

EICSTES DELIVERABLE D1.3

EICSTES INTERNAL MEETINGS PROCEEDINGS

MINUTES OF THE KICK-OFF MEETING

IDESCAT, Barcelona. January 18-19th, 2001

1. Agenda of the meeting

Thursday 18	
10:00-11:00	Reception, informal meeting and distribution of documentation
11:00-11:15	Introduction of the host and brief presentation of the partners of the consortium
11:15-11:45	Report of the coordinator. Consortium agreement and related topics Payment and Cost Claims procedure Reporting, Deliverables Mailing lists Rich modes of communication (eg videoconferencing) Dissemination plan Future meetings
11:45-12:00	Report about Luxembourg meeting (EPROS-DOSIS conference)
12:00-14:00	Open discussion
14:00	Lunch
15:00-16:00	Discussion (and preparation of) about end users session
16:00-18:00	Website: Design and contents. Web document repository
21:00	Consortium Diner

Friday 19	
9:30	Pre-meeting of consortium
10:00	End-user session I: Spanish statisticians Bienvenida (Welcome) <i>Enric Ripoll, subdirector de Asistencia Técnica Estadística, IDESCAT Catalunya, la nueva economía y el Idescat.</i> Experiencias en la recopilación de datos. (Catalonia, new economy and IDESCAT) <i>Vicenç Meléndez, IDESCAT</i> Presentación de algunos resultados de MS Study II. Mercado de la información electrónica (Presentation of the MS Study II results) <i>Tomàs Baiget, IDESCAT</i> EICSTES <i>Isidro Aguillo, CSIC</i>
14:00-15:00	Lunch

15:00-16:00	WP2 and WP3. CTI presentation
16:00-16:30	WP4 and WP6. University of Surrey previous experiences
16:30-18:00	Other WP. The "state of art" review

2. Internal meeting.

Tomàs Baiget welcomed the participants and made a short introduction of IDESCAT, the regional statistical office of Catalonia (Spain).

Guided tour through the building. Besides the own computer centre for producing statistics, in the same building there is the enormous computer network system of Meff, a company distributing financial information and providing continuous e-stock exchange services only for a selection of big investors (banks, agents and other).

Introduction of participants:

Isidro Aguillo, CINDOC: Biologist with interest in documentation. Special field is study of internet for last 10 years, electronic market. Quantitative methods for description of internet. Director of Cybermetrics e-journal. Coordination of EICSTES.

Gaston Heimeriks, Univ. van Amsterdam: Biochemist. Working in an European project about self organisation of the information society, using bibliometric methods to analyse activities, mapping technosciences.

Xavier Polanco, Institut National pour l'Information Scientifique et Technique. Applies AI-methods for documentation of scientific activities, knowledge representations, mapping, clustering, neural maps. These methods now should be applied on web data.

Thanasis Poulakidas, Computer Technology Institute: Computer science engineering in Patras. Studied computer science in Santa Barbara (CA, USA). Expert on compression algorithms.

Giorgos Ikonou, CTI: Computer engineering in Patras, works on PhD, information retrieval, network routing, network administrator, knowledge about physical internet.

Nigel Gilbert, Univ. of Surrey: Prof. of sociology in UniS. After a first degree in engineering, doctor in scientific knowledge. 20 years in academia. Project using first microcomputers for social security matters –this gave him a wide national recognition-. Computer science projects. In the European Framework Programme 2 or 3: speech recognition project, but allways contributing to social aspects. Structure of dialogue. Computer simulation in sociology. Digital world research centre (interdisciplinary). Involved in starting up the Digital Word Research Centre, 1976, Surrey. ICT's social aspects. First electronic journal "Sociological Research Online (SRO)", 1996.

Moysis A. Boudourides, Computer Technology Institute (CTI): Chemist. Considers himself working in a mixture os sociology and computing. Mathematics about dynamical systems, complexity algorithms. Later works about sociological topics. Internet analysis. In his opinion, to understand the new society it is necessary to study its social aspects.

Peter van den Besselaar, Univ. van Amsterdam: Degrees in mechanical engineering, mathematics and philosophy. Works in a computer science group at UvAm. Technical and sociological research. Social informatics. Informetrics. Scientometric research. European Framework programs.

Clemens Widhalm, ARCS: Technical physicist, experiences with stone mason industry (geology). Works at Austria's biggest non-universitarian research center (700 employees). Department of Technology Management, performs technology monitorings for companies analysing literature and patent databases. Developed bibliometric visualisation tool, mapping based on concept of co-occurrences.

Tomàs Baiget, IDESCAT: Degree in electric engineering and telecommunications. Expert in online databases since 1973. Later works on organisation of documentation. Statistics about databases and online information (e-content). Produces a journal about libraries, information science & documentation. Course director and teacher at a technological institute.

Vicenç Meléndez, IDESCAT: Degree on economics and expert on documentation. Good knowledge of international statistical data sources. In recent years has worked on information society indicators. Teacher at various courses and seminars.

Report of the coordinator:

Project Gantt-chart. Since project is delayed by one month, an extension by one month will be applied.

Isidro suggests to invite a Commission officer at the next EICSTES meeting.

Consortium Agreement:

Standard contract, except 11.5 (Software results) that was added by legal department of CINDOC. Text has to be shown to legal representatives of partners soon (before end of January).

Isidro surrenders original contract page (original signatures) to each partner.

Payments: First funding (40% advance payment) already should had been received.

Cost claims procedure: **Isidro** shows a standard form (model) by the EC. See handout. Isidro needs such forms only. Partners have to keep all the receipts, etc., for at least 2 years after the project has finished. Original forms will be sent in the near future.

Reporting: Each 6 months. **Isidro** proposes just summaries plus enclosed deliverables. Deliverables should be published in the project internal network electronically.

Mailing lists:

Moysis: Weborg is first mailing list from times of project preparation. 2nd mailing list is for CTI-group. Should be mailing lists subdivided in workgroups? Weborg should become an open forum for discussion.

Isidro: Weborg will be public forum. Private forum is needed.

Peter: 2 private lists should be installed. One restricted for organisational and administrative matters (EICSTES) and one for technical matters (ECSTASIS) more open, but only extended to partners' collaborators.

Other forms of communication:

Giorgos could help installing a video conference via MS NetMeeting. IP-addresses have to be known. **Nigel** is afraid that the audio-transmission could be not good enough. Time discipline for all partners to perform a online meeting could be a problem as well.

Moysis: Videoconferencing could be developed to better contact with end users.

Moysis and **Giorgos** will send technical specifications to experiment videoconferencing among the consortium members.

Dissemination plan:

Isidro prepares brochure about EICSTES. Draft will be surrendered very soon. Parts of projects may be presented at conferences. ISSI-conference in Sydney (in July) could be an interesting date for presentations. **Isidro** plans a general presentation about EICSTES there.

Moysis would like to have relevant information about events, etc., on a project web site.

Report on EPROS-DOSIS conference in Luxembourg, 4-6 Dec. 2001
(**Moysis**, see also handout)

EPROS: Eurostat Research Plan in Official Statistics
DOSIS: Development of Statistical Information Systems
SINE: Statistical Indicators for the New Economy

Current projects (SINE and others):

BEEP: Best eEurope Practices.

ECaTT: Benchmarking Progress on Electronic Commerce & New Methods of Work

EICSTES

EMMERGENCE: Estimation and Mapping of Employment Relocation in a Global Economy in the New Communications Environment.

ERMIS: Electronic Commerce Measurement Through Intelligent Agents

FAMILIES: Families, Work and IST

NewKInd: New Indicators for Knowledge Based Economy

SIBIS: Statistical Indicators for Benchmarking the Information Society

STAR: Socio-economic impacts of new technologies and services

STING: Evaluation of Scientific & Technological Innovation and Progress in Europe through Patents (data from patents automatically will be processed in order to do factor analysis.)

ETK and NTTS 2001, Crete (see also handout)

ETK: Exchange of Technology Know-how

NTTS: New Techniques and Technologies for Statistics.

NEW CALLS in FP5:

CPA 7 and CPA 8 merge to new CPA 7: Socio-economic issues and indicators. There will be 1 or 2 calls under new CPA 7; publication of the first one after January 15.

They are building projects clusters.

Lunch

Future EICSTES meetings:

Second:

Crete could be a good place because of above conferences in June. Time will be the weekend June 22/23/24. **Moysis** to arrange for a meeting place.

Third:

EUROSTAT will plan a meeting beginning of December 2001 in Luxembourg. A project meeting therefore would be convenient in Nancy (organised by **Xavier**) may be 10th till 11th of December.

Discussion about and Preparation of End Users Session I (statisticians)

Proposal by **Isidro**: To develop a draft questionnaire and distribute it tomorrow at the End-User Session. Answers could be collected tomorrow or allow people send later per email.

End users expected for the tomorrow morning session are professional statisticians, people from public services and universities.

Discussion about possible questions:

Xavier: Kind of information needed about physical internet (internet infrastructure indicators).

Nigel: In what areas do you have problems in getting data on at the very moment?

The group discusses different issues about defining "physical internet".

- static (e.g. number of lines, bandwidth, topology...)
- dynamic (e.g. data flow, links, portals...)
- economics
- quality (e.g. performance...)
- who is doing what in European networks?

Giorgos tells about RIPE, where the location of each European site is stored.

RIPE (Réseaux IP Européens) is a collaborative organisation open to organisations and individuals, operating wide area IP networks in Europe and beyond. The objective of RIPE is to ensure the administrative and technical coordination necessary to enable operation of a pan-European IP network. RIPE does not operate a network of its own.

<http://www.ripe.net>

Nigel: There are 4 different types of geographical locations according to:

- Administrator
- Machine / server
- IP domain

- Content providers

Moysis believes that the situation has to be studied in a whole with all sociological aspects: gender, migrations, education levels, income, etc.

Vicenç insists on the importance of the geographical locations that will allow to compare countries and regions.

Discussion about physical internet will continue (see below).

Final design of preliminary questionnaire.

It is decided that a short pre-questionnaire will be given to the end users right away with the following 3 questions:

1. Please describe your current experience with internet related indicators and statistics of the information society and the new economy:
2. What are your main needs for statistics in this field? Please provide as much detail as possible:
3. What do you expect from the EICSTES project?

Later, a longer questionnaire will be e-mailed both to the end users attendants, as well as to other statisticians that showed interest in participating at the meeting but were not able to come.

Consortium dinner.

19. 1. 2001

Joining people:

Carmen Vidal (Director, CINDOC). Graduated in Philology, master in Public Administration.

Víctor-Manuel Pareja (CINDOC). Webmaster.

Pre-meeting at 9:30

End user session at 10:00

Vicenç: The New Economy and the Idescat: Experiences in Data Collection (handout)

Tomàs: MS Study II

The markets for electronic information service in the European Economic Area (handout)

Isidro: EICSTES (handout)

Moysis and **Giorgos** (CTI)

Web data collection, agents etc.

Clemens

Software demonstration: Visualising networks

Lunch

The project EICSTES group continues the work.

Real data to support web results.

Isidro rises the question about which kind of no-web data could allow to correlate web measurements and to validate the trends discovered through the use of intelligent agents on the internet. There is a need to compare internet results with other "real-world" measurements. This also would allow to see which interactions exist between the two worlds.

Gaston comments that sometimes there is a feedback between two phenomena. For instance technologies have an influence in how science is communicated but this has an influence in how scientific information is produced.

Moysis has some concern about "case studies" and suggests to be much more specific with services like e-commerce, e-health, e-education...

Tomàs observes that EICSTES seems to potentially cover too much broad areas and that perhaps some specific indicators should be selected.

Vicenç offers to the group a list of more than 100 information society indicators that he has compiled and in some cases tested practically.

Data presentation or visualisation.

Raw data obtained by the project will have to be entered into databases and then visualised.

Isidro says that EICSTES will publish its results on the web, on cd-rom and also on paper. After the publication there will be a need for data or results interpretation.

Vicenç stresses the need of written statements. **Tomàs** says that it depends on the self-explanation nature of each graphical representation. **Moysis** has the opinion that now it is too early for this discussion.

State of the art.

Nigel is worried about the current situation in other countries (i. e., in the USA). How do we know that we are not reinventing the wheel?

The group discusses which work packages include to carry out an state of the art (6, 8 and 9). **Isidro** and **Moysis** have already discussed how to implement a repository of documents, in some cases available with some restrictions, which will facilitate the above studies.

Physical internet (continued)

Thanasis:

- Connectivity (bandwidth)
- Measurements from a point (bits, traffic...)
- End-to-end characteristics
- Web caching (how web is accessed)

Thanasis continues explaining the topology and the characteristics of TEN-155 backbone and the connectivity of GRNET (Greece).

All partners should supply router data (detailed tables)

Example see handout "Measurement from point to point (2)" and "Asking from partners".

>sh ip accounting
Source, Destination, Packets, Bytes

Giorgos plans to send a software (agent) to the consortium-partners.

Isidro (WP3):

MS Site Analyst

Sites are described by e. g.

- Unesco codes (thematic), partly comparable with NACE-codes
- NUTs 3 (geographic)
- Type of institution (univ...)

Isidro states the difference between "web site", "web server" and "web location" (f. i., group of people working together, like an university dept., that can not be distinguished automatically).

Work started at CINDOC with (around 7.000) university sites.

Nigel (WP 4, Site analysis tool):

Explains a tool they have at UniS for monitoring web usage:

- Log files in the server are analysed using an open software analysis program.
- Record data base is obtained, with record linkings.
- Transition matrix.
- Identify popular routes, typical user behaviours... using cluster analysis, factor analysis and multidimensional scale analysis (MDS).
- Correlation with the page features and drawing conclusions.
- Design recommendations to improve the web.

Problems and issues

- Web caching. If a user connects a 2nd time he receives the page from its local proxy and then it is no possible to know this. A solution is to set the page expiration time to zero.
- Markow chain. f true simplifies some analysis.
- Competition. Commercially available valuable analysis tools.
- Integration with other WP (e. g., visualisation).
- Experience with 2 journals: JASSS (20.000 pages/month) and SRO (3.500 pages/day). Other possible experimentation with ICL (a company offering hosting and other IT services) and Cordis.

Nigel (WP 6, Intermediaries in the digital economy):

The two main areas here were:

- Conceptualisation, functions of intermediaries.
- Data collection and analysis.

Tomàs is worried about the future "intermediaries" definition inside the project, because this is a very general word that can cause confusion. He thinks in the many kinds of intermediaries in the information chain or added value chain, not only webs.

Moysis points out the small difference between Amazon, a books intermediary, and any other e-shop selling, f. i., appliances.

Any web page providing a link o referring to something could be an intermediary?

Nigel explained having difficult personnel resources (two researchers previously appointed for the project are now not available and still has not been possible a

replacement). This could cause some delay but **Isidro** proposes some flexibility for waiting for one of them that is ill.

Xavier (WP4, Web database study)

Explains Miri@d software that allows to carry out descriptive statistical analysis on databases (PASCAL), which data are retrieved by users, time spent, authors, descriptors used... and see the evolution along the years. They have 10.000 user searches registered.

Isidro:

Proposes to discuss "new economy" topic in detail in Crete.

MINUTES OF THE 2ND PROJECT MEETING

Crete. June 23-24th, 2001

1. Participants

The meeting was hosted by CTI at Hersonnisos, Crete the 23rd and 24th of June 2001.

The participants were:

Name	Email	Organisation
Isidro F. Aguillo	isidro@cindoc.csic.es	CSIC
Victor Pareja	vmpareja@cindoc.csic.es	CSIC
Clemens Widhalm	clemens.widhalm@arcs.ac.at	ARCS
Ivana Roche	roche@inist.fr	CNRS
Moysis Boudourides	mboudour@upatras.gr	CTI
Giorgos Ikonou	georgeik@cti.gr	CTI
Manolis Mavrikakis	manolism@master.math.upatras.gr	CTI
Athanasios Poulakidas	poulak@cti.gr	CTI
Peter van den Besselaar	peter@swi.psy.uva.nl	UvA
Gaston Heimeriks	gaston@swi.psy.uva.nl	UvA
Nigel Gilbert	Nigel.Gilbert@soc.surrey.ac.uk	UNIS
Peter Morris	P.Morris@surrey.ac.uk	UNIS
Janet Vaux	j.vaux@soc.surrey.ac.uk	UNIS

No IDESCAT people attended the meeting because their main role is latter in the project. The Project Officer attended unofficially the beginning of the meeting.

2. Agenda of the meeting

The agenda of the meeting was shaped as follows:

Saturday, 23 June 2001			
Start	End	Topic	Speaker
9:30 AM	10:00 AM	Welcome by Project Manager and briefing by Project Officer	Isidro F. Aguillo and Project Officer
10:00 AM	10:45 AM	WP2 presentation	Athanasios Poulakidas and Giorgos Ikonou
10:45 AM	10:55 AM	Website presentation	Isidro F. Aguillo
10:55 AM	11:00 AM	CTI's website presentation	Moysis Boudourides and Manolis Mavrikakis
11:00 AM	11:20 AM	Analysis of CTI's collected data (demo)	Ivana Roche
11:20 AM	11:35 AM	Visualization and analysis of UNIS log data about web usage	Clemens Widhalm
11:35 AM	12:00 PM	<i>Coffee break</i>	
12:00 PM	12:25 PM	Presentation on WP4 and WP6	Nigel Gilbert
12:25 PM	12:55 PM	State-of-the-art in	Gaston Heimeriks

		Scientometrics and Webometrics, empirical analysis	
1:00 PM	3:00 PM	<i>Lunch break</i>	
3:00 PM	3:35 PM	MIRI@D	Ivana Roche
3:35 PM	4:05 PM	Intermediaries, transit sites and navigation on the web	Janet Vaux
4:05 PM	4:30 PM	New directions of Internet research	Moysis Boudourides
4:30 PM	5:00 PM	<i>Coffee break</i>	
5:00 PM	5:45 PM	Cybermetrics: Theoretical basis and preliminary results from EICSTES project	Isidro F. Aguillo
5:45 PM	6:30 PM	Web data collection (demo)	Manolis Mavrikakis
6:30 PM	7:00 PM	Discussion	All

Sunday, 24 June 2001			
Start	End	Topic	Speaker
9:00 AM	11:00 AM	Managerial meeting	All
11:00 AM	11:20 AM	<i>Coffee break</i>	
11 :20 AM	11:40 AM	Discussion on Eastern European partner	All
11:40 AM	12:30 PM	Technical discussions	All

3. Physical Internet data

CTI presented their collection of data and tools. The Project Officer commented on the unavailability of data from public but not educational networks. It was noted that these data are hard to get but, once available, the methodology to collect them is ready. Also, non-physical data on this area of the Internet are collected in other WPs.

4. Web data

CTI demonstrated their scripts and tools for collection of web data and described what data are collected. CNRS demonstrated their analysis of the web data collected by CTI. In the future, it is planned to automate the process and verify it by checking the content of the web pages.

ARCS stated that it prefers the data in another format, in order to process them with their visualization tools, and it was found that transforming one format to the other is easy.

CSIC noted that periodic re-searching is more important than trying multiple search engines since the web grows rapidly. It was decided to evaluate the search engines by checking their "performance" on specific sites: first a list of items to check will be compiled and then the checking will be performed.

UvA raised the issue of granularity for co-citation analysis on the web: if large sites are checked (e.g., universities), a co-citation may have little significance. They suggest lower granularity and use of thresholds to get rid of "insignificant" links. CTI argued that all links are significant and granularity may not be a big issue. It was suggested to perform an analysis for 2-3 case studies at varying granularity levels.

CTI presented an approach for Internet research, by using social network analysis, ethnography and chaos theory analysis.

UvA presented their contribution to the state-of-the-art in Scientometrics and Webometrics. They also discussed their empirical analysis on universities' linking (citation) and the issues involved.

UNIS presented an overview of WP6 and the issues involved. They noted that checking the content is a second approach to the problem, in contrast with the other approach that automatically collects, processes and visualizes the data. Aim for the two approaches is to converge. CSIC noted that the web is not uniform and needs to be "diced". UNIS suggested Britannica as a case study.

5. Usage data

UNIS has a collection of logs about the users' accesses to <http://jasss.soc.surrey.ac.uk/> and <http://www.surrey.ac.uk/>. ARCS has inserted the web usage data collected by UNIS in their visualization tool and demonstrated the results.

CNRS presented the architecture of MIRI@D.

6. Web sites

The project's website (<http://www.eicstes.org/>) was presented. The site will be copied to CD-ROMs and distributed to the partners. Then, a mirror of the main pages of the site, translated to local language, should be operational at each partner by September 15th, 2001.

CTI's website about the project (<http://nicomedia.math.upatras.gr/eicstes/>) was presented.

7. Managerial issues

A financial report, a report of activities for months 1-6 and a progress report are due by July 6th. The Project Manager will distribute templates to the partners. Three signed copies and photocopies of all receipts are required.

The Project Manager will distribute new version of the Consortium Agreement, after checking ARCS' suggested modifications.

An Exploitation Plan should be prepared before October's review meeting. The Project Manager will take the survey performed during the 1st meeting at Barcelona. The question of doing another survey on the Internet was discussed. In preparation for the October review, collecting tables and maps was also discussed.

8. Deliverables

The Project Manager suggests that all deliverables should be ready a month in advance their due date and distributed internally to all partners for comments.

D2.1 has been distributed to all partners. Comments are expected till the 3rd week of July. The plan is to deliver it that week.

D3.1 is already on the web.

A draft of D10.1 will be prepared by the 2nd or 3rd week of July by CSIC.

9. Next meeting

UvA will host the next meeting at Amsterdam. Proposed dates are 10 (all day) and 11 (till afternoon) of December 2001.