

EICSTES DELIVERABLE D3.1

DIRECTORY OF THE EUROPEAN R&D PUBLIC SYSTEM

Isidro F. Aguillo
CINDOC-CSIC
isidro@cindoc.csic.es

January 2002

Presentation of the Directory that it is available from the central website of the project

1. Introduction.

The European R&D public sector has been the object of several studies involving description and evaluation using traditional bibliometric and scientometrics methodologies. Today it is possible to add to these tools new quantitative methods based on the diffusion and utilization of the Internet in R&D organisations. The cybermetric indicators derived from the data about the web presence of such institutions could improve our knowledge of the scientific communication and innovation processes.

One of the objectives of the EICSTES project is to discover and describe the new economy patterns using academic and research institutional websites as source of data. New indicators are needed, specially those derived from the web that are not linked with the traditional units, scientific publications. The point is to use a wider scenario, involving all the activities of the R&D institutions and not only those leading to publication.

That means a global approach, using institutional presence in the web as a whole. We defined a new unit, the institutional website, or simply the site as a group of hierarchically linked pages that represent an organization with an autonomous documental presence in the Web, clearly recognised from other similar units by formal and content characteristics.

The site is not an easy unit to identify and describe due to the Internet general problems of lack of standardization and different levels of formal and content quality heterogeneity:

- Sometimes there is no association between physical address and autonomous presence.
- The authorship or mention of responsibility of the websites are not clearly stated in many pages.
- Research groups, virtual groups (consortia) and undefined organisations have good quality presence with an autonomous site.
- Nested websites are a frequent situation.

Due to these problems we decided to compile first by hand a Directory of all the websites developed by public institutions related with R&D topics in the EU.

2. Methodology.

Obviously there is no previous list of institutions with indication of their web address covering in detail all the EU countries. Some directories were available but with

obsolete data and covering only some regions and even without the level of depth required.

The Universities and Research Councils homepages offered listings of their departments and units, but mixing true websites and non-independent pages. Sometimes, new or informal sites were not included.

For these reasons, an exhaustive review was done using as a basis printed Directories (World of Learning) and web search engines and directories. Each main institution was carefully scanned to recover all the websites and the links were followed for extra sites. Additional effort was made to identify other websites not cited in the official directories. All the identified units were double-checked for website status.

The websites were described using human indexing that involves:

- Subject classification according to the UNESCO thematic codes
- Geographic classification (Eurostat NUTS codes)
- Institutional level

The original name of the site is preserved, but a new coined descriptive one in English is provided as entry point for the Directory.

Besides, we obtain automatic-generated statistics using agent technology, including values for website size (pages) and number and type of hipertextual links. The data is obtained with a minimum of load for the servers as the pages are not downloaded. However the level of quantitative description obtained is very high, as the reports contains the complete listing of all the outgoing links of the site.

It is scheduled that a full complete new quantitative description of the websites will take place at the end of 2002. Previous to that new sampling, several link checking tests and updates made with the feedback of colleagues and the website comments will be made.

3. Results.

At the end of 2001, with an almost complete coverage of University websites and a good sample of the non-Universities websites the database consists of more than 24.000 items. That figure is about a 20% over the estimated population in the proposal.

It is presumed that the final value will be over 26.000 sites with more than 40 million web-objects collected.

The Table shows that exists some inconsistencies among countries, some of them explained by the different models of university organisation, but there is plenty of other possibilities awaiting further analysis.

The websites are large, some of them really very large with a big number of external links, so the population scanned offer enough data for deriving cybermetric indicators (WP 8).

| Country | University Sites | Pages | Pages per Site | External Links | Total Objects |
|---------------------|------------------|-------------------|----------------|-------------------|-------------------|
| Germany* | 6.539 | 3.800.287 | 581,17 | 8.798.680 | 9.522.935 |
| United Kingdom* | 6.504 | 2.795.208 | 429,77 | 5.624.001 | 9.503.321 |
| Spain | 2.158 | 693.764 | 321,48 | 456.555 | 2.238.265 |
| Italy | 1.188 | 1.023.415 | 861,46 | 444.413 | 1.723.412 |
| Netherlands | 890 | 1.810.831 | 2034,64 | 484.738 | 2.574.169 |
| France | 875 | 469.578 | 536,66 | 234.351 | 1.322.134 |
| Finland | 815 | 533.540 | 654,65 | 387.178 | 1.897.078 |
| Austria | 800 | 458.749 | 573,44 | 220.083 | 1.158.275 |
| Belgium | 761 | 218.759 | 287,46 | 120.907 | 605.527 |
| Portugal | 570 | 157.723 | 276,71 | 127.610 | 466.305 |
| Sweden | 323 | 532.628 | 1649,00 | 270.954 | 1.393.740 |
| Ireland | 322 | 109.737 | 340,80 | 69.984 | 296.108 |
| Greece | 216 | 137.932 | 638,57 | 64.320 | 301.694 |
| Denmark | 191 | 213.894 | 1119,86 | 124.270 | 553.911 |
| Luxembourg | 24 | 9.935 | 413,96 | 3.556 | 26.239 |
| Total | 22.176 | 12.965.980 | 584,69 | 17.431.600 | 33.583.113 |
| *(work in progress) | | | | | |

4. Dynamic website.

Due to the large volume of data currently available from the Directory, the access has been made public through a database gateway from a dynamic website. This gateway is part of the contents of the project central webserver. Its address is:

<http://www.eicstes.org/eustat.htm>

All the records are easily accessed from this URL address, that provides tables with the websites ordered according a series of criteria.

This system allows a easy correction and immediate update directly from the database to the dynamic website.

Figure shows a screen capture with two records:

Choose a Country Go

Choose a Section Go

Choose a University Go

Description

Workplan

Results

Resources

Internal
Forum



NETHERLANDS



R&D PUBLIC SECTOR

- ▶ Universities
- ▶ R&D Institutes
- ▶ Hospitals
- ▶ Libraries
- ▶ Other Bodies

R&D PRIVATE SECTOR

- ▶ Profits
- ▶ Non Profit

REFERENCE RESOURCES

- ▶ Portals
- ▶ Directories
- ▶ Virtual Libraries
- ▶ Databases

GENERAL

R & D

INTERNET

Help

First

Last

Previous

Next

Records 1 to 10 of 105

| SITES | nat | pages | links | objects |
|--|-----------------|-------|--------|---------|
| University of Groningen | n1110 Groningen | 66260 | 163653 | 368721 |
| Department of Molecular Dynamics <small>Orig. name: Department of Molecular Dynamics Spanish name: Grupo de Dinámica Molecular URL: http://zagradf1.chem.rug.nl/tod.html</small> | | 649 | 149 | 1226 |
| Faculty of Law <small>Orig. name: Faculty of Law Spanish name: Facultad de Derecho</small> | | 372 | 89 | 747 |