more of the harvested databases did not have these two fields separated.

There are discrepancies in the number of fields available in the data providers selected for the feasibility study. For example, Kilden has over 70 fields while CEC-WYS has a little over 25 fields. Discrepancies are not only about how the information is broken down –i.e. if the complete name is represented in one or two fields– but also in other content. For example, while femdat has a field for mother tongue and other languages, the CEC-WYS database has no fields for the languages spoken by the data contributor. Kilden puts a great emphasis on publications, which allows the searching users to look for publications from specific persons, but no other database offers this service explicitly.

We now discuss the fields which are meaningful for the DATAWOMSCI Service Provider given the fields in the existing databases:

**name:**

The field "name" will be displayed in the DATAWOMSCI Service Provider. Since some of the databases do not distinguish between first and last name, name will represent the complete name in the feasibility study. See Table 4 for a listing of the existing name field in each database. It is still recommended to keep the family and first names of persons separated in the future.

Table 4: The name fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	name	vorname	first name	firstname	first name
		nachname	familyname	lastname	surname

Privacy plays a role in FemConsult and CEC-WYS. This means that the data contributors have the right to keep their names hidden from the searching public. In such a case the names of these persons can not be searched in the database. Instead, a solution for hidden names for the DATAWOMSCI Service Provider is discussed in Section 6.1.4.

**country:**

The field country does not exist in all the databases but we shall analyze it here since it is a field which we recommend. In FemConsult the country is implicit in the address of the data contributor, therefore is currently not available. femdat has two country fields. One refers to the country of residence, while the other refers to the country of the workplace. The same holds for the Kilden database. In CEC-WYS the field is not well defined, but seems to refer to country of origin, since the database covers four different countries. See Table 5 for an overview of the existing and relevant country

fields. The complete address of the contributors will not be included in the DATA-WOMSCI Service Provider.

Table 5: The country fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	country		institution. country	person. work- country	country

The country field would allow searching users to do a general search on scientists on a geographical basis. Country should refer to the country in which the person is currently working. We discourage the use of nationality as an alternative field. The country field is recommended although under the current conditions it may be difficult to implement.

#### email:

The field email exists in all the selected databases. In CEC-WYS the contributor may decide to hide her email from the general public. If this information is hidden for privacy reasons one of the solutions discussed in Section 6.1.4 can be considered. An overview of the field email is given in Table 6. We recommend the use of the field email.

Table 6: The email fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	email	email	email	person. work email	email

#### institution/organisation:

Institution or organisation is included in all of the databases. In the case of FemConsult this data is collected in two fields, "Inst1" and "Inst2". These two fields are to capture a hierarchy if it exists, like Inst1: University of Bremen, Inst2: Department of Computer Science. These fields are hidden from searching users of FemConsult for privacy reasons.

Table 7: The institution fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	institution	Inst1	Institution. Name	person employer	organisation
		Inst2		occupational experience institution	

The privacy regulation of FemConsult in this case results in a conflict of interests. If we include the institution field and FemConsult institutional information is not included, a search on the "institution" field will exclude all scientists listed in FemConsult. If FemConsult decides that this is important, then CEWS needs to contact all data contributors of FemConsult to ask for their permission to release these data. One last possibility is to exclude the "institution" field from the DATAWOMSCI Service Provider, in which case, there is increasingly less data to search for. In such a case, it is worth considering if it makes sense to build a database like the DATAWOMSCI Service Provider.

A socio-political aspect of this field is unemployment. The data contributors may have difficulties filling out this field considering current unemployment among women academics in Europe. As seen in Table 7 Kilden provides a solution to this problem. They include the current institution of employment, but they also include a list of the institutions as a part of the "occupational experience" field. SET has an employment history field, CEC-WYS an experience field, and femdat a professional experience field which are similar. Unfortunately, these fields usually contain more than just the institutional information and therefore are inadequate.

Regardless of all these problems, we recommend the inclusion of an institution field.

**qualification/discipline:**

Defining this field is difficult. Professional qualification is not necessarily an academic degree or discipline. The definition of the "professional qualification" field in SET says:

*[Professional qualification] includes sponsored qualifications from a company, training courses, degree, etc Please include: course title ...*

As far as we could interpret, the "degree.academic subject" fields in the Kilden database clearly represent academic discipline tied to a specific degree. femdat, FemConsult and CEC-WYS define these fields as academic fields and are to be selected from a predefined list of disciplines. A similar field exists in SET as "keywords" with a list of disciplines. See Figure 8 for an overview of discipline and qualification related fields. The semantics of this field are a matter of further discussion among the partners. The existence of such a field for search is essential and therefore is recommended.

Table 8: The qualification/discipline fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	discipline	discipline	training	degree. academic subject	professional qualification

**experience:**

The "experience" field is included in all the selected databases other than FemConsult. FemConsult has two additional fields: "Experience"(Schwerpunkte) and "Further disci-

plines and experience" (Weitere Fächer/Schwerpunkte). The first one is selectable from a predefined list of disciplines and the latter is free text. Table 9 gives an overview of existing fields related to experience.

femdat breaks down the experience fields to the following: further education, fields of specialization, experiences (selected from a list of values), and professional experience. It is a matter of further negotiation if "experience" should refer to "qualifications within the field, regardless of employment" or if the emphasis is to be on solely on past "employment". The prior definition would include activities related to the content of the expertise, whereas the latter would emphasize the various organizational roles the scientist had during her employment i.e. decision making, project management etc.

Table 9: The experience fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	experience	emphases	experience	occupational experience description	application of expertise
field name		further disciplines/emphases	fields of specialization		

Currently, in all databases experience can refer to work experiences which also include non-discipline related qualifications like management, teaching, mentoring etc. Despite the discrepancies these fields can be summed up under the field experience within the DATAWOMSCI Service Provider. We recommend the use of the experience field.

#### publications:

The field publication exists in all the databases other than in FemConsult. Table 10 gives an overview of these fields in the selected databases. Since the DATAWOMSCI Service Provider is about women in research and science it is highly recommended to include publications. The problem here is similar to that of the country and institution fields. Since FemConsult does not include publications, a search on publications would exclude all women in the FemConsult database. We recommend the use of the field publications.

Table 10: The publications fields

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	papers	N/A	publications	all publication fields (12 in total)	publication

Kilden is the only database that stores all the publication details (date, authors, isbn etc.). It is highly recommended in the future to use the MARC library standards. Further information on this standard can be found in the CERIF 2000 toolkit described in Section 7.2.5. The use of such a standard would contribute to a further service in women scientists' databases of the future: the ability to search for publications by women in Europe.

### 6.1.2 *Further fields*

Further fields which are implemented in the five databases are too divergent or too detailed for use in the DATAWOMSCI Service Provider. For example, the field "degree", if we defined it to be the educational level and not the academic discipline, could be very difficult to implement. The degrees available within the German higher education system vary greatly from that of the United Kingdom education system which probably differ from those in the Czech Republic. For example, which of the following keywords would be appropriate for searching a scientist with a doctorate degree: PhD, Dr., Doctor, Doctoeur, Doktorand etc.

Clearly, some of the fields analyzed in Section 6.1.1 have similar divergences in content, and with the agreement of the institutions further fields can be standardized in the future. For the purpose of this study, we remain only with the common fields listed in Section 6.1.1.

The DATAWOMSCI Project Consortium partners emphasized the importance of the "language skills" field for women scientists' databases. In addition, they recommended a distinction between "qualification" which sometimes is also called "academic degree" and the disciplines the women scientists work in. Both of these recommendations are key to the quality of a women scientists database, although in our analysis of the five databases the implementation of both these fields in the DATAWOMSCI Service Provider is currently not possible.

Further fields may be defined as important and negotiated with the administrators and maintainers of existing databases. We identify the extension of the list of fields as a task beyond the limits of the feasibility study and one which rests on future negotiation among the interested institutions.

### 6.1.3 *Redundancy of Data*

The redundancy of data can occur if a data-contributing woman scientist registers in two or more databases. Unambiguous identification of persons is highly difficult since the fields are not standardized. Even if all the fields were standardized with predefined lists a data contributor could give slightly different information, making it impossible to identify redundancy automatically. In addition, if a data contributor has registered in two or more databases, it will not be easy to identify which set of data are most up-to-date or most accurate. Because of all these difficulties the possibilities of redundancies will be ignored. If in the future, redundancy of data becomes a problem then institutional as well as technological solutions will have to be considered.

### 6.1.4 *Hidden data*

Some of the databases hide parts of the records on persons for reasons of privacy. Respecting the privacy needs of the data contributors should be a priority of the DATAWOMSCI Service Provider.

Hidden data may be confusing to users who are searching for scientists. There needs to be a clear explanation for users searching the DATAWOMSCI Service Provider about hidden data. If scientists with hidden data appear in a search result then the name of the database where the scientist is originally registered as well as the email of the institutional person responsible for such contacts should appear.

It is rather desirable that the searching users avoid manual work. Ideally, the user can select a set of scientists whose data are hidden, like the "selected scientists shopping cart" of FemConsult. The system would then deliver a message to the respective organizations with the information of the selected scientists and a short notice from the searching user. The implementation of such a service may be costly but would cause user satisfaction for data contributors as well as those searching scientists. It would also prevent the possible complications which may occur while relating ambivalent summaries of search results to the institutional person.

#### 6.1.5 Character Encoding

An overview of the character sets used by the databases is given in Table 11. Other than SET all databases have confirmed that they could export the necessary data in XML format using Unicode. In this case, the problems with the differences in character sets are taken care of and cease to be an issue for the DATAWOMSCI Service Provider. Details on the character encoding problem are listed in Section 7.2.1.

Table 11: The Character Sets used by the Databases

	CEC-WYS	FemConsult	femdat	Kilden	SET
field name	ISO-8859-2	ISO-8859-1	ISO-8859-1	ISO-8859-1	UN

#### 6.1.6 Security

Since adding and updating for data contributors remain solely with the original databases, there are no password related security issues for the DATAWOMSCI Service Provider. There is no reason to introduce passwords for searching in the DATAWOMSCI Service Provider. If additional services are defined then it may be necessary to introduce membership or passwords. See also Section 5.2.3.

The security of the DATAWOMSCI Service Provider server is imminent. The services should be secured against breaches of integrity of data and of the availability of the server.

#### 6.1.7 Data Protection Act

The use of data in the DATAWOMSCI Service Provider is the use of the data for a purpose other than originally intended and requires the permission of the data contributors. We recommend the formulation of a legally well-formed letter which adheres to

European as well as local Data protection acts to be sent out to the data contributors. It may be interesting to find out if it is legally and technically possible for contributors to give consent online to the use of their data within the DATAWOMSCI Service Provider.

#### *6.1.8 Future Changes*

Four of the databases are in a state of transformation. CEC-WYS is a freshly built database still under construction and currently receiving fresh registrations. The SET database will be completely renewed in the coming months and will be limited to women from the UK. The femdat database will migrate to a new open source middle-ware in the year 2005. FemConsult is planning a revision of their database services by the year 2005/2006. Plans at FemConsult include an English version of the database as well as a new privacy policy in which current anonymous listing practice will be discontinued.

Certainly these transformations will have effects on the data models of the existing databases. Still, we hope that the semantics of the fields will not change tremendously or along the lines of our recommendations. In all cases, other than the SET database, we have informed the administrators of the databases of the CERIF 2000 toolkit, described here in Section 7.2.5.

The good thing about the OAI-PMH is that underlying structural changes, like the restructuring of the data model, do not affect the application, as long as fields semantically corresponding to the fields listed in Section 6.1.1 remain. Therefore, we expect the future changes to cause some work in reproducing the XML export of the databases with the right fields, but not to affect the technological feasibility of the DATAWOMSCI Service Provider.

## **6.2 Conclusions on the feasibility of the DATAWOMSCI Service Provider**

There are many issues to be cleared, especially concerning the fields of the DATAWOMSCI Service Provider, which are beyond the scope of this study. We have shown technically that fields with similar information exist and that these could form the basis of a DATAWOMSCI Service Provider. The negotiation of these fields and finding appropriate names for the fields on the DATAWOMSCI Service Provider are assignments for the future. Once these fields are agreed upon, the technical possibilities exist to harvest the selected databases.

Harvesting the data collections is one side of the coin, producing meaningful services within the DATAWOMSCI Service Provider is another. In Section 7 we analyze the possibilities for multilingual information retrieval. In Section 11 we shall analyze interface issues like multilingual interfaces, display of search results, help documents, and further end-user services.

## 7 Analysis and the Evaluation of Cross-Lingual Information Retrieval for the DATAWOMSCI Service Provider

In this section we evaluate existing technologies for cross-lingual information retrieval. We first describe what we mean by cross-lingual information retrieval and then list existing methods.

### 7.1 Introduction to cross-lingual information retrieval

With cross-lingual information retrieval we mean searching document collections in one or more languages other than the language of the search. We shall now describe the complexities of cross-lingual information retrieval in the context of the DATAWOMSCI Service Provider.

Ideally, once harvested using the OAI-PMH, the DATAWOMSCI Service Provider will contain data in several languages. According to our study, the languages used in the five databases that come into question are:

- English
- German
- Norwegian
- French
- Italian

In the future it is imaginable that as more databases are harvested using the OAI-PMH further languages are included in the DATAWOMSCI Service Provider.<sup>1</sup>

Let us imagine an ideal DATAWOMSCI cross-lingual information retrieval scenario:

An English speaking user searching for a "marine biologist" starts a search within the DATAWOMSCI Service Provider. She enters as keywords "marine biology". The search engine queries the database using the keywords "marine biology" as well as the translation of "marine biology" in German, Norwegian, French and Italian. The search returns all the relevant results regardless of the language of the content. The resulting entries which were originally in one of the languages other than English are translated into English.

There are two aspects which are central to the cross-lingual information retrieval in this scenario:

1. the user enters keywords in his/her own language yet all entries that contain a translation of the keywords are also searched
2. the user is able to make use of the returned results which were originally in another language, since the content is translated.

<sup>1</sup> The CEC-WYS project is planning on having all their data entries restricted to English. Currently they have migrated the content of the Zenyaveda database to the CEC-WYS database which is mainly in the Czech language. This data will be excluded from the database in the near future.

Unfortunately, to this day few technological tools come close to providing this high-quality service. Refer to the CELEX and EURLex projects of the European Union for public projects that come close. Accurate translation of natural languages through computers is complex and costly and is the topic of current research. The same holds for the development of [cross-lingual] information retrieval systems.<sup>[23]</sup>

Yet, there are more moderate solutions of cross-lingual information retrieval, which do not deliver such high-quality services but offer partial solutions to the problems that arise with cross-lingual information retrieval.

## 7.2 Cross-lingual information retrieval methods

In the following we introduce some of the existing solutions for cross-lingual information retrieval and evaluate the usefulness and feasibility of implementing these technologies within the DATAWOMSCI Service Provider.

### 7.2.1 Transliteration

#### The method:

The data in a database are always encoded according to a character encoding standard. We mention the different character encoding standards used by the selected databases in section 6.1.5. The use of different character encodings usually causes several problems:

- search problems: which character encoding will the search query be in?
- search problems: how will relevant data collections be searched properly when they are encoded in numerous character encodings?
- problems displaying results: if the browser is unable to automatically detect the character encoding of the search results, the results may be displayed erroneously or not be legible at all.

#### Transliteration:

The process of converting the characters of a script, could be used to transfer all the characters to one character set which is then searchable. The drawback of transliteration is that this technique leads to loss of information.<sup>[23]</sup> Examples of such losses are accents, umlauts and other language-specific mark variant forms of words.

#### Feasibility for DATAWOMSCI:

The OAI-PMH harvests data in XML from all the data providers. All XML documents whatever language or writing system they employ, use the Unicode character encoding. Unicode is an international standard for representing the thousands of discrete symbols making up the world's writing system. Hence, if the data providers are able to provide the XML Unicode output necessary for harvesting, the character encoding problem ceases to be an issue for the DATAWOMSCI Service Provider.

It should also be noted that the pages need to be served in Unicode (UTF-8) so that the search masks' fields also do not have to be transliterated.<sup>[18]</sup> Unicode may still represent a problem for machines that do not accommodate this standard. We expect this rarely to be a problem.

If an alternative to the use of XML is selected, then the appropriate transliteration standards will be implemented for the records in the DATAWOMSCI Service Provider. An international forum, ISO/TC46/SC" (International Technical Committee 46/ Subcommittee Conversion of Written Languages) creates standards for conversion methods of writing systems.<sup>2</sup>

### 7.2.2 Dictionary-Based Approach

#### The method:

In the dictionary based methods queries are translated into a language in which a document may be found. Through the ambiguities of natural languages these methods sometimes achieve unsatisfactory results. Many words in natural languages do not have only one meaning and the alternative translations may have very different meanings. In addition the scope of dictionaries is limited, especially when it comes to technical and topical terminology. According to research, electronic dictionary based approaches for query translation have achieved an effectiveness of 40-60% in comparison with monolingual retrieval.<sup>[23]</sup> One possibility of overcoming these weaknesses is to combine this method with other types of methods.

#### Feasibility for DATAWOMSCI:

There are three reasons why we believe that the dictionary based methods are not adequate for the DATAWOMSCI Service Provider. First of all, the effectiveness of dictionary based methods applied alone are too low. Second, implementing a dictionary-based method in combination with other methods can be costly and complex. Third, dictionary-based methods are especially weak when it comes to technical and topical terminology, which make up the main body of the content that will be included in the DATAWOMSCI Service Provider.

### 7.2.3 Corpus-based Technologies

#### The method:

In the corpus based methods, large collections of existing texts (corpora) are analyzed and information on which the translation will be based are extracted. Corpora are collections of information in electronic form. These collections are then used to support spelling and grammar checkers, hyphenation routines or in order to evaluate the performance of cross-lingual information retrieval results.

---

<sup>2</sup> <http://www.elot.gr/tc46sc2/>

A collection may contain parallel and comparable corpora. A parallel corpora is a collection which may contain documents and their translations. A comparable corpora is a document collection in which documents are aligned based on the similarity between the topics which they address.

**Feasibility for DATAWOMSCI:**

Since the documents within the DATAWOMSCI Service Provider are in their content diverse and are not made up of full text but rather keywords, the application of corpus-based technologies is not recommended.

*7.2.4 Indexing by Latent Semantic Analysis (LSI)*

**The method:**

Latent Semantic Indexing methods allow users to retrieve documents by concept and meaning and not only by pattern matching. For example, if the query words have not been exactly matched, this does not mean there are no relevant documents in the data collection. Usually, the opposite is the case. There are many relevant documents which do not contain the query term. This is the problem of **synonymy**.

Computer scientists for example formulate their queries different than linguists. Documents can not contain all the terms that users may enter as a query. Using thesauri to overcome this issue remains ineffective, since expanding a query to unsuitable terms decreases the precision drastically.<sup>[23]</sup>

In the Latent Semantic Indexing method a matrix of terms and documents are produced which allows the analysis of "semantic neighbourhood." The result of LSI analysis is a reduced model that describes the similarity between term-term, document-document, term-document relationships. So once a query word is found, documents in the semantic neighbourhood may also be returned as relevant documents.

LSI has been shown to yield better results than pattern matching. Some experiments have shown that LSI achieves 75% average precision in comparison with a monolingual system on the same data collection.<sup>[23]</sup> In this way, LSI methods are highly promising. The drawback is that adding new terms and documents have proven to be time consuming and costly.

**Feasibility for DATAWOMSCI:**

The latest and most experimental multi and monolingual information retrieval LSI is exciting but costly. The application of LSI to the DATAWOMSCI data collection in cooperation with a research group which is interested in indexing documents on scientists may be a solution. Certainly the Language Technology research group at the **Institute for the Protection and Security of the Citizen (IPSC)** or the **Current Research Information Systems Project (CRIS)** of the EU could be such partners. The results would then benefit the DATAWOMSCI Project as well as these projects, the former being

interested in applying their latest research in cross-lingual indexing, and the latter interested in Research Information Systems.

Without such a partnership it is not recommended to use LSI methods on the DATA-WOMSCI Service Provider. Some commercial software exists that offer custom-tailored services. If this is of interest then a market analysis of existing products needs to be made. We did ask one company for an estimate, and the suggested price was approximately 8000 Euros plus 17% of the price annually for support and updates.

### 7.2.5 *Controlled Vocabulary*

#### **The method:**

In the method of controlled vocabulary terms are assigned to all documents in the data collection. Descriptors are used to represent documents. The terms are translated and mapped to each other in thesauri. The complexity of a thesaurus emerges since the translation of terms is not isomorphic meaning a term may have many translations and may mean different things in different contexts. For example the word "free" may mean "liberty" but as well "with no cost or charge", or the word "bar" may mean a place to drink, but also "lawyers collectively".

Thesauri can be built manually or automatically. In order to generate a thesaurus automatically documents are collected, classified and organized. The semantic relationships between the terms (i.e. synonyms, neighbourhood etc.) are deduced. In multilingual thesauri, terms in different languages are mapped onto each other according to language mapping standards.

The controlled vocabulary method remains limited to applications whose vocabulary is manageable. Once the size of the vocabulary increases, the efficiency and effectiveness degrade radically. The approach also requires the indexing of every document using a selected vocabulary. This could be done by the users entering the data or by a qualified person. But even then, the persons entering their data and those searching the data collection are restricted by the categories available in the thesaurus. This method hence lacks flexibility and does not allow for free text search.

#### **Existing applications:**

Numerous projects intensively work on standards for multilingual thesaurus development. The European Union initiated the EUROVOC project in order to build a multilingual thesaurus based on the established standard ISO 5964. The EUROVOC thesaurus supports 11 of the official European Union languages.<sup>3</sup> Unfortunately, technical and scientific terms, especially vocational terms are not a strength of the EUROVOC thesaurus.

Another similar project is the multilingual term bank Eurodicautom. The project is not a thesaurus but could be useful in developing one since it is particularly rich in technical and specialized terminology related to EU policy.

<sup>3</sup> <http://europa.eu.int/celex/eurovoc/> the languages that are currently included are: Spanish, Danish, German, Greek, English, French, Italian, Dutch, Portuguese, Finnish and Swedish

The CERIF initiative makes use of these two projects. Now a subproject of CRIS, the **Common European Research Information Format (CERIF)** is a set of Guidelines, Code of Good Practice, formats and a set of software tools to develop and maintain Current Research Information Systems (CRIS). CERIF was developed

1. to improve accessibility and comparability of research information from various sources
2. to assist in CRIS development and maintenance
3. to help the use of research information

In order to reach these goals the CERIF 2000 Toolkit offers data models and controlled vocabulary. The controlled vocabulary is based on international standards and resources like the EUROVOC, Eurodicautom, and most valuable of all the ORTELIUS thesaurus. The ORTELIUS thesaurus offers a classification of higher education in Europe. In addition, classification schemes for Economic Activity and Products by Activity are included.

Controlled attribute value lists are also offered by the CERIF 2000 Toolkit in order to allow for a guided look-up of elements in their data model. The existing lists are presented in Table 12.

Table 12: List of CERIF Standards

• Language	• Country
• Currency	• Address
• Role of a person in an organisation	• Qualification of a person
• Organisation/Company size	• Type and status of a patent
• Type of publication	• Type of event
• Type of multimedia item	• Role of an organisation in a project
• Role of a person in a publication	• Role of a person as a scientist
• Type of organisation	• Role of a person related to equipment
• Role of an organisation related to a product	• Type of equipment

#### Feasibility for DATAWOMSCI:

We believe that from the existing methods for cross-lingual information retrieval the solution that is most likely to be adapted to the DATAWOMSCI Project is the controlled vocabulary method. Yet, the implementation of a thesauri specific to the DATAWOMSCI Service Provider may be too costly within the context of this project. At the same time, it is a meaningful and innovative application field for this method and may lead to an interesting cooperation between existing European Union projects which are researching cross-lingual information retrieval.

For example, within the Language Technology research at IPSC, one of the Joint Research Centres of the European Union work is carried out on document retrieval, text analysis and information extraction, as well as visualization of textual information. One of their services includes the development of cross-lingual thesauri. In addition,

in their presentation they state their interest in cooperating with other EU projects by applying the technologies they are developing for cross-lingual information retrieval. Such a cooperation may be fruitful for the implementation of the DATAWOMSCI Service Provider.

A more realistic solution is the use of the CERIF 2000 Toolkit. The CERIF 2000 Toolkit is meant for the compatibility of **Current Research Information Systems (CRIS)**. The data model included in the CERIF 2000 Toolkit could be used as the underlying data model of the DATAWOMSCI Service Provider, which would achieve interoperability with other CERIF data model compatible projects. The controlled vocabulary could be used as descriptors of the multilingual documents.

There is one important obstacle, namely the CERIF 2000 Toolkit includes the 11 languages of the European Union, which represent the languages before the expansion of the EU. This means that this Toolkit does not provide support for Norway or any of the countries that entered the EU May 2004. This needs to be taken into consideration in the design of the DATAWOMSCI Service Provider. Other search possibilities will need to be studied for databases with data collections in languages not included in the CERIF 2000 Toolkit.

Another incompatibility of the CERIF 2000 Toolkit is that it is actually centred around projects rather than persons. Therefore, the CERIF data model needs to be analyzed and those fields which are relevant to DATAWOMSCI need to be determined and the model accordingly extended. This analysis process could be enriched by using participative requirements elicitation methods, to find out which fields should be mandatory and which fields are desired in women scientists' databases across Europe.

In addition, for women scientists' databases which are now being planned, it is recommended to use the CERIF 2000 Toolkit in order to develop future women scientists' databases which are compatible with European standards. The CERIF data model can be applied as it is, or the mandatory fields can be included and further fields can be added locally. The use of the CERIF model is currently a precondition for data exchange and information harmonization in Europe. A short analysis of the CERIF 2000 Toolkit with respect to future women scientists' databases and the DATAWOMSCI Service Provider is made in Section 10.2.<sup>4</sup>

Last but not least, CERIF 2000 Toolkit offers standards, but these may be incompatible with local standards, like the one used in femdat for Switzerland's classification of academic disciplines. Therefore, the use of standards needs to be negotiated with the partners and can only take place with their acceptance.

---

<sup>4</sup> The ERGO project is supposed to give an overview of CERIF compatible CRIS projects. Unfortunately, their search facilities were not working during the period of this feasibility study. <http://www.cordis.lu/ergo/>

## 8 Introduction to International Web Sites

An international web site is a web site which has an international audience. It may be monolingual or multilingual. In addition, the sites in the different languages may be adjusted to the cultural contexts of the languages of the different audiences.<sup>[2]</sup>

When we speak of culture, we don't refer to the "ethnic" culture, but rather cultures of communication, technology, organization, science etc. An example here is the order of the day, month and year in writing down a date, in which case 01/02/03 could mean three different dates, depending on the "cultural" context. Or the word "science" may only refer to the natural and technical sciences, whereas in other contexts refer to all disciplines, as in the German "Wissenschaft".

In this Section, we will briefly observe the advantages and disadvantages of each: monolingual international sites, multilingual international sites, and culture-sensitivation of sites. We will refer to content that is culture-sensitised as "changed content," and where this is not done, as "same content."

Decisions in multilinguality and internationalization will have effects on file structure of the site, on the software and interface requirements, and hence on the hardware. For example, a multilingual site may allow the users to navigate through languages, whereas an international site may allow the users to navigate through countries or geographic regions. Therefore, it is extremely important to be aware of these decisions at the very beginning of the implementation of the project.

The W3C, the standards organization of the internet makes the following distinctions among web sites:

- monolingual international site
  - monolingual changed content
  - monolingual same content
- multilingual international site
  - multilingual changed content
  - multilingual same content

We will shortly introduce these categories and analyze their advantages and disadvantages.

### 8.1 Monolingual international site

Web sites that present the same monolingual content to an international audience may be acceptable for some communities (i.e. technical communities) with an agreed standard language.<sup>[26]</sup> In other communities, with varying languages and standards, the language picked for the site needs to be kept simple and should be checked for culturally troublesome content.

Nevertheless, the use of simple language may result in "cultural blandness" affecting the quality of the site negatively. In any case, building a monolingual site same content

site risks alienating users who do not understand the content or can not relate to the cultural specificities of the content.

**monolingual same content:**

Monolingual same content sites may be the only option as a result of the costs and complexities of multilinguality and internationalization. In such a case, the site may include help services for the users, for example through the availability of permanent key texts (i.e. introductory texts or help texts) in multiple languages.

**monolingual changed content:**

Some international sites, although they remain monolingual, vary their content according to the needs of the "local" audience. For example, it may be plausible to arrange the content according to the different legal requirements of the different locations. This approach is more costly and complex, but still may result in more user satisfaction and less confusion based on content and context.

*8.1.1 Recommendations for monolingual sites*

If monolingual same content is the choice, then key texts, like the introduction of the site as well as help documents should be available in multiple languages. If monolingual changed content is possible, then a further study needs to investigate the needs of the different user groups based on location, language, technological and communication culture, etc. to find out where variations are possible.

## 8.2 Multilingual international site

There are many cultural and regional differences in the way information is presented. Earlier we gave the example of dates. Clearly, the naming of the different fields in a search mask, or the language used on the different navigation buttons are all a part of this decision-making process in multilingual pages.

**multilingual same content:**

Sites that provide multilingual content address understandability, but not necessarily cultural differences or relevance.<sup>[26]</sup>

Here, issues of translation need to be attended to. Hillier (2002) lists some of the issues that come up during translation:

- agreement regarding the meaning of words: in many languages the same word may mean different things. In English for example the word "trunk" means in an Australian context "a large box" whereas in the US American context it is the luggage compartment of a car.

- agreement on terminology: this is especially true for terminology of technical nature. In English the abbreviations "No" and "" can mean "number" but in Chinese there are no such abbreviations. In naming fields of a search mask or a database, especially if the use of abbreviations is considered, such differences need to be accounted for.

Certainly there are further issues which have not been listed here. With multilinguality the issue of character codes for example also becomes central. Technology needs to be picked to deal with the different character encodings, and if possible to allow for multilingual pages.

Although numerous technologies have been developed, for some languages the character encoding schemes are still in development. Those technologies that are developed sometimes require computers which are state of the art with the most contemporary operating systems.<sup>[3]</sup> Since the audience of a future women scientists' databases is diverse in technology and will not always have the latest computer technologies available, decisions on such issues are important.

#### **multilingual changed content:**

Especially in sensitive fields, adapting to the cultural and contextual needs of the different user groups may pay off. Often, content adaptation also increases relevance to the users. Again, this decision needs to be weighed against the costs and complexities of implementation. Studies on the contextual needs of the users are recommended, in order to find out which contextual needs are relevant to the project.

Formats of such things as dates, times and names require attention. In some countries it is the case that people first write their family name and then their first name. This can be different in other countries. There may also be different sensitivities with hierarchies, organisational or personal presentations. Therefore, the representation of such information needs to be identified before the development of future women scientists' databases.

#### *8.2.1 Recommendations for multilingual sites*

Multilinguality requires extra expertise. First of all, it requires persons with knowledge of the languages included in the site. In addition, it benefits from expertise or studies on the contextual cultural needs of the audience. In addition, it requires an analysis of work processes that are necessary to maintain the site. In computing language the results of such analyses are called "workflows". Workflows need to be developed for processes of creation of pages (conceiving original text, translation, drafts, adding to the site under the appropriate navigation points, keywords etc.) and maintenance of pages (if changes are made in one language, the same content also needs to be updated in the other language).

As a result of such complexities, many sites introduce multilinguality at a minimal level. They only translate central static/non-changing pages and keep the rest in a common language. This decision depends on the resources of a future DATAWOMSCI Provider which may be financial (to buy multilingual services) or in cooperation with

the different existing institutions which are currently maintaining women scientists' databases.

In Section 9 we shall analyze alternative solutions for a DATAWOMSCI Provider. In all three cases, we shall consider the recommendations for monolingual and multilingual sites relevant to the DATAWOMSCI Project.

## 9 Three DATAWOMSCI Provider Solutions

At the beginning of the DATAWOMSCI Project three alternatives were introduced for a future DATAWOMSCI Platform. The alternatives were:

- federated solution: integrate a small number of best-practice databases with central search and navigation services. Link the rest of the data collections in a systematic way.
- distributed solution: prepare a structured link list, possibly with further supporting services
- central standardized solution: develop a central data collection solution with a specific definition of women scientists and a data model for a database of women scientists. Existing data collections migrate and newcomers adhere to the standard solution while developing data collections.

The "federated solution" is the underlying concept of the **DATAWOMSCI Service Provider (DSP)**, the feasibility of which we studied in Section 6 and later in Sections 9.1 and 9.2. The "distributed solution" is a database of databases or a register of databases. This solution we shall call the **DATAWOMSCI Information Provider (DIP)**.

The last solution, the central standardized solution which we suggested at the beginning of the project we would like to vary. We do not find the development of a central database meaningful. A central database would lose the heterogeneity and richness of the existing databases. In addition, it would lose the effects that local databases have on local and global communities. A central standardized solution would in addition cause duplications as women would want to register in local databases as well as in a Europe-wide database.

Instead we suggest bilingual registration. In the **bilingual registration** model, the women continue to register in the local databases. In addition to the fields that they have been providing so far, they also fill in a few extra fields in a common language, most likely in English. Only these fields are then harvested for the DATAWOMSCI Service Provider. This solution we shall call the **Monolingual DATAWOMSCI Service Provider (MonoDSP)**.

Since we dismiss the central solution and recommend a Monolingual DATAWOMSCI Service Provider as an alternative, we shall be evaluating two federated solutions and one distributed solution. These can be listed as follows:

- the multilingual federated solution: Multilingual DATAWOMSCI Service Provider (MultiDSP)

- the monolingual federated solution: Monolingual DATAWOMSCI Service Provider (MonoDSP)
- the distributed solution: the DATAWOMSCI Information Provider (DIP)

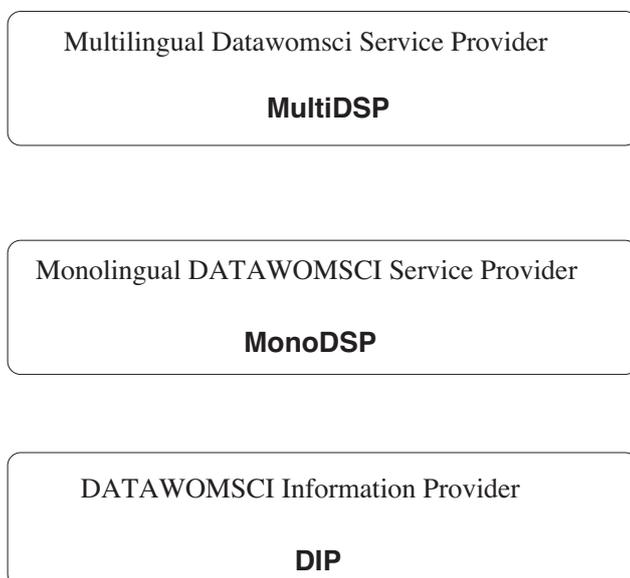


Figure 5: The three DATAWOMSCI Provider Solutions.

Figure 5 gives an overview of the three alternatives. In the coming sections we study these three models with the extra services that could be offered with each model.

### 9.1 The multilingual federated solution: Multilingual DATAWOMSCI Service Provider (MultiDSP)

We studied the feasibility of the multilingual federated solution in Section 6. It was also evident from the questionnaires received from the administrators of the databases considered for the feasibility study, that the harvesting of the datasets from the different databases would be technically possible.

The next question is if the multilinguality of the harvested datasets should also be reflected in the services of the multilingual DATAWOMSCI Service Provider or shortly the **MultiDSP**. There are various alternative models to a MultiDSP. We shall consider three of these here:

- MultiDSP with a monolingual interface
- MultiDSP with multilingual interface
- MultiDSP with cross-lingual search services

These models can be developed as follows: a multilingual DSP may either have a monolingual or multilingual interface. In both cases, additional cross-lingual search services can be implemented, as shown in Figure 6. We now go onto define and study these different models.

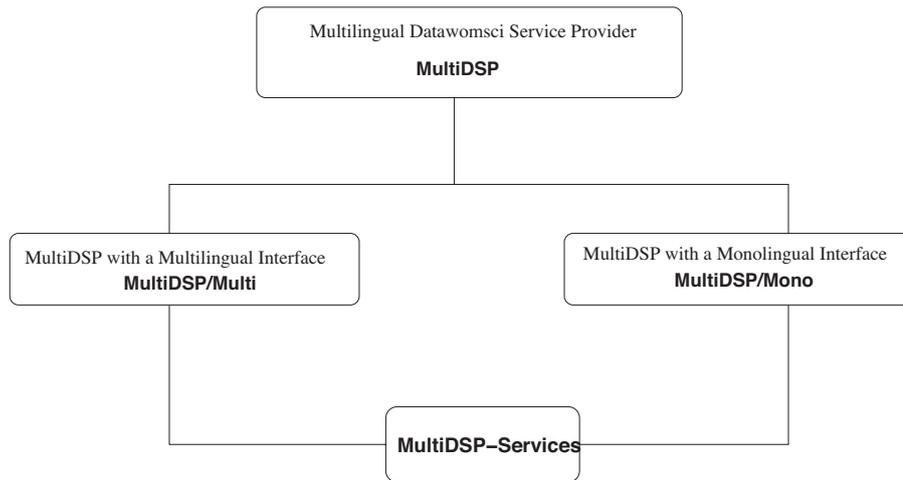


Figure 6: The multilingual DSP variations.

9.1.1 *MultiDSP with a Monolingual Interface (MultiDSP/Mono)*

The multilinguality in the multiDSP refers to the multiple languages of the harvested datasets. According to this model the MultiDSP is itself a monolingual site, most likely in English, and is called MultiDSP/Mono. This requires that the users of the **MultiDSP/Mono** are proficient enough to navigate and search in English. If the user wants to search datasets in different languages, then the user needs to have some proficiency in the language of interest.

The most important advantage of the central search mask of a MultiDSP/Mono is that users do not have to acquaint themselves with different search masks in different databases. In addition, they are able to search datasets from multiple databases at once. In the search results the users may identify the databases that the datasets are originating from.

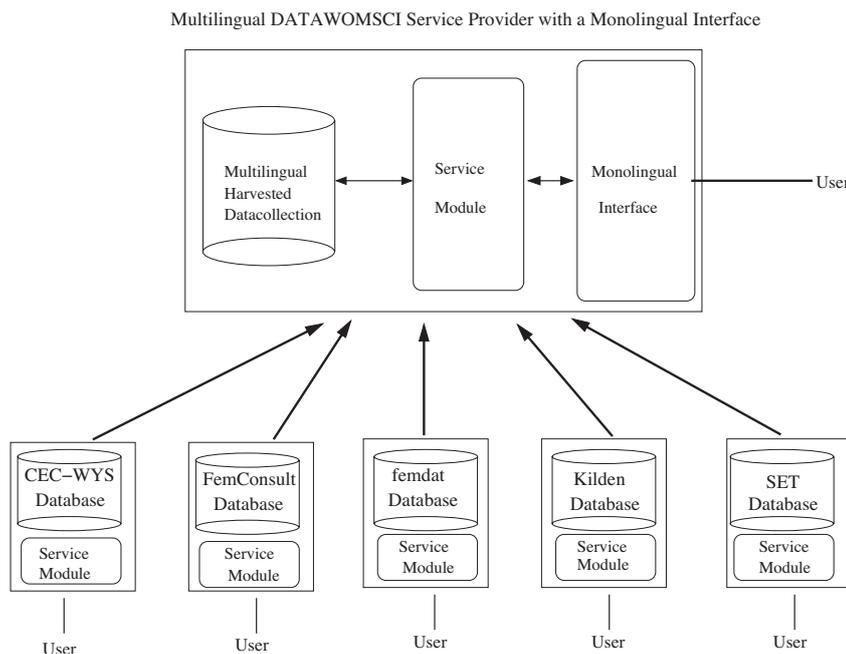


Figure 7: The MultiDSP with a monolingual interface.

The disadvantages of the MultiDSP/Mono are the limitations that it imposes on its users. A MultiDSP with monolingual navigation is only accessible to those who are proficient in the language selected for the site interface. In addition, unless the persons are multilingual or can find out the translations of the keywords they are searching for, they are not able to search numerous databases in multiple languages.

The advantage of the central search mask MultiDSP/Mono is that it is easy to implement and it is not extremely costly or complex. Despite all the limitations, it is a good gateway solution to the existing women scientists' databases in Europe. If a MultiDSP/Mono is of interest, attention needs to be given to the issues raised in Section 8.1 on monolingual international sites. In addition, some multilingual support can be provided to the users of the MultiDSP/Mono by translating the help texts on using the site into multiple languages.

### 9.1.2 MultiDSP with a Multilingual Interface (MultiDSP/Multi)

It is possible to implement the MultiDSP with a multilingual interface and multilingual search masks; this model is called the MultiDSP/Multi. The MultiDSP/Multi would be accessible to persons in different languages.

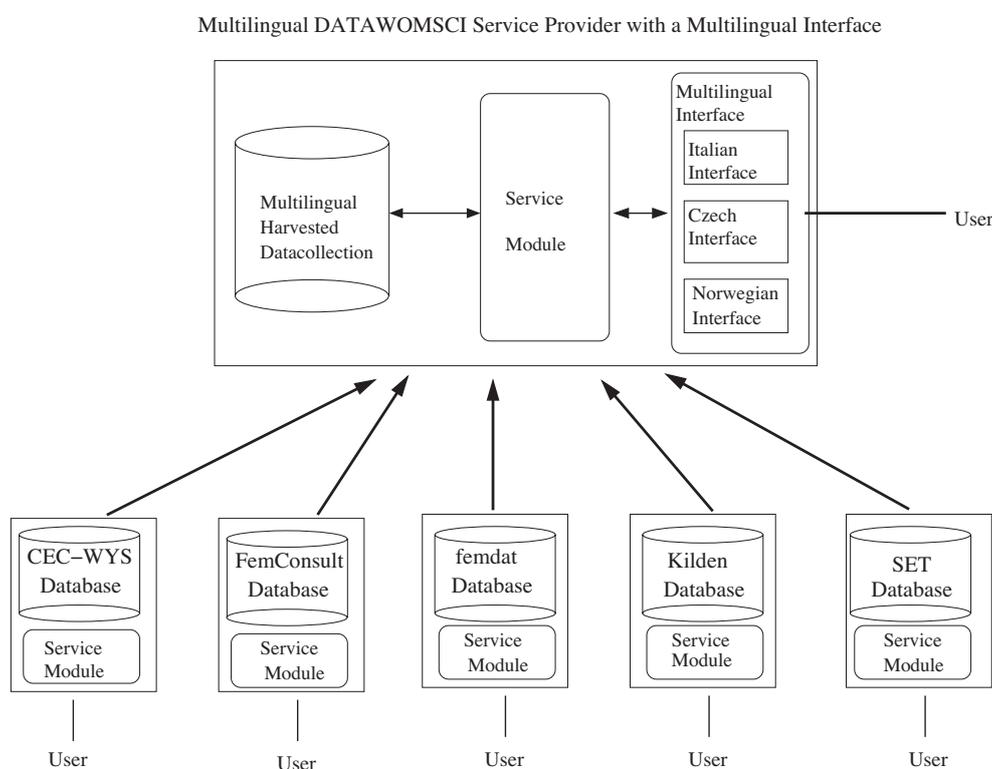


Figure 8: The MultiDSP with a multilingual interface.

The implementation of a multilingual interface and multilingual search masks brings up the issues of multilingual content which we described in Section 8.2. The maintenance of a multilingual site is complex and can be costly. We shall shortly list some of the main issues to be considered:

**translation:**

Translation is central to running a site with multilingual navigation. Qualified personnel need to be employed who assist the translation of the relevant texts, navigation trees, and the search mask fields or results. This is tedious work and can be costly.

**localisation:**

A multilingual website usually has global and local content. Local content is content relevant to a specific location, a specific geography or a specific language. For example, some of the information in Norwegian may include information specific to Norway's specific non-EU status and the different privacy laws there. On the other hand, the global content needs to be translated into all the languages, contains general information, which nevertheless needs to be adapted to the various differences between languages and cultures. Dealing with these subtle differences can be costly and time consuming.

**scientific, academic, and technological culture:**

Graphical conventions, socially acceptable forms of address and hierarchies, and issues of privacy all vary from place to place. These differences will have to be discussed during the design phase as well as when new content is added.

**feedback:**

Responses to any website feedback will need to be addressed in the language of the initial communication. User feedback should not be solicited in a language if it cannot be routed to a suitably qualified person, who can answer in the appropriate language.<sup>[41]</sup>

**design:**

Text or graphic labels that fit the design constraints in one language may not work well in translation. The same holds for navigation. The best way to tackle this issue is to discuss these issues in-depth in the design process, and to set some standards at the beginning of the project.

**workflow:**

Translation workflow requires that changes to a page trigger appropriate notification of required changes to the other language versions of that page.<sup>[41]</sup> A multilingual content management system, which integrates translation workflows is highly recommended.

**character-sets:**

We handled the issue of character-sets in Sections 6.1.5 and 8.2. The possibility of mixing languages on a single page should also be considered, allowing linking of other language versions of content to be handled simply.

Notice that, implementing and testing a multilingual site requires expertise in all the languages of the site. This may be very difficult to have in one space. Therefore, if a multilingual interface for the MultiDSP is of interest it is recommended to ask the exis-

ting institutions maintaining women scientists' databases how far they could support the multilingual interface. The support could be requested for maintaining the site or simply for responding to feedback from the users and communicating the feedback to the maintainers of the MultiDSP/Multi. Without such a support the quality or the cost of a MultiDSP/Multi may not be easy to handle.

### 9.1.3 MultiDSP with cross-lingual search services (MultiDSP-Services)

Cross-lingual search services can be added to both a MultiDSP with a monolingual or multilingual interface as shown in Figure 9. Cross-lingual search services allow the users to search documents by formulating their queries in a given language (called the **source or query language**) and finding results in another language (called the **target language**). The following is a list of possible cross-lingual search services:

- a document acquisition module which performs language identification, document classification and document summarization.
- indexing the content for cross-lingual search.
- translation of the user's query, with optional interactive disambiguation (sorting out ambiguous translations) and expansion.
- simultaneous search in data collections in multiple languages and the merging of the results.
- automatically translated document summaries.
- automatically translated search results.
- restriction of the search by language and thematic category.
- filtering of search results according to language and category.
- search by navigation through classification systems which return results in multiple languages.

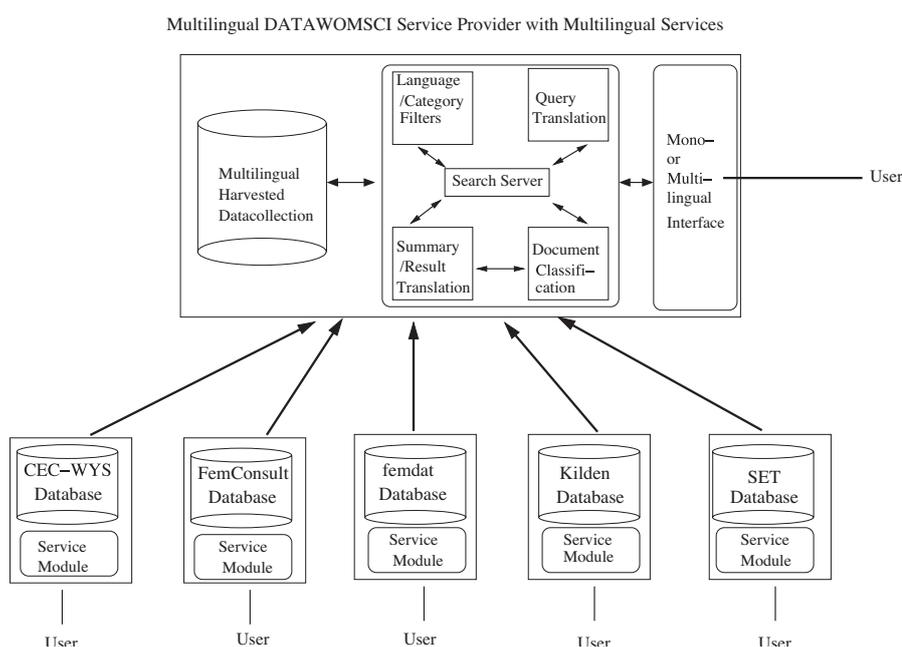


Figure 9: The MultiDSP with cross-lingual search services.

Some of these technologies we discussed in Section 7. Some of the interfaces for cross-lingual information retrieval we discuss in Section 11.

Cross-lingual search services are highly recommended for the European project DATAWOMSCI Service Provider. These services are costly to implement and require various kinds of expertise. It is highly recommended to survey the future users' multilinguality needs before implementing such costly technologies.

A participatory development model can be adapted to elicit user needs for the MultiDSP. As part of the development process of the DSP, a diverse group of users can be engaged to accompany the development of the project. These users are then asked to test the technologies that are in development and evaluate the importance of the different services. Prototyping, meaning producing a mock-up of the desired system to be tested by the users, is also a possible participation model. Without the analysis of the user requirements it is not recommended to implement complex and costly multilingual services.

## 9.2 The monolingual federated solution: Monolingual DATAWOMSCI Service Provider (MonoDSP)

The realization of the Monolingual DATAWOMSCI Service Provider (MonoDSP) requires that all local databases agree to collect data from their contributors in a common language. When registering, the scientists fill in extra fields in a given language, most likely in English. The data in English are then harvested by the MonoDSP. If a user searching the MonoDSP finds a scientist in the database, s/he can be directed to the complete record of the scientist in the local database which appears in the language of this database.

**Registrierung für Wissenschaftlerinnen: (German)**

Vorname:  
 Familienname:  
 Adresse:  
 Erfahrung:  
 Forschung:  
 Andere Interessen:  
 Email:  
 ....

**Data for the European DATAWOMSCI Service Provider: (Please enter data in English)**

Qualification:           [Look up qualification in Subject Index]  
 Degree                   [Lookup degrees in classification]  
 Organization:  
 Country:  
 Experience:             [Look up experience in Subject Index]  
 Language Skills:

Figure 10: A local registration form for data in English to be harvested by the MonoDSP

Figure 10 shows a simple overview of a possible registration form in one of the local databases in the German language. The registration form has the usual fields of the database, listed in the upper half of the screen in German. Further, the registering scientist is asked to fill in information in English. Help tools like multilingual thesauri

or subject indexes may be used to assist the non-English speaking user in filling in the fields in the English language. The use of subject indexes can also enhance the search and navigation capabilities of the MonoDSP.

The MonoDSP has no further differences when it comes to the harvesting of the data from the local databases. This remains the same as in the definition of the DATAWOMSCI Service Provider in Section 4.2 and as in the definition of the multilingual DATAWOMSCI Service Provider in Section 9.1. Figure 11 shows an English based Monolingual DSP for the five databases of the feasibility study. In femdat, FemConsult, and Kilden the users enter bilingual information. From all databases only the common fields in English are harvested.

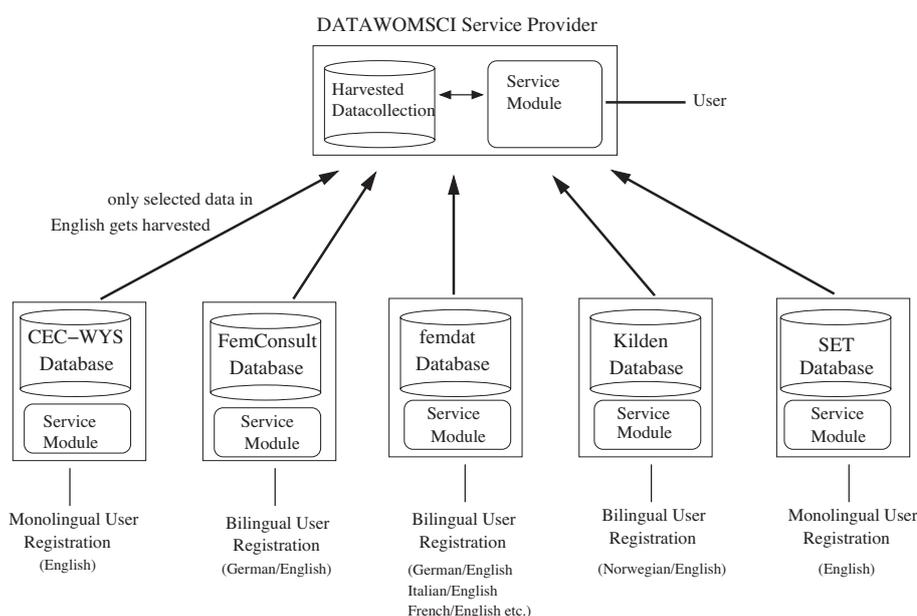


Figure 11: A local registration form for data in English to be harvested by the MonoDSP

One of the advantages of the MonoDSP model is that the complexity and costs of developing a multilingual service provider is avoided. The translation work which would be required of the maintainers of a multilingual service provider or of the users searching the DSP data collection is shifted to the scientists contributing their data. The registering women may then decide how they classify their work in the given classification models and subject indexes. The women also decide if they want their data to be available in a further European database.

There are disadvantages to the MonoDSP model. First of all, women scientists whose proficiency in English is not good enough to fill in the fields may be excluded from the MonoDSP. Assisting services like multilingual thesauri or subject indexes may be helpful with this problem. In addition, the fields that need to be filled out in English may still be labelled in the language of the database as well as the help texts describing the fields.

Next, the cost of implementing an addition to the local databases may be costly and complex. In addition to the problems that may arise when changing the data model, the conception of a new registration form or the binding of thesauri may cause more

work than the institutions are willing to do or pay for. The maintainers of the databases may also find it confusing or discouraging to ask registering scientists to fill in some of their information repetitively.

In order to respond to these problems, it is recommended to do a user survey or prototyping in order to find out if users agree with the bilingual model, have objections to it or further suggestions.

Once the decision is made for a MonoDSP it is also highly recommended that a MonoDSP-implementation package is prepared for the local databases and a support desk is created to support the database administrators in realizing the MonoDSP registration fields. Such a package would be useful for existing women scientists' databases as well as future women scientists' databases as they are developed in different institutions in Europe. This package should also include benefits of a MonoDSP for women scientists, and it should explain why the scientists would have to enter some of their information twice.

The following steps are recommended for the realization of the MonoDSP:

- determine if the current institutions maintaining women scientists' databases and their administrators accept the extension of their databases to include certain fields in English to be harvested by a MonoDSP.
- survey users about entering their information in a second language and repetitively either through a questionnaire or through prototyping.

If all stakeholders agree and support a MonoDSP, then:

- negotiate with the current institutions that maintain women scientists' databases on the content and labelling of the MonoDSP fields, as well as thesauri and subject indexes to support the registering users.
- find out the legal framework in order to formulate a "data protection" paragraph within each of the local databases for the use of the data in a further database, namely the MonoDSP.
- develop an open source support and implementation package on "implementing the data provider for the MonoDSP" which also includes information on the data model to be implemented, the services to be implemented with the registration form, the standards that are recommended, and a list of best practice criteria relevant to the extension. The CERIF 2000 Toolkit can be taken here as an example. It is recommended that such an open source package should be revised periodically and the revision should take feedback from developers and users into consideration.
- create a support desk which can be used by the institutions maintaining databases for questions, problems and feedback.

### 9.3 The distributed solution: the DATAWOMSCI Women Scientists Data Collections Information Provider (DIP)

In this model a site called the DATAWOMSCI Information Provider (DIP) offers its users an overview of existing women scientists data collections in Europe. Depending on the

services available the **DIP** can offer detailed information on the databases as well as support in the use of the databases.

The DATAWOMSCI Information Provider can have one or more of the following objectives:

- offer a catalogue of data collections on women scientists
- offer a classification of data collections on women scientists
- inform users of the databases' scope
- inform users on how the databases can be accessed
- support trouble shooting with the maintaining institution and deliver contacts
- evaluate and present the strengths and weaknesses of the databases
- improve the quality of databases by giving advice and examples on how quality can be increased
- become a space to develop a community of practice around women scientists database maintainers and users

An additional benefit of a DATAWOMSCI Information Provider is that it allows users to get an overview of the existing databases on women scientists around Europe. This means that the information is not only interesting in terms of finding out about the local databases but also in finding out about patterns in women scientists' databases across Europe.

Furthermore, a DATAWOMSCI Information Provider could also be useful for the database maintainers to compare their resources, to develop a community of practice, and to share experiences. The users of the DIP can make comparisons between sources and this could assist the users in making conscious decisions on which sources fit their needs best, without having to visit each database sequentially.

In the following Sections we shall analyze various aspects of a DATAWOMSCI Information Provider. We shall first analyze the possible objectives of a DIP and how these can drive the requirements for services of the DIP in Section . Next, in Section we shall analyze the global content that should be included in the DATAWOMSCI Information Provider in order to reach the objectives of the DATAWOMSCI Information Provider in a transparent manner. In this section we shall also shortly review some of the additional interactive services that can be offered by the DATAWOMSCI Information Provider. Then in Section we analyze the content on the data collections and institutions registered in the DATAWOMSCI Information Provider.

### *9.3.1 The objectives of the DATAWOMSCI Information Provider and the resulting Services*

The objectives of the DATAWOMSCI Information Provider can drive the decision on which services to include. There are two objectives which are not exclusive. These are the following:

- The DATAWOMSCI Information Provider will act as an independent source on women scientists' data collections.
- The DATAWOMSCI Information Provider will act as a source for self-organisation among the maintainers and users of women scientists' data collections.

A DATAWOMSCI Information Provider with the first goal would require that the institution maintaining it would periodically collect information and evaluate the services and quality of existing women scientists' databases. This could just be done either by simply organizing and classifying information on women scientists' databases or by going as far as delivering usability reports and including user reviews. The descriptive authority in representing the existing data collections would lie in the hands of the DIP maintainers. The content would be produced centrally by the institution maintaining the DIP.

A DATAWOMSCI Information Provider with the second goal would include services which would allow the maintainers of the data collections to self-define. The preliminary information on existing data collections could be extracted from the questionnaires sent out during the DATAWOMSCI feasibility study.<sup>5</sup> Services of the DIP would allow maintainers of the data collections to register their data collections on DIP and to describe and evaluate their own services. Users of the data collections as well as the maintainers could keep the DIP lively through their contributions and updates.

We recommend that the DATAWOMSCI Information Provider is a combination of these two goals as shown in Figure 12. For example, the document "Women Scientists in Europe - Databases and other Resources" which gives a structured view of existing data collections is the kind of information and analysis that belongs to a DATAWOMSCI Information Provider with the first goal. Usability reports on existing women scientists' databases also fit this goal. Nevertheless, the collection of such information is highly time consuming and requires access to networks in numerous countries. Therefore, introducing services which allow institutions that are maintaining data collections to register their resources is useful in distributing the task of information collection.

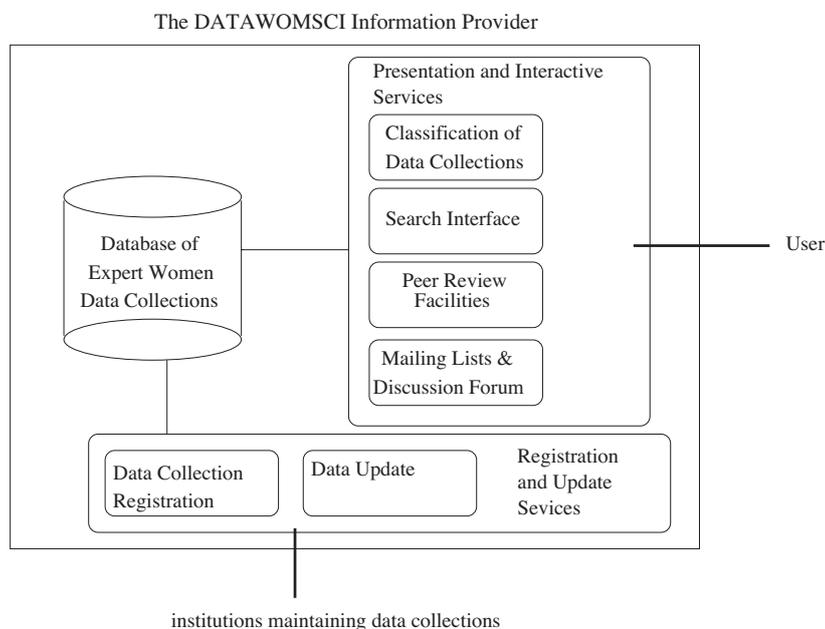


Figure 12: A preliminary model of the DATAWOMSCI Information Provider

Institutions that register their data collections in the DATAWOMSCI Information Provider can also decide how they want to classify their data collections and enter help

<sup>5</sup> Here we refer to the DATAWOMSCI questionnaire coordinated by the CEWS and not the technical feasibility questionnaire.

texts. The DIP maintainers could periodically check the DIP database for quality of data and run questionnaires on the data collections.

Further, the DATAWOMSCI Information Provider can publish further studies run by the institutions maintaining the data collections or refer to these publications. Last, peer reviews could be encouraged by offering services that allow evaluative interaction among the maintainers of the institutions. Similarly, tips and critique from the users of the data collections can be encouraged. If peer reviews are of interest, then services shall have to be developed which support these processes with the participation of the interest groups.

The interaction around the DATAWOMSCI Information Provider can also be enhanced by offering services for a community of practice. This could be through discussion forums or mailing lists set up for the different communities involved in the field of women scientists data collections. In all these models, it is important to have the support of the institutions maintaining collections either through data contributions or through dissemination in their own networks and in their communities of practice.

In the next two sections we make recommendations on the global content that should be available on the DATAWOMSCI Information Provider about its organisation and objectives, and the content on the databases and institutions registered in the DATAWOMSCI Information Provider.

### *9.3.2 Global content and services on the DATAWOMSCI Information Provider*

By global content we mean the content describing the complete DATAWOMSCI Information Provider. Transparency should be a guiding principle in describing the DATAWOMSCI Information Provider. The users should be able to inform themselves of the objectives of the provider. Most important of all, documentation and the organization of content should support the users of the DATAWOMSCI Information Provider in utilizing the services according to their needs.

Next, we list some of the topics which are highly recommended as global content. Later we make suggestions on services that could support interaction on the DATAWOMSCI Information Provider. We believe that it is through strong interaction that the DATAWOMSCI Information Provider can grow with its user community by responding to their interactions and hence their needs and requirements.

We suggest the following issues to be handled as part of the global content of the DATAWOMSCI Information Provider. As users formulate their needs this content will have to be revised and reorganized.

#### **background information:**

Provide the users of the DATAWOMSCI Information Provider with a general overview of what the database covers, how it links with other databases and other services of the information provider.

**criteria for inclusion:**

Explain which data collections are included in the DATAWOMSCI Information Provider and why. For example, in some cases institutions may be listed rather than data collections, as was the case in Chapter 3. Such decisions can be made transparent to the users of the DATAWOMSCI Information Provider.

**the process of data collection and data update:**

It is recommended to have a central statement as to how the original DATAWOMSCI Information Provider came to being. In addition, it is also recommended to document the process of updating the DIP datasets. If self-registration and update of the information on the registered data collections is possible, then the date of update also needs to be made transparent.

**geographical areas covered by the data collections:**

It is highly recommended to include an overview of the geographical areas covered by the various data collections. Here one should have numerous navigation possibilities, i.e. by country, by region, and also by the language of content.

**what is the DATAWOMSCI Information Provider useful for:**

Here the different purposes of the site can be included depending on the possible objectives of the DATAWOMSCI Information Provider.

**help with using the DATAWOMSCI Information Provider:**

Here various documents can be included on the services offered by the DIP and how to use them. If search and navigation facilities are part of the services the use of these can be described here.

**overview of the data collections:**

The analyses that were done in the Chapter "Women Scientists in Europe - Databases and other Resources" can be presented as an initial overview of data collections in Europe and can be extended as new institutions and data collections register in the DIP. It is recommended to do such an overview periodically to show some of the interesting aspects of the different data collections as well as to show best practice examples and gaps in women scientists' databases. In this sense, the DATAWOMSCI Information Provider could influence policy regarding women scientists data collections and institutions interested in networking women scientists in Europe.

**how to contact the maintainers of the DIP:**

Information on the institution maintaining DIP and contact information should be listed on the site.

**useful links and publications:**

Part of the DATAWOMSCI Information Provider could include the publications of the DATAWOMSCI Project. In addition, the best-practice criteria for women scientists' databases, which also attends to technical aspects, can be included. Links can be set to other resources of information.

**quality of data:**

In order to guarantee the quality of data on the DATAWOMSCI Information Provider the links to the data collections need to be checked for correctness. Measures like automatic link checks as well as date of last update need to be implemented to guarantee these quality checks.

In addition, to the topics described above it is recommended to have search facilities that allow the users to search through the registered data collections. The following search possibilities are highly recommended.

- search through fields: depending on the data model of the DATAWOMSCI Information Provider it should be possible to search through the different fields.
- search geographically: a classification of regions and countries should allow the users to find out about the different data collections or institutions in the regions or countries of their interest.
- search by discipline: if users are interested in searching for specific discipline then this should also be possible through a classification of the disciplines included in the different data collections.

Further, the following technologies can be considered for encouraging interaction on the DATAWOMSCI Information Provider:

- mailing lists: mailing lists can be used to build a community around the DATAWOMSCI Information Provider. These could be mailing lists for institutions maintaining data collections, mailing lists of administrators of women scientists' databases, mailing lists for users interested in being informed about relevant news on the DATAWOMSCI Information Provider through a newsletter.
- discussion forums: if users want to comment on issues around the DATAWOMSCI Information Provider or around "women scientists data collections" or further possibilities of networking, discussion forums can be handy. In addition, a list of relevant announcements related to the following or further topics can be implemented through discussion forums (i.e. open positions, call for papers or conferences, or events taking place around Europe or elsewhere). Also, in order to allow peer reviews, the discussion forums can be linked to descriptions of data collections in an organized manner.

Last but not least, extra and short documentation like FAQs and Tutorials should be included in the site to assist the users in using the DATAWOMSCI Information Provider, to guide them through the services, and to make the site transparent.

### 9.3.3 Information on the registered data collections

In the following we list the attributes which can be used to represent the different data collections in the DATAWOMSCI Information Provider. Where it is possibly unclear we describe the purpose of the given attribute. The list is not ranked in order of importance but rather is a list of recommendations. We consciously use the word "data collection" here instead of "database" in order to emphasize that data collections can also be data resources which are not online, based on a database software, or even digital.

**history of the data collection:**

Provides a short history of the data collection and the maintaining organization.

**objectives of the data collection:**

Lists the main objectives of the data collection and the audiences of the data collection.

**the main services of the data collection:**

Contains a short list of the services offered by the institution maintaining the data collection and the online services (if they exist).

**the languages of the data collection:**

The language of the interface and the datasets are provided. If cross-lingual search is possible this is also mentioned under this heading.

**who can register:**

Describes the preconditions for registering in the data collection. Who are the targeted scientists? How is "scientific expertise" defined?

**which disciplines:**

Which disciplines are included in the data collection?

**where do I find information on how to register:**

If the data collection is online and offers online registration services, then a link to these services can be provided here. In addition, any comments or tips on registration can be included in this field. For offline data collections contact information and a description of the registration procedure can be provided.

**quality of data:**

How often is the data collection updated? What is happening to outdated data? Are the data validated for consistency and correctness of information? Are letters of recommendation or proof of expertise in the discipline of the scientist required?

**the size of the data collection:**

How many datasets are included in the data collection? If the number of datasets are not calculated dynamically, then the date at which the dataset was given in can be included. Further, information on how many women register per month and year can also be included.

**data update feedback:**

Is it possible to be informed by the data collection holder of new entries to the data collection? Where can the user interested in such a notice register for such a service?

**are national or local subject indexes or classifications used:**

This is recommended for data providing users to know what classification possibilities exist. This information is useful for an overview of the different subject indexes or classifications used around Europe, as well.

**geographical coverage:**

Which geographical regions are covered by the given data collection? Which language space?

**is the data collection accessible online:**

If the data collection is online, then the url can be provided. Online and offline data collections may have to be represented differently in the DATAWOMSCI Information Provider, since some of the information for online data collections are irrelevant for offline data collections. If the data collection is offline then information needs to be provided as to how scientists who want to be included in the data collection can contact the institution, and how users searching women scientists can contact the institution.

**data privacy:**

Inform the users on which legal data privacy framework, if any at all, is used for the registered data. It is also important to inform all users of which data shall be hidden for privacy reasons.

**privacy and archiving:**

The data providing scientists may want to know how their information is archived. Is it only on computers or is it also on paper. If data privacy concerns result in the hiding of some data, the users may want to find out who has access to the complete record.

**data analyses:**

Are there filters or services of the data collection which allow for overall analyses of the data (i.e. the number of women scientists registered with a specific discipline).

Users searching scientists in the data collection can analyze the data through search (search for women with a given discipline). Further, services may be provided that dynamically output such information.(i.e. a service that provides an overview of all the data for all the different disciplines: in this data collection n women scientists with discipline x and m women scientists with discipline y can be found etc.) Are there any reports produced by the maintainers on their data collections, which include information like who is registered, which disciplines are most represented, which locations, and which interests.

**the fields of the data collection:**

The list of the data collected from the data providing scientists.

**who maintains the data collection:**

Which institution or persons maintain the data collection? A link to the organization may also be included, if it exists.

**feedback and questions:**

Is it possible to contact persons or institutions for feedback and questions? If this is the case, contact information should be included.

**help and documentation:**

Where can the users wanting to search or register in the data collection find help documents?

**further remarks:**

Particularly institutions registering their data collections may want to add particular information on their data collection which is not covered by the above headings. In such a case, a free text field called "further remarks" is very useful.

#### 9.4 Developing a DATAWOMSCI Platform: A Service Provider, an Information Provider, or both

In this section we conclude our findings on the alternatives to a DATAWOMSCI Provider and make recommendations on the development process and maintenance of the future DATAWOMSCI Provider.

First of all we would like to mention that the DATAWOMSCI Service Provider and the DATAWOMSCI Information Provider can be developed together as shown in Figure 13. We have defined the DSP and the DIP as two components which can be included in a DATAWOMSCI Platform, developed sequentially or at the same time. The decision on the development of one or more of these modules depends on the finances and resources available to the implementation of the DATAWOMSCI Project.

### DATAWOMSCI Platform

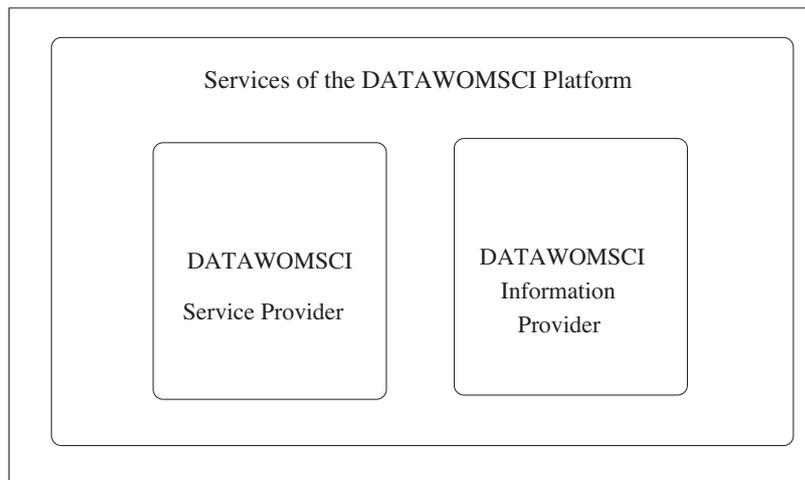


Figure 13: A DATAWOMSCI Platform which contains both a DATAWOMSCI Service and Information Provider

Regardless of the components to be developed, we highly recommend the use of participatory development models in developing a DATAWOMSCI Platform. Participatory development models encourage the participation of all the stakeholders in the development process of a platform. In addition, participatory development models are not product but process oriented. This means that there is always the possibility of extending or updating the platform according to the needs of the users of the platform.

Keeping a platform lively usually requires the organization of events online or offline. This could be the initiation of a platform, a topic based online discussion in a given time interval, or the organization of a real-life meeting. In addition, the maintainers of the DATAWOMSCI Provider need to be aware of the changing needs of their users and adapt the services of the platform to these needs.

Once the DATAWOMSCI Platform is up and running its existence only makes sense if it can be maintained and updated regularly. As soon as the information is out of date, the DATAWOMSCI initiative can be endangered. We recommended in the definition of the DATAWOMSCI Information Provider two possibilities for updates: on the one hand self-organized updates of the institutions maintaining data collection, in that they register their data collections on the platform and disseminate news and important information; next, a central management and support team which takes care of organizing and updating the site and validating the new content, evaluating the quality of incoming data and reorganizing the services of the platform. Keeping communities of practice around the DATAWOMSCI Platform active also requires the organization of events and services which support their interests and encourage them to actively take part in the DATAWOMSCI Platform.

It is therefore central that in planning the DATAWOMSCI Platform sufficient resources are organized for an extended requirements analysis process with user participation. In addition, resources should be secured for future real-life activities and for a continuous support desk and maintenance team that can respond to the requirements elicited throughout the development process and during the use of the DATAWOMSCI Platform.

## 10 Metadata for the future DATAWOMSCI Service Providers and Data Providers

One of the objectives of this Section is to analyze the use of metadata in order to make the data exchange between distributed heterogeneous databases possible. The Open Archives Initiative Protocol for Metadata Harvesting requires the use of unqualified Dublin Core at a minimum and further metadata can also be used like qualified Dublin Core or others. We recommend the use of unqualified Dublin Core as defined within the OAI-PMH for a DATAWOMSCI Service Provider as a minimum.

Nevertheless, during our technical feasibility study we came across the CERIF 2000 Toolkit which offers a data exchange model. Since, the goals of the CERIF 2000 Toolkit overlap strongly with the goals of the DATAWOMSCI Project, and the CERIF data model is recommended as a data model for research related databases, we concentrate our analysis here on the CERIF data exchange model. This analysis is useful for understanding data exchange possibilities between existing women scientists' databases as well as for future women scientists' databases.

In order to do this, we first introduce the CERIF concepts. Next, we give an overview of the data models offered by the CERIF 2000 Toolkit, and last we analyze the CERIF data exchange metadata model for adaptability to future women scientists' databases and to a potential DATAWOMSCI Service Provider (DSP).

### 10.1 CERIF

CERIF resulted from the need to collect research information from heterogeneous distributed research information systems. It is argued that often more than one system contains the research information that a user is looking for. This brings about a demand for the integration of research resources.

Numerous solutions are developed and maintained by the CERIF Task group in order to reach the goal of resource integration. Until now these are:

1. simple portal pointing to CRISs (e.g. DRIS)
2. central catalogue and contact information for individual CRISs e.g. ERGO
3. central catalogue and automated retrieval from CRISs (ERGO2++ Proposal)
4. central catalogue and advanced knowledge-assisted retrieval from CRISs (ERGO3 proposal)<sup>[31]</sup>

In its goals and its suggested solutions the CERIF initiative has commonalities with the DATAWOMSCI Project. As part of the technical feasibility study we study three alternative models, as introduced during the DATAWOMSCI Project. The "distributed model" which we introduced at the earlier stages of the project corresponds to the CERIF model number 2, and the "federated model" corresponds to the model number 3 above. The models number 1 and 4 are not considered, the former for its simplicity, and the latter for its complexity. The central standardized solution is not of interest to

CERIF, the main goal of the different CERIF models is to integrate "distributed" heterogeneous databases and not to provide "central" solutions.

In their paper, Lopatenko and Anderson (2002) study three CERIF approaches. We introduce these approaches in Sections 10.1.1 through 10.1.3 in order to sketch the possibilities considered by the CERIF task force. Adopting the CERIF data model would make the introduction of such approaches possible in the future. Next, in Section 10.2 we analyze in-depth the "CERIF exchange data model" from the perspective of the DATAWOMSCI Service Provider to show the feasibility of using the data model in the DATAWOMSCI Service Provider or in future women scientists' databases.

#### *10.1.1 Distributed Database Approach*

In the distributed database approach each database exports a set of views conforming to the requirements of CERIF (signature of view and semantics of attributes). In order to guarantee this export JDBC access to database views for select operations must be provided. For each network one central server repository exists which registers all databases and provides a search over all databases to the end user.

When views are exported for a database, the database must be registered at the central server of network, with description of data source, quality, sector of research and access information. After the database is registered it will be used by the central search facility to answer queries of users. The data is registered at the central server through SOAP web service interface by sending the RDF description of data source.<sup>[31]</sup>

A number of preparations need to be made in order to realize this model. The views need to be defined and provided by each database. In some cases certain standardized vocabularies or look-up values may also have to be imported. Each of the distributed databases may be using their own sets of vocabularies. In such a case, these vocabularies have to be mapped onto the standardized vocabularies. These mappings also have to be included in the view definitions.

These preparations also make up the disadvantages of the distributed database approach. This approach has a high demand for manual configuration (view definitions, vocabulary mapping, look-up tables). Also, security problems may arise as a result of direct database access.

The distributed database approach harvests information from the distributed databases very much like the DSP model we suggest in Section 4.2. Yet, with the introduction of views, the model is much more complex and hence more flexible in relation to the DSP. In addition, we have planned the DSP such that the data providers do not have to do any extra work such as mappings and view definitions. Hence, the distributed database approach is bundled with extra workload which is not appropriate for the DATAWOMSCI Project.

### 10.1.2 Semantic Web Approach

The semantic web approach aims to solve problems of distributed information retrieval when semantic heterogeneity is high and sophisticated information retrieval operations are demanded. The semantic web approach is based on CERIF as the core ontology to describe the meaning of research data. CERIF is the base vocabulary which provides a common set of terms and which must be understood by any system conforming to CERIF-SW approach. <sup>[31]</sup>

An advantage of the use of the semantic web is that the same information can be represented in information systems in different ways, described from different point of views due to diversity of data sources, policy restrictions, or even the position of the author. The views of the same information can be very different for different categories of information consumers.

The disadvantage of this approach lies in its complexity and costs. In return, the efficiency of this approach is currently very low. Search is very inefficient. Further, the approach requires developers with knowledge of the semantic web, ontologies, and special tools.

For the DATAWOMSCI Project there is an additional reason why current implementation of the semantic web is not realistic: the quality of existing ontologies. Few ontologies include sufficient information on interdisciplinary research areas, where many women are active. For example, it is rare to find women's studies, gender studies, queer studies, feminist studies etc. in many ontologies.

### 10.1.3 Web Services Approach

CRISs with a schema different from CERIF can easily participate in the CERIF network by publishing (through a wrapper) their data in CERIF XML. CERIF-WS has been built in such a way that it is very easy for developers to create or change the XML encoding of their data in the wrapper. Thus the technique allows CRISs with non-CERIF encoding to be accessed as if they were CERIF-encoded at the expense of building and maintaining a wrapper.<sup>[31]</sup>

The web services approach is not as efficient as the distributed approach. Web services are still not very mature and the current CERIF implementation only allows very simple queries.<sup>[31]</sup>

The web services with wrappers has a similar approach to the DSP model that we introduced in 4.2. In addition, it contains complex web services which are beyond the needs of the DATAWOMSCI Project.

### 10.1.4 Prospective of the CERIF approaches in the DATAWOMSCI Platform

All of the approaches listed above are far more complex than the DATAWOMSCI Service Provider we recommended in Section 4.2. Nevertheless, they are approaches with

long term technological perspectives within the European Union. The implementation of any of the models described above is therefore exciting and yet costly, even though the CERIF 2000 Toolkit is freely available. This is the case, not only but also because a detailed application of CERIF in DATAWOMSCI will require the cooperation of the different administrators of the existing databases.

We now go onto analyzing the CERIF data model, which is independent of the implementation of any of the three approaches accounted for above.

## 10.2 The CERIF Data Model

Large numbers of documentation exist on the CERIF data model. Here we shall only give a short overview and analyze parts of the model that are relevant to a future DATAWOMSCI Service Provider.

In the CERIF documentation, the goals of the CERIF 2000 data model are listed as follows:

- a full Current Research and Information Systems (CRIS) data model with flexibility to allow the majority of existing CRIS to accommodate their own database structures
- a base framework for data exchange

In order to reach these objectives three data models are defined in the CERIF 2000 Toolkit:

- a CRIS data model, covering the majority of the CRIS database structures
- a set of data models that provide examples for data exchange
- a metadata data model to provide a uniform summary level view over heterogeneous information sources

For those databases that will be built in the future, we recommend the adoption of one of the first two data models in developing the new databases. Such an adoption will however depend on the local contexts and is beyond the scope of this study.

We shall analyze the exchange data model to investigate if enough overlap between this model and the suggested DSP model exist. The analysis will take place rather at the content level, meaning, we shall see if the data collected by the DSP could fulfil the minimum content necessary for the exchange data model to be applicable. The CERIF 2000 Toolkit additionally offers a list of metadata, which can be used to realize the exchange data model. We shall not analyze in-depth a specific list of metadata or a specific data model for DSP or its mapping to the CERIF exchange data model, since this goes beyond the boundaries of this study. Instead, we shall point out problem points while adopting a data model and make suggestions on how to deal with these problems.

### 10.2.1 Overview of the CERIF Data Model

The exchange data model is a subset of the CERIF full CRIS data model. The hierarchy of the data models offered by the CERIF 2000 Toolkit are shown in Figure 14. The

metadata data model offers a summary view of the exchange data model, and can be used to extract the necessary metadata for the DSP, once the data model is decided. This process is left for the implementation phase of the DATAWOMSCI Project. We shall shortly analyze the basic building blocks of the full CRIS data model and then analyze the use of the exchange data model for DSP.

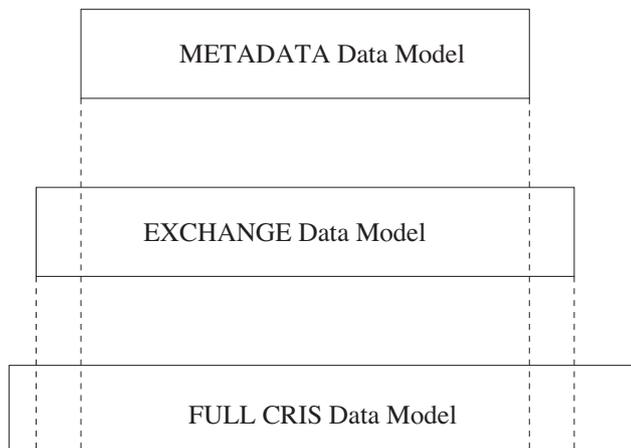


Figure 14: The different data models offered by the CERIF 2000 Toolkit

#### The full CRIS Model:

The full CRIS data model is developed around three major entities: projects, persons, and organisational units. The relationships between them are many-to-many and controlled by role and time interval. This means that a project may have relationships with many persons and organisational units within a given time interval. Similarly, a person may have relationships with many projects and many organisational units. And organisational unit may have relationships with many persons or projects.

Note that the CERIF data model covers more than the data that existing women scientists' databases offer. In none of the databases which we studied are there many-to-many relationships –i.e. which persons are part of a specific organisation or project–represented. The CERIF model is based on Current Research Information Systems which are usually run centrally for a given organisation. In such a case, the relationships between the projects of the institute, the persons working there and the organizational units can be centrally and completely identified. This is not the case for any of the data providing women scientists' databases nor for the DSP.

In addition, the CERIF data model has an equivalent emphasis on projects and organisational units as it does on persons. The databases we have studied in the feasibility study so far are mainly only about persons. Few of the databases include explicit information on projects the scientists are involved in. Organisational units in which the women scientists are currently employed or were employed in the past are identified. But further details, unless entered in one of the free text fields by the data contributing scientist, are often not available.

Although these differences exist, there are numerous reasons why the analysis of CERIF for utilization in a future DSP is important:

- it is possible to only utilize a subset of the CERIF data models.
- the CERIF 2000 Toolkit is a collection of experiences over the years (the first CERIF initiative was already launched in the 90s. CERIF is an attempt to develop a model for research information systems. Scientists, and their research and scientific activities are a subsystem of the research information systems.) By studying CERIF we can identify how women scientists' databases can be made visible, explicit, and findable within the research landscape.
- there are numerous databases already built with the CERIF data model. Designing the DSP to be compatible with the CERIF standards would make it possible to harvest the women scientists listed in these databases.
- by analyzing the use of CERIF 2000 in the context of women scientists' databases, suggestions and recommendations may be formulated that enrich the CERIF 2000 Toolkit.

We now go on to study the relevant parts of the CERIF data model for the DSP.

#### **The five levels of abstraction:**

In order to explain the CERIF data model with all its complexities, the authors of the CERIF 2000 tool kit have introduced five levels of abstraction:

1. level 1 only covers the most important or core entities: person, project, organisational unit.
2. level 2 includes also secondary entities such as results, classification, contact, event.
3. level 3 includes for all entities covered by level 2, the translations.
4. level 4 includes "look-up entities" such as for "academic title", for which a look-up table usually contains a defined set of values.
5. level 5 deals with the many-to-many relationships, such as between person and "project or person" and "organizational unit".

We shall first analyze the top level entities and then take a look at an "expertise exchange data model" and study the further levels of the abstraction.

#### **level 1: Top level entities:**

The top level entities are made of person, organisational unit and projects as shown in Figure 15. All other entities, within the scope of CERIF 2000, are considered to be related to these top level entities. All possible relations (meaning many-to-many: many persons can be involved in a project, many projects can be part of an organisational unit, a person may be part of many organisational units etc.) between the three entities are represented with a modelling language (similar to UML) used in the CERIF 2000 Toolkit.

Already the top level entities of the CERIF data model contain vast amounts of information. Only a subset of the entities and attributes defined here are meaningful for the DSP. After all, in all the databases that we studied during the technical feasibility study only Kilden database has the same kind of emphasis on organisational units and projects. None of the databases contain the many-to-many relationships that exist between the three entities of the CERIF data model.

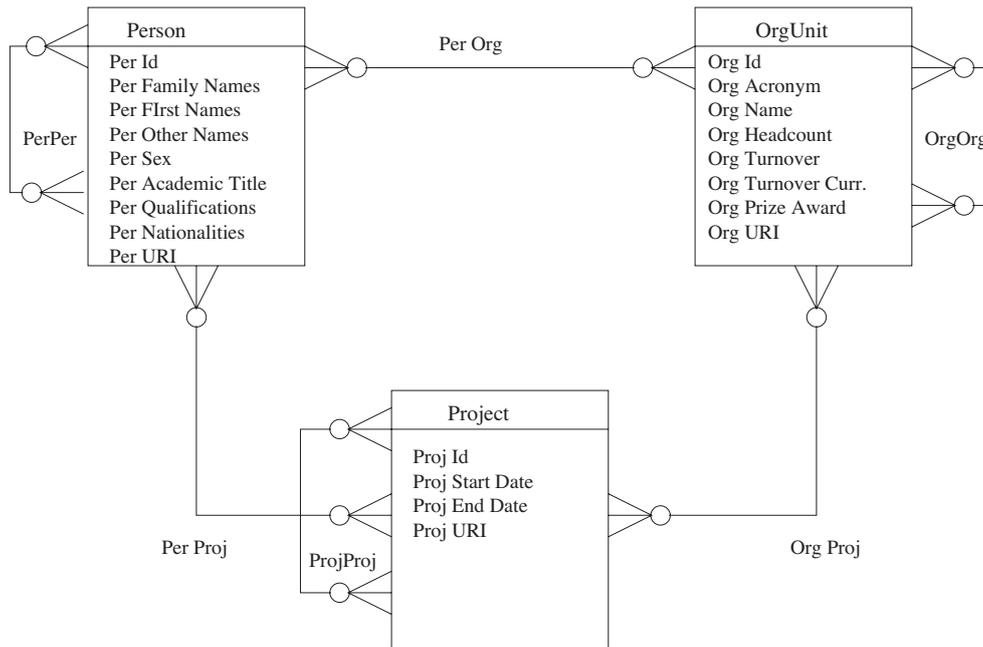


Figure 15: The top level entities in CERIF

Still, we shortly analyze the “person” and “organisational unit” entity with respect to DSP. The person entity contains almost all the fields we are interested in and some more. The relevant fields in the entity “person” are:

- **per Family Names, First Names, Other Names:** in Section 6.1.1 we recommended the inclusion of the name field in a DSP. We believe that it is essential to break down the Name field in order to distinguish between Family Name and First Name. This is for reasons of compatibility with the CERIF data model as well as for necessary consistency in databases. Naming these fields in the plural is also of value for those persons who may have multiple first and family names. The “Other Names” field is recommended but is not essential.
- **per Sex:** the databases we selected for our feasibility study all either contain only women or allow the distinction between males and females. It is recommended to have a sex field for two reasons. The first is for data exchange so that external databases that harvest the DSP recognize that the scientists listed are all women. Second, it is possible that institutions decide to allow a specific group of men to be part of their databases (this was the case for example in a number of the “gender studies” databases that we studied in the DATAWOMSCI Project). In the case of such an institutional decision to include scientists of both sexes, the DSP can hold onto its goal of promoting scientist “women” by sorting the female scientists. This decision will have to be taken with the consent of the institution maintaining the database. Such scenarios make the “sex” field relevant for the DATAWOMSCI data providers as well as the DSP.
- **per Academic Title:** this degree corresponds to the Qualification/Discipline field that we analysed in Section 6.1.1.
- **per Nationalities:** in Section 6.1.1 we objected to the use of nationality for reasons of possible discrimination. Therefore, we do not recommend the use of a nationality attribute.

Of interest to us is also the Organisational Unit. In Section 6.1.1 we recommend the use of an "institution/organisation" field. The CERIF 2000 data model recommends the inclusion of numerous attributes like headcount, turnover, prizes and awards of the institution. For future databases these attributes may be of interest. But, for the databases we have analyzed within the feasibility study, the attribute "Org Name" fulfils all the required information.

There is an important distinction between the relationships that can be modelled in the CERIF data model and the DATAWOMSCI Service Provider or the DATAWOMSCI data providers. The "Per Id" or "Org Id" are both keys in the CERIF 2000 data model. It assumes that each person or organisation within the database appears once and has a unique id number attached to it. Please see discussion on "redundancy of data" in Section 6.1.3 for some potential problems here. Redundancy in the databases that we studied may occur, not because of anomalies in the design of the database, but because the data contained in the database are not contributed to a central system with redundancy check but to distributed systems. Therefore, the relationship between a person and an organisational unit may only be one-to-many: a scientist may be part of many organizations but not the other way around. It will not be possible to represent who all the women are that belong to a specific organisation. This means that even if a number of women scientists from the same institution register within any of the women scientists' databases, this information will be lost.<sup>6</sup>

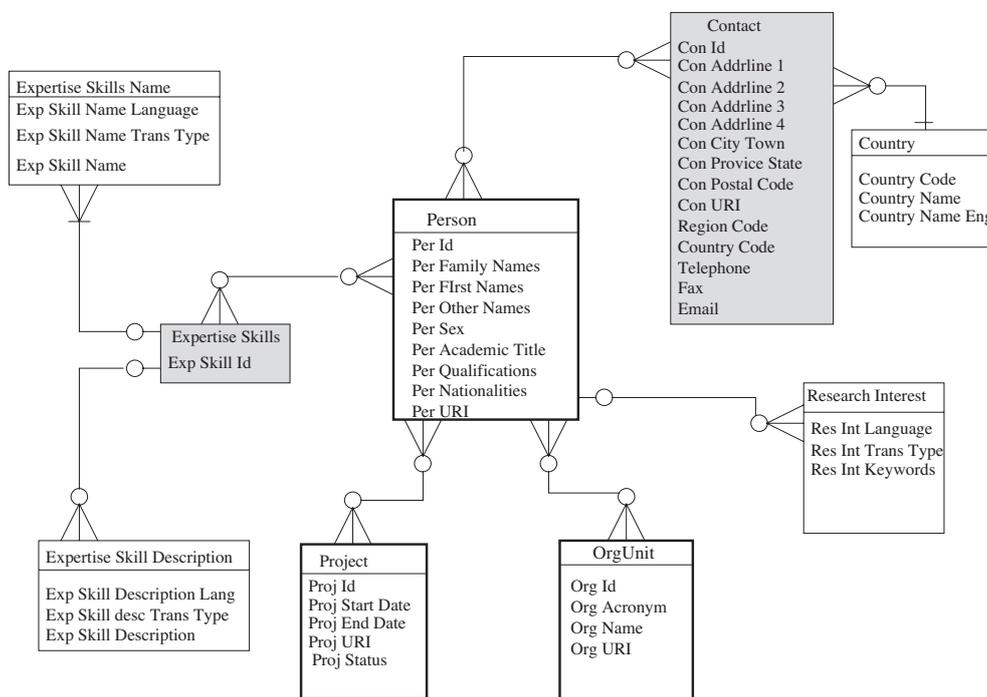


Figure 16: The expert exchange model in CERIF

6 For example, two scientists from the University of Bremen may enter their data. They may both describe the University of Bremen with the same vocabulary. But it may also be the case, that one enters "University of Bremen" while the other enters "Bremen University". Identifying each entry with a unique key implies these are two different organisations. Yet, unless somebody checks the data for consistency, there is no way for the system to recognize that they mean the same organization. Redundancy of data is unavoidable unless manual human work, possibly supported by a list of all organisations, is available.

Before we move to the analysis of the abstraction levels 2 through 5 we introduce an "expert exchange data model" based on the CERIF exchange data model. The analysis of the abstraction levels shall then follow with respect to this specific "expert exchange data model".

#### *10.2.2 The Analysis of the Exchange Data Model for the DATAWOMSCI Service Provider*

The idea behind the exchange data models is to negotiate a transfer of data between a data provider and a requesting end user, which in our case is the DATAWOMSCI Service Provider. Figure 16 gives an overview of an expert exchange model:

The analysis of level 1 we concluded in Section 10.2.1. We now analyze the abstraction levels 2 through 5 of the CERIF data model in the example "expert exchange data model". Next, we present a version of the expert exchange data model that could be applicable to the DATAWOMSCI Service Provider.

##### **level 2:**

Level 2 of the model includes secondary entities. These are the main "associative" entities with the top-level entities. In the expert exchange model these are "expertise skills" and "contact". These are slightly highlighted in Figure 16. Both are of interest to the DSP. Contact is limited only to email within our analysis of DSP.

##### **level 3:**

Level 3 is where the translated entities are introduced. If multilinguality is of interest to the DSP, this model can be used to include possible translations of harvested data. An example of the translated entities is given in the expert exchange model with the entities "expertise skill name" and "expertise skill description".

##### **level 4:**

Level 4 is where Look-up entities are introduced. Look-up entities usually contain a defined set of values. The value of an attribute of an entity is checked against these values to make sure it is a valid entry –for example codes for countries in the expert exchange data model. CERIF makes recommendations for what lists of values or classification codes should be used for some entries or attributes.

##### **level 5:**

Level 5 deals with the many-to-many relationships between entities. We mentioned the problems with many-to-many relationships within the existing databases and the DSP in Section 10.2.1. Therefore, we do not introduce many-to-many relations here.

If we only display the relevant data, the expert exchange data model for DSP would look as in Figure 17.

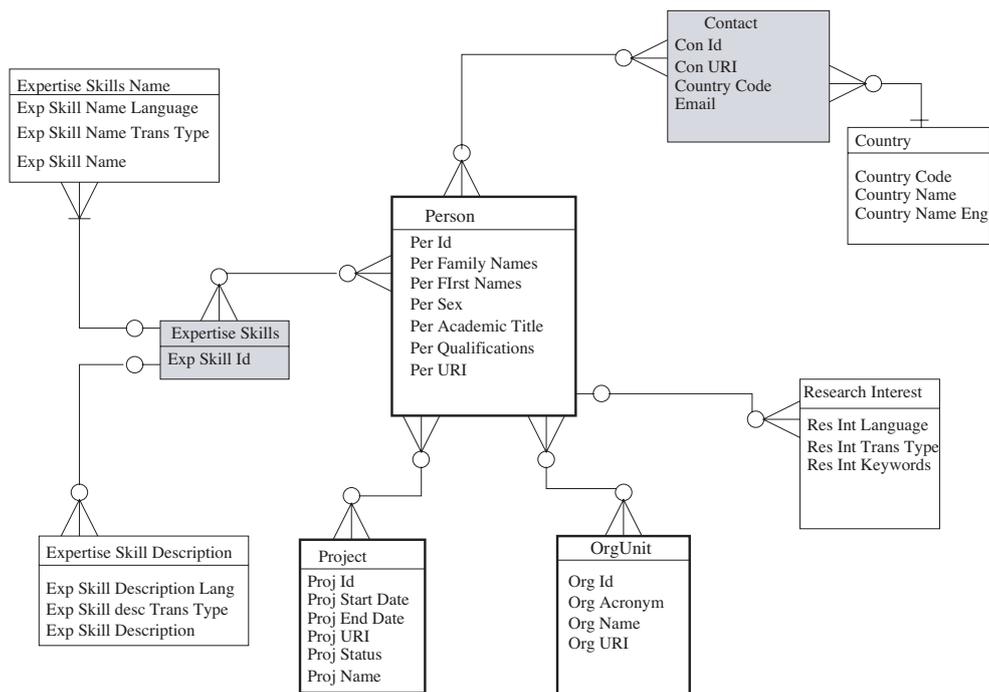


Figure 17: The expert exchange model for the future DSP

In addition to the expert exchange data model it is possible to make use of the look-up tables, or in other words classification systems used in CERIF within the future DSP. Therefore, in the next section we analyze the use of look-up tables. Last, we make suggestions on how to adopt the CERIF recommendations within the DATAWOMSCI Service Provider.

#### CERIF look-up tables and the DATAWOMSCI Service Provider:

The CERIF Toolkit recommends the use of standardized look-up tables, or in other words classification systems, in order to support data exchange as well as multilinguality. Standard look-up tables can be integrated into the DSP system. There are advantages and disadvantages to this. We shall shortly count some of the main issues related to DSP and look-up tables.

The use of controlled vocabulary or classification systems, as in look-up tables has the following advantages:

- helps with browsing
- enables the broadening and narrowing of searches
- gives context to search terms being used
- simplifies multilingual access to collections of information
- and allows the partitioning and manipulation of large databases.<sup>[20]</sup>

Automatic indexing techniques are still no substitute for the conventional use of classification systems or thesauri. Appropriate subject indexing is crucial in a multilingual environment.<sup>[20]</sup>

In the CERIF documentation different subject classifications are recommended. We identify those classifications relevant for the DSP:

- ortelius thesaurus: for research subjects
- ISO 639 standard: for languages
- ISO 3166 standard: for countries
- ISCED: for qualification of persons
- UNIMARC: type of publication

Ortelius and ISCED are classification systems that are highly relevant for scientists databases. Unfortunately, both classification systems are weak when it comes to contemporary or interdisciplinary subject areas. In ISCED for example there is no mention of gender studies or women's studies, cultural studies, or any other newer interdisciplinary fields. In Ortelius, Women's Studies and Gender Studies are listed but no mention is made of Feminist Theory. In both, few expressions of the newer and important digital media related disciplines are listed.

CERIF recommends the use of the Ortelius thesaurus of research subjects and offers support for enriching the Ortelius thesaurus. In Chapter 6 of the CERIF document a section is dedicated to "Enriching Thesauri for research and development". The instructions and contacts can be used by the maintainers of the DSP to make the developers of CERIF aware of the absence of gender and feminism related studies or other interdisciplinary and contemporary topics in which many women have expertise.

### *10.2.3 Adopting the CERIF Exchange Data Model*

The CERIF 2000 Toolkit offers numerous advantages. It is a flexible data model that can be adapted to the different needs of women scientists' databases as shown in the previous sections. It is easy to implement since a free-Toolkit and a support desk is available which can be used in the development process. Most important of all, CERIF promotes the use of open standards and therewith supports Europe-wide data exchange.

Even though a toolkit exists, we have shown in our analyses that there are problems of integration which can be complex and costly (i.e. mapping of subject indexes, use of semantic or multilingual technologies). The adoption of the CERIF exchange data model requires an understanding of the toolkit and its implementation by all stakeholders, which includes all the institutions maintaining databases and those interested in data exchange.

If the adoption of the CERIF exchange data model is of interest then a development process needs to be developed which allows the participation of the different stakeholders. Here we note two possibilities.

#### **A meeting with all the administrators and the maintaining institutions:**

All of the administrators we interviewed during the feasibility questionnaire were interested in our findings and were curious about the technologies and methods the administrators of the other databases were using. A meeting together with all these administrators and the maintaining institutions to decide on a model would foster a Europe-wide discourse on women scientists' databases and scientists' database tech-

nologies. Here the administrators, together with the institutions could discuss the viability of the different models and agree upon work like subject listing word-mappings (i.e. mapping the disciplines subject listing of femdat to the subject listing suggested by CERIF), multilinguality, existing and additional fields and attributes etc. They could also discuss the extension of these scientists databases to include publications and projects, as suggested in CERIF.

#### **A requirements analysis model with interviews:**

It may be difficult to bring all the administrators and institutional persons together. In such a case, project coordinators could travel to the different institutions, introduce the CERIF data exchange model, and interview the maintainers and administrators for their interests and perspectives. The interview results could be made available online for further discussion, leading to a concluding model and assignment of activities.

Without meetings with the administrators, the maintainers of the institutions and without the consent or support of these stakeholders, the implementation of a complex model like CERIF is not possible. Nevertheless, our analysis of CERIF here is still important for a deeper understanding of exchange data models for future DATA-WOMSCI Service Providers as well as for future women scientists' databases. We recommend the use of the CERIF exchange data model in future women scientists' databases.

## **11 Concepts for Search Interfaces on a Meta-database**

There are various reasons why a user starts a search on the web. In classical information retrieval **search** is defined as "driven by information need." Information need is defined as the "perceived need for information that leads to someone using an information retrieval system in the first place."<sup>[7]</sup> Yet, the intentions behind a search may be other than informational. In his article Broder (2002) defines a taxonomy of web search:

- **navigational:** the immediate intent is to reach a particular site
- **informational:** the intent is to acquire some information assumed to be present on one or more web pages
- **transactional:** the intent is to perform some web-mediated activity

The goal of a navigational search on women scientists database would be to reach a particular person. The searching person would most likely search the scientist by name or if the name of the person is not available through a refined search. On a DATA-WOMSCI Information Provider a navigational search example is a search for a specific institution or a specific database - e.g. what was the link to the Norwegian women scientists database.

An example of an informational search on a women scientists database would be "how many scientists have experience in computing" or "what further qualifications does scientist x have". On the DATAWOMSCI Information Provider a query on regional data collections in Eastern Europe can be given as an example.

The intention is what distinguishes the transactional search from an informational search. For example, on FemConsult, users search women with a specific expertise, if the search delivers positive results then the searching user may decide to contact CEWS, the maintaining institution, to ask for the contact details of the scientists. The use of the DATAWOMSCI Information Provider to find online databases to search in is also a transactional search.

The taxonomy of web search tries to distinguish the intentions of the user, which may have an effect on the structure of the services offered by a provider. The intention of the users while searching is difficult to determine without asking the user personally. Analyzing the log files for search may point to some hints, but still does not contain information on intentions.

In this Section we recommend user-friendly search and navigation services for a meta-database. Nevertheless, the usefulness of these services depends on the intentions of the users. Therefore, it is recommended to occasionally run surveys with the users of the women scientists' databases and the users of a future DATAWOMSCI Provider in order to find out the intentions and requirements of the users, and the resulting services that may attend to these requirements.

In Section 11.1 we give an overview of user-friendly search services. In the following Section we make recommendations related to the display of search results.

## 11.1 Search and Navigation

Search is the most important functionality of a DATAWOMSCI Provider since it is the main form of navigating through the data collected on the site. In that sense, search allows the user to prioritize his/her interests in the data collected in the site and in a sense customize his/her navigation according to these priorities.

In the following we suggest services which can support users in developing their navigation through the data on the site through search. In the text **search terms** are defined as character strings separated by space.

### 11.1.1 Suggesting search terms (query expansion)

Experienced searchers use more search terms than less experienced searchers.<sup>[4]</sup> The users can be supported by suggesting further search terms. This can be done through the use of thesauri or clusters. In the clustering method group of the results that share the same words or phrases are used to deliver topic filters. These filters can then be used to navigate to these results. See also the topic "use of filters" below.

The suggestion of search terms can be useful in multilingual as well as monolingual search. The Mulinex Project (1997) offers a wonderful application of multilingual query expansion and translation. The search starts with a simple interface, which requires the user to enter the language of query(source) and the target languages. A search mask in a DATAWOMSCI Service Provider would additionally ask the user which databases s/he is interested in searching as shown in Figure 18.

Figure 18: An example multilingual search mask based on the Mulinex search mask.

Once the user decides on the target languages a query assistant can be used to identify the translations of the term, as well as expansions of the original search term. Figure 19 shows an example query assistant based on the Mulinex query assistance. The first column in Figure 19 suggests alternative or contextualizing search terms, whereas the second and third columns lists possible translations (we did not include all the possible alternatives and translations for purposes of simplicity). In addition, the meaning of the translated terms are recapitulated in the source language so that the user can select translated terms according to the context.

English query terms	French translations	German translations
search term: <input type="text" value="fair"/> <input type="checkbox"/> blond <input type="checkbox"/> fair-haired <input type="checkbox"/> moral <input type="checkbox"/> righteous <input type="checkbox"/> bazaar <input type="checkbox"/> just <input type="checkbox"/> sales activity <input type="checkbox"/> cute <input type="checkbox"/> equitable <input type="checkbox"/> regular <input type="checkbox"/> light <input type="checkbox"/> mess <input type="checkbox"/> trade fair	<input type="checkbox"/> <b>blond</b> blond, golden, fair haired <input type="checkbox"/> <b>moral</b> fair, righteous, moral, just <input type="checkbox"/> <b>marche</b> bazaar, marketplace <input type="checkbox"/> <b>kermesse</b> bazaar, fair	<input type="checkbox"/> <b>Jahrmarkt</b> fair <input type="checkbox"/> <b>Messe</b> fair, mess <input type="checkbox"/> <b>blond</b> fair, light <input type="checkbox"/> <b>gerecht</b> equitable, fair, just <input type="checkbox"/> <b>hübsch</b> cute, fair, handsome

Figure 19: A query translation and expansion assistant.

The feasibility of such a query assistant depends on the number of languages contained in a multilingual site. After all, bilingual dictionaries are used for each language pair. The addition of high numbers of bilingual dictionaries could be detrimental to the performance of the search and hence the usability of the provider.

For a monolingual search mask a query assistant would only be used to suggest further search terms. This could be a great support to novices of search as well as to advanced users of search services in refining their queries.

#### *11.1.2 Document classification*

In the previous Sections we mentioned the importance of classifying. Classification supports the users in finding the search results systematically, helps in navigating through the search results, and can be the best solution for multilingual search. The implementation of a classification system requires the implementation of metadata which attributes the corresponding categories to each record. Hence, implementing classifications are costly and complex. Further, classifiers need to be trained beforehand in order to categorize the documents.<sup>[13]</sup> Machine learning algorithms can be investigated to generate a hierarchy of classes for web document classification. Yet, such methods still require significant improvements.

We highly recommend the use of metadata within the DATAWOMSCI Provider. We have dealt with issues of metadata briefly in Section 10. If classifications are of common interest to all the institutions that are maintaining databases, then the resources for such an implementation need to be secured and an agreement made on the classification of the documents.

#### *11.1.3 Search tutorials*

A search tutorial can be prepared after analyzing the log files for most popular queries and often made mistakes. Such an analysis can be useful in understanding patterns of user interaction with the search mask. A search tutorial can attend to questions like formulating and refining queries, using boolean operators, finding the right search terms etc.

A search tutorial additionally can include general tips for the searching user. If a search history is available on the site, then users could be encouraged to make use of this service, which allows them to save their queries or search results.

#### *11.1.4 Explain search operators in natural languages*

Especially in monolingual search the use of boolean operators can be extremely useful. Unfortunately, many users have misconceptions on how these logical operators work.<sup>[4]</sup> We recommend the explanation of the boolean operators in natural languages. Aula and Kaki even recommend the interpretation of a query put together using boolean operators. For example the query "Slovenia AND "computer science" -engineer" would return:

You are now searching for documents that have both the term "Slovenia" and the phrase "computer science". Documents that have only one of these terms or the term "engineer" will not be considered.

Such an interpretation may be more confusing than helpful as users formulate longer queries. It is therefore recommended to use natural language terms –i.e. all, either, not– and to provide users with examples on the use of boolean operators.

### 11.1.5 Use of filters

If a DATAWOMSCI Service Provider comes to being, it may be interesting to implement filters. Filters would allow users to navigate the complete data of the provider or the search results according to given criteria.

The following criteria can be used as filters:

- country
- originating database
- language of original record
- by date (date of data creation or update)
- classification of disciplines

An alphabetical listing of the scientists may also be of interest for the users of the provider. An alphabetical listing would serve navigational searches, as defined at the beginning of this document. An alphabetical listing of institutions and data collections within a DATAWOMSCI Information Provider is highly recommended.

The screenshot shows a search interface with the following elements:

- Language selection: English (selected), German, French.
- Navigation: Search, Advanced Search, Help.
- Search History (dated 30.11.2004):
  - "computer engineer" (with a link to [computing](#))
  - [computer engineer](#)
  - [computer science](#)
- Search input: Search for "computer science"
- Language of query: English (dropdown)
- Find documents in:
  - English
  - German
  - French
- Find in databases:
  - Femdat
  - Kilden
  - cec-wys
  - SET
  - Femconsult
- Buttons: Start Search, Query Assistant.

Figure 20: A query search history integrated into the search mask. The user can use the search history to refine queries.

### 11.1.6 Search history

A search history allows the users to maintain their footsteps during a search as shown in Figure 20. This could be done by saving all queries in a session or by offering customization services that allow the user to save their history on the site. Similarly, the user may be able to save search results.

### 11.1.7 Search agent

Users who register may be allowed to configure their own search agent. This search agent would then periodically query the data collections, collect information of interest to the user, prepare a report and communicate it to the user. The communication can either be through an email or through a web page the user can login to. In order to support users popular agents configurations can be extracted from the logs and offered to the users as generic agents.

An agent can contain the following configurations:

- the periods in which information should be delivered
- how to communicate results: per email or web page
- notification on changes on the site or with search
- options of summary (if the search results should contain the complete datasets or just summaries)
- notification from other services like discussion forums or peer reviews

### 11.1.8 Site search box

A site search allows the user to search the contents other than the database of women scientists. Site search box is a simple field for search on the contents of the site (i.e. a DATAWOMSCI Provider) which allows the users to quickly search the site for further information. A central site search facility is especially useful if the content of the site grows rapidly and dynamically through user interaction or if the user is looking for a short cut to information related to the site. The site search is especially useful for searching discussion forums or archived mailing lists or peer reviews. The query language of the search box should work with logical operators.

If many interactive content services like archives of mailing lists or discussion forums exist on the DATAWOMSCI Provider then it is recommended to offer an extra site search page for site search. The following functionalities can be offered for such a site search:

- select which services: it should be possible for the user to search specific documents on the site, i.e. a selected discussion forum, only mailing lists, all articles etc.
- updated date range: search for data contributions in a specific time frame. For example "the discussion forum contributions of the last 24 hours".

- posted by: it should be possible to search the postings by author.
- sort order: the user decides in according to which criteria the search results are ordered: by date, by author, by relevance.

## 11.2 Display of Search Results

The browsability of search results needs to be improved in order to meet the increasing quality demand of the users and to support them in dealing with the vast amounts of information available through a search. Hence, search results should be presented in a way so that the searching user can determine what the document is about and how it is relevant to the users search.

In the case of a meta-database like the DATAWOMSCI Service or Information Provider the search results should make visible the databases from which the documents are originating. This contributes to making the data provider databases visible and also locates the scientist in her context. It also emphasizes the "gateway" role of the DATAWOMSCI Provider.

### 11.2.1 Summary of results

Search result display includes summaries of the information on the scientists whose data match the search. Studies on search result summaries emphasize the importance of query search terms reappearing in the summary. In other cases, for usability purposes in a specific field like women scientists' databases it may make more sense to have a standard summary format, which lists the most important information. If this is the case, it may still be useful to highlight the search terms in the complete record of the scientist.

If a static format for display of search results is decided on, then we recommend the use of the following fields:

- name of the scientist
- discipline
- last update or creation date of the data
- the language of the original record (in case of a multilingual data collection)
- the database in which the scientist is registered (including an icon or link)

For the format of the summary the following elements are recommended:

- the document's title: this will most likely be the name of the scientist in the case of a search on women scientists, and the name of the institution or the data collection in the case of a data collection search.
- no more than three lines from the record
- if the search terms are in the title or appear in the summary these should be highlighted
- if subject categories or keywords are associated with the record, these can be displayed

- the url: it is possible to discuss if showing the url, especially in the case each record is assigned its own url, is useful or an information overflow.

If multilinguality issues are of concern, machine translation of summaries may also be an option to support users.

### 11.2.2 Highlight search terms

Highlighting search terms in documents shows the context of the search terms in the documents found. Users may utilize such context information to refine their search or to quickly scan documents to find out if the document fits the criteria they are looking for.

### 11.2.3 Links to related content

When the user finds an article on the CiteSeer Scientific Literature Digital Library<sup>7</sup> s/he is also provided with links to articles that have similar content or links to other articles that cite the original article. This is another version of Amazon's "people who bought x also bought y" links. Such technique can be developed with women scientists' databases. Such links may be especially useful in multilingual search.

### 11.2.4 Exporting search results

It is highly recommended that the user can save, print, mail or export the results of a search. Saving search results on a customized page may also be of interest as we mentioned in the "search history". A saved search should include the query for the results and the date of the search.

## 12 Conclusion

The three models described in Section 9 were discussed in detail during the DATAWOMSCI Project. The recommendations of the DATAWOMSCI Consortium partners concerning the three alternatives are as follows:

**Multilingual DATAWOMSCI Service Provider:** the linking of the multilingual content from existing databases is the easiest implementation of the DATAWOMSCI Service Provider. Yet, the implementation of multilingual interfaces as well as cross-lingual search is complex and costly. The implementation of a Multilingual DATAWOMSCI Service Provider without multilingual interfaces and cross-lingual search is not recommended. If multilinguality is a priority then funding should be provided for the implementation after a survey of the interests and requirements of the users and stake holders of such a multilingual DATAWOMSCI Service Provider is completed.

---

<sup>7</sup> <http://citeseer.ist.psu.edu/>

**Monolingual DATAWOMSCI Service Provider:** the implementation of the Monolingual DATAWOMSCI Service Provider is highly recommended. The implementation of a Monolingual DATAWOMSCI Service Provider would require the cooperation of institutions and organisations that maintain women scientists' databases. This cooperation would also include extra costs for the maintainers to extend their databases (See Section 9.2). Therefore, funding should be secured and allocated to the maintainers in order to bring about the necessary extensions to their databases. In addition, long term funding should be provided for a open source support package for the implementation of a Monolingual DATAWOMSCI Service Provider. This package can be freely available to all interested parties and revised according to the feedback of the implementers and the changing women scientists' databases landscape.

**DATAWOMSCI Information Provider:** the implementation of the DATAWOMSCI Information Provider is also highly recommended. The DATAWOMSCI Information Provider is especially important for data collections on women scientists which are not online. This is the case for different regions in Europe. Many institutions and organisations do not have the funding and resources for developing such databases yet have many resources on women scientists. The DATAWOMSCI Information provider may contribute to making such institutions and organisations visible and improving cooperation among them.

If possible, the implementation of both a DATAWOMSCI Service Provider and a DATAWOMSCI Information Provider is highly recommended.

In addition, three of the partners also connected to the databases part of the feasibility study mentioned that their users show great acceptance in filling out their data in English, when necessary. This strengthened the arguments for a Monolingual DATAWOMSCI Service Provider.

The multilinguality of a DATAWOMSCI Information Provider may also be considered. The same recommendations as for the Multilingual DATAWOMSCI Service Provider mentioned above also hold here.

In any case, we are looking forward to using the future DATAWOMSCI Provider and hope that the feasibility study can function as a starting point for developing such a provider.

## References

- [1] ANSI/NISO. Information retrieval (z39.50): Application service definition and protocol specification. Technical report, National Information Standards Organization, 2002. revision of Z39.50-1995.
- [2] Phil Arko. Faq:international and multilingual web sites, 2004.
- [3] Cliff Ashcroft, Njomeza Kartalozzi, and Dan McQuillan. Lasa multikulti project: Multilingual websites and information in community languages, 2004.
- [4] Anne Aula and Mika Käki. Understanding expert search strategies for designing user-friendly search interfaces. In Proceedings of IADIS International WWW/Internet 2003 Conference, Algarve, November 2003.
- [5] Nuala Bennett, Katie Clark, Esther Gillie, and Lura Joseph. Access task force 4 report and recommendations. Technical report, May 2003.
- [6] Marshall Breeding. Understanding the protocol for metadata harvesting of the open archives initiative. *Computers in Libraries*, 22(8), September 2002.
- [7] Andrei Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
- [8] Michael Buckland, Aitao Chen, Hui-Min Chen, Youngin Kim, Byron Lam, Barbara Norgard, and Jacek Purat. Mapping entry vocabulary to unfamiliar metadata vocabularies. *D-Lib Magazine*, 5(1), January 1999.
- [9] P. Buitelaar, K. Netter, and Feiyu Xu. Integrating Different Strategies for Cross-Language Information Retrieval in the MIETTA project. In Proceedings of TWLT14, Enschede, The Netherlands, 1998.
- [10] Joseph Busch. What are the best practices in web site search?, 2004.
- [11] CeLeX. Celex, expert search service, visual quickstart guide. Technical report, Europa, 2001.
- [12] Antony Corfield, Mathew Dovey, Richard Mawby, and Colin Tatham. Z39.50 and xml - bridging the old and the new. In *WWW2002 Alternate Paper Tracks Proceedings*, 2002.
- [13] Hang Cui and Osmar R. Zaiane. Hierarchical structural approach to improving the browsability of web search engine results. In Proceedings of the 12th International Workshop on Database and Expert Systems Applications. IEEE Computer Society, September 2001.
- [14] Herbert Van de Sompel, Jeffrey A. Young, and Thomas B. Hickey. Using the oai-pmh ... differently. *D-Lib Magazine*, July 2003. <http://www.dlib.org/dlib/july03/young/07young.html>.
- [15] Susanne Dobratz and Birgit Matthaei. Open archives forum: Project deliverable. Technical report, Humboldt University, Berlin, September 2002.
- [16] DoCDat. DoCDat Data Collection Manual. London School of Hygiene and Tropical Medicine, 2002.
- [17] DoCDat. Docdat development, 2002.
- [18] Martin Dürst. Faq:multilingual forms, 2003.
- [19] EuroCRIS. Code of good practice for current research information systems. Technical report, EuroCRIS, 1998.

- [20] EuroCRIS. Cerif 2000 toolkit. Technical report, EuroCRIS, 2000.
- [21] Peter Evans. Z39.50: Part 1 - an overview. Biblio Tech Review, April 1999. [http://www.biblio-tech.com/html/z39\\_50.html](http://www.biblio-tech.com/html/z39_50.html).
- [22] Sonya Finnigan and Nigel Ward. Z39.50 made simple, 1997. <http://archive.dstc.edu.au/DDU/projects/ZINC/zsimple.html>.
- [23] Hachim Haddouti and Rene Cluzel. Survey: Multilingual text retrieval and access, 1999.
- [24] Sebastian Hammer and John Favaro. Z39.50 and the world wide web. D-Lib Magazine, March 1996.
- [25] Mathew Hillier. Multilingual website usability: Cultural context. In Proceedings of the International on Electronic Commerce, 2002.
- [26] Richard Ishida. Faq:monolingual vs. multilingual web sites, 2004.
- [27] Wessel Kraaij. The EU project 'Twenty-One' and cross-language IR, 1997.
- [28] Carl Lagoze and Herbert Van de Sompel. The open archives initiative: Building a low-barrier framework. JCDL'01, 25-28. June 2001.
- [29] Carl Lagoze, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. Implementation guidelines for the open archives initiative protocol for meta-data harvesting. Technical report, Open Archives Initiative, 2002. Version 2.0.
- [30] Ray R. Larson, Fredric Gey, and Aitao Chen. Harvesting translingual vocabulary mappings for multilingual digital libraries. In The Second ACM/IEEE Joint Conference on Digital Libraries, pages 185–190, July 2002.
- [31] Andrei Lopatenko, Anne Asserson, and Keith G. Jeffrey. CERIF – Information Retrieval of Research Information in a Distributed heterogenous environment. In Proceedings of the 6th International Conference on Current Reseach Information Systems, Kassel, Germany, 2002.
- [32] Sigfrid Lundberg. St. laurentius digital manuscript library: An excursion along the border between resource discovery and resource description. Technical report, netLab, Lund University Libraries, August 2002.
- [33] Clifford A. Lynch. The z39.50 information retrieval standard, part i: A strategic view of its past, present and future. D-Lib Magazine, April 1997.
- [34] V. Lyytikäinen, P. Tiitinen, and A. Salminen. Xml metadata for accessing heterogeneous legal databases. In Proceedings of the XML Europe 2001 conference, 2001.
- [35] Gerry McKiernan. E-profile: Open archive initiative data providers, part 1. Library Hi Tech News, (3), 2004.
- [36] William E. Moen. Resource discovery using z39.50: Promise and reality. In Bicentennial Conference on Bibliographic Control for the new Millennium: Confronting the Challenges of Networked Resources and the Web, November 2000.
- [37] Sandra Payette and Oya Y. Rieger. Z39.50, the user's perspective. D-Lib Magazine, April 1997.
- [38] Maragaret St. Pierre. Z39.50 for full-text search and retrieval, 1995.

- [39] Stephen Pinfield, Mike Gardner, and John MacColl. Setting up an institutional e-print archive. *Ariadne*, April 2002. <http://www.ariadne.ac.uk/issue31/eprint-archives/>.
- [40] Pavel Smrz, Anna Sinopalnikova, and Martin Povolny. Thesauri and ontologies for digital libraries. In *Proceedings of the 5th Russian Conference on Digital Libraries RCDL2003*, St.-Petersburg, Russia, 2003.
- [41] Danny Sofer. *Lasa knowledgebase: Multilingual content management*, 2003.
- [42] Heinrich Stamerjohanns and Bruno Klotz-Berendes. *Tutorium: Die open archive[s] initiative*, March 2003.
- [43] UNESCO. *International standard classification of education*. Technical report, United Nations Educational, Scientific and Cultural Organization, 1997.
- [44] Simeon Warner. Exposing and harvesting metadata using the oai metadata harvesting protocol: A tutorial. *High Energy Physics Libraries Webzine*, 4. June 2001. <http://library.cern.ch/HEPLW/4/papers/3/>.
- [45] Yiming Yang, Jaime G. Carbonell, Ralf D. Brown, and Robert E. Frederking. Translingual information retrieval: Learning from bilingual corpora. *AI Journal Special Issue: Best of IJCAI-97*, 103:323–345, 1998.
- [46] Eric H. Zimmerman. *Cris-cross: Current research information systems at a crossroads*. In *Proceedings of the 6th International Conference on Current Research Information Systems*, University of Kassel. EuroCRIS, August 2002e



## Chapter 6 – Follow up of the Project

This publication is recommended to all institutions, organisations etc. wanting to develop or maintain women scientist databases. It may also be useful in case of databases which do not only contain women scientists. Here, we would like to make further recommendations for the enhancement of the European women scientists' database landscape.

On the European level further funding especially for building up new international databases or database co-operations is necessary. As a result of the DATAWOMSCI feasibility study, the development of a meta-database which would allow the linking of existing databases is recommended. Additional funding for the development of such a meta-database is necessary.

The possibilities for linking existing or future databases are described in Chapter 5 (feasibility study) through different meta-database models. In Chapter 4 (best practice guideline) we have taken the development of the future meta-database in our recommendations into consideration. All recommendations on "data exchange" reflect these considerations.

Further, in order to strengthen European wide standardisation of women scientist databases and in order to improve the quality of existing and future women scientist databases we recommend an external expert group founded by the European Commission. Similar to the group of experts working on Current Research Information Systems (CRIS) an expert group of software programmers, database maintainers and international experts with knowledge about the scientific institutions in the different European countries and gender equality programmes should be established.

The objectives of this expert group may be the exchange of experiences, development of standards for women scientist databases, the review of subject indexes to include areas in which women scientists are active (i.e. gender studies, gender equality etc.), harmonisation of national and international classification systems etc.

The Chapter on "Women Scientists in Europe – Databases and other Resources" (Chapter 3) shows that the state of women scientists databases or data collections in the different regions and countries varies immensely. In some countries no information on existing databases was available. This may be due to the absence of such databases or the lack of European visibility. Therefore, in these countries, funding is required to either build new databases or to enhance the visibility and accessibility of existing databases. We recommend that national governments provide funding for such projects.

In other countries databases exist which may be improved according to European standards and for data exchange. Existing databases also require funding for maintenance and for revisions. We recommend that national governments provide funding for the maintenance, revisions and harmonisation with European standards in those countries where they are already available.

Due to the fact that Germany has the greatest number of established databases we recommend no more funding for building up new databases but further funding for supporting co-operations between the existing databases, for merging with other databases and about developing synergies.

In the Northern European countries (i.e. Sweden, Finland and Iceland), the United Kingdom and Ireland should be put more effort on building up further interdisciplinary databases with online access. For Norway and Denmark we suggest that the existing databases are strengthened with funding for development and maintenance.

In the CEE countries there is a recently (2004) developed database which is funded by the European Commission covering the Czech Republic, Hungary, Slovakia and Slovenia. We recommend observing the further evolution of this CEC-WYS-database in order to decide in a foreseeable period of time how far this model works and how far an extension of this database towards other countries is required. Nevertheless, it is necessary to use already existing experiences with databases of women scientists in other European countries in order to take advances in this geographical area.

Last but not least we recommend more cooperation between the different national and international databases. Currently, a European Platform of Women Scientists is being developed. The objectives of this Platform are to network the networks, promote women scientists and to function as a communication channel between women scientists and the EU. We therefore recommend that the institutions and organisations maintaining women scientists' databases use the European Platform of Women Scientists as a co-ordination and first contact point in the future. In addition, if a DATAWOMSCI Platform consisting of a DATAWOMSCI Service Provider and/or a DATAWOMSCI Information Provider (see Chapter 5, Section 9.4) will be developed the European Platform of Women Scientists would be the ideal location for the structural integration.