

Author	Gian Piero Zarri	Organisation	CNRS		
Title	Report on the EUFORBIA kick-off meeting				
Publication Date	9/2/2001	Document Reference	CNRS/Note/1	Task Reference	T1.1/1.2/1.3
Circulation (C,R,P)	<b>R</b>	Document Type	Note	Document Version	1.1
Deliverable Reference	D1				

## EUFORBIA KICK-OFF MEETING

*(Maison des Sciences de l'Homme, 54 boulevard Raspail, 75006 Paris)*

# REPORT

*January 22<sup>nd</sup> – 23<sup>rd</sup>, 2001*

## Delegates :

Centre National de la Recherche Scientifique, France : Gian Piero ZARRI;  
Maison des Sciences de l'Homme, France : Peter STOCKINGER;  
AXON, Portugal : Gilberto MACHADO, Fatima PIRES, Pedro ABREU;  
PIRA International, United Kingdom : Andrew BRASHER, Emma BROOMBY;  
Department of Computer Science of the University of Milan, Italy : Elena FERRARI.

## Monday, January 22<sup>nd</sup>, 2001

The meeting opens at 11am, in the Room 214, 2<sup>nd</sup> floor, of the Maison des Sciences de l'Homme (MSH). Professor Stockinger welcomes the delegates; the delegates introduce briefly themselves and their parent organisation.

Dr. Zarri sums up the main points of the project as they emerge from the contract that the Commission and the CNRS (on behalf of the consortium) have signed in the last week of December 2000.

EUFORBIA is the IAP (Internet Action Plan) project n. 26505. IAP, which derives from a European Parliament Resolution of April 1997, is now fully integrated in the Fifth Framework context. However, it has its own history and proper characteristics. Even if, e.g., one of its four IAP Action Lines concerns "filtering" — and we can find "filtering" in the Action Line III.4.1 of the IST 2001 Workplan — IAP is not IST, and EUFORBIA is not a pure technological (IST) project. This implies that the consortium must pay careful attention to the 'sociological' aspects of the problem we have to deal with (illicit use of Internet) and, in particular, must pay careful attention to a quick and accurate creation of the 'EUFORBIA user group', active during the whole life span of the project.

Passing now to the peculiarities of the EUFORBIA approach, this project must contribute to the production and use of new generations of Internet filtering systems that:

- can support a computer-effective description of the semantic content (the 'meaning') of Web documents that could be simultaneously i) very precise and complete in the description of the issues at stake in a given document; ii) neutral as much as possible with respect to any specific doctrine, ideology or value system;
- can provide the users — both the individual consumers or institutional users — with software tools able to make use directly of the neutral descriptions above to set up filtering policies and filtering schemata according to the most different cultural, political, religious etc. options.

A prototype, running software system must be realised by the consortium under the control of the EUFORBIA user group.

From a more technical point of view, the EUFORBIA strategy will be based on the systematic utilisation of two tools, NKRL and the 'Milan model'. The above objectives will then be realised according to the following principles:

- NKRL (and, in particular, the NKRL tools realised in the framework of the Esprit CONCERTO project) will be used to supply a neutral and exhaustive description of the meaning of Internet documents. 'EUFORBIA labels' under the form of NKRL annotation will then be associated with the chosen Web documents.
- The filtering tools will be based on two sort of 'generalised if-then rules' (similar to the Concerto inference rules) that will be utilised to process the EUFORBIA (NKRL) labels. A first class of rules will be used to calculate directly, automatically or semi-automatically, the rating categories and the associated values needed by the Milan model. The rules of the second class will be employed autonomously to filter the Internet documents.

The partners agree on the previous points: in particular, Professor Ferrari confirms that the UNIMI team agrees about the distribution of the rules into the two classes expounded above. The meeting is adjourned at 12.30 to allow the delegates to have lunch at the MSH canteen.

The meeting resumes at 2.00 pm. Several practical problems concerning the start of the project are discussed: a brief summary of the main points evoked and of the main decisions taken follows.

- CNRS will send to the partners a certified copy of the original EUFORBIA contract signed by CNRS and the Commission. At the same time, CNRS will ask the partners to specify the name and address of their bank in order to be able to transfer to them the advance payment. According to the contract, the Commission should send the money by the end of February at the latest.
- It is confirmed that, in an IAP framework, specific 'monthly reports' detailing the work accomplished each month by the partners for the different tasks of the project are not required. End of May/beginning of June, CNRS will ask the partners to produce a global document illustrating, for the first months of the project, the man/months they have devoted to each of these tasks. A template for producing this global document will be sent by CNRS to the partners in due time.
- All the documents and deliverables produced by the consortium will be prepared according to a word.doc or word.rtf format. A standard label to be reproduced on the first page of each document/deliverable, and a format for the page header, will be sent by CNRS to the partners. For simplicity's sake, only four sorts of documents (and of document identifiers) will be admitted: Admin., Note,

Tech.Report, Deliverable. 'Note' corresponds to the document that, such as the present 'Report', are not of a strict administrative nature and do not concern a very precise technical point. The five EUFORBIA teams will be differentiated in the identifier of the document ('Document type') by using their 'natural' labels: AXON, CNRS, MSH, PIRA, UNIMI. The documents will be progressively numbered according to i) the author team; ii) the type of document. For example, this document is labelled as CNRS/Note/1; the hypothetical technical document n. 27 produced jointly by AXON, MSH and UNIMI will be labelled as AXON/MSH/UNIMI/Tech.Report/27. Deliverable D1 etc. will be simply labelled as Deliverable/D1.

- A centralised mail server for the consortium will be installed at CNRS. MSH will develop the Web of the project: this will include a public and the private section. It is hoped that an on-line demo could be installed as soon as possible (end of 2001?) in the public part of the server.
- The delegates confirm the role their team has within the project. CNRS and AXON will develop the software that concerns mainly the NKRL aspects of the project. UNIMI will improve the Milan model; they will assure also the Java implementation of this model, in order to assure the compatibility with the NKRL software. A rough analysis shows that 2/3 of the EUFORBIA software will concern NKRL, and 1/3 the Milan model. As already stated, MSH will realise the Web of the project. PIRA will take in charge, mainly, the analysis of the user's (both content providers and content users) needs with respect to the 'friendliness' of the final software; they will also specify the 'external' interfaces of this software. They will host the 'EUFORBIA tools and resources' server, which will be a prototype of a commercial service. Pira will develop the business model for exploitation of the software and resources (e.g., ontologies etc.), and this 'Euforbias tools and resources' server will be the testing and development bed for these ideas. This server will be realised in WP5.
- CNRS will send to the partners the final version of the CONCERTO software as soon as a common declaration will be signed detailing all the contributions (software, general know-how, contacts etc.) the partners will put at the consortium's disposal. This document will be used to determine the shares of the partners with respect to the (possible) profits derived from the commercial exploitation of the results of the project.
- Windows NT, Java and XML are already assumed as standard tools for the project. CNRS and UNIMI will evaluate the possibility of using in the new project the tools developed in CONCERTO for translating NKRL expressions into RDF format. The consortium will also evaluate the possibility of making use of the software developed by UNIMI in CONCERTO for the set up and the management of a base of NKRL annotations (in this case, of EUFORBIA labels).

- Next meetings: i) in April (precise date to be confirmed), an EUFORBIA technical meeting will be held in Milan at UNIMI, to enable UNIMI to give a 'tutorial' about the Milan model: ii) in June/early July (date to be decided), an EUFORBIA consortium meeting will be held in Leatherhead at Pira International.
- The creation of the 'user group', and the possible contributions of all the partners in this context are evoked at length in the discussion.

The meeting is adjourned at 5h50 pm.

### Tuesday, January 23<sup>rd</sup>, 2001

The meeting resumes at 9h30 am, in the Room 7, 2<sup>nd</sup> lower ground floor, of the Maison des Sciences de l'Homme. All the teams are represented. The delegates debate informally about the technical and non-technical points raised the previous day.

In particular, PIRA emphasises again the need for really friendly interfaces that go beyond, in this domain, the somewhat unsatisfactory results obtained in a CONCERTO context. On a more technical ground, they advocate the use of a mixed strategy, both top-down and bottom-up, for the adaptation of the H\_CLASS and H\_TEMP ontologies developed in CONCERTO to the needs of the new project. Top-down means that most of the concepts proper to the new domain of application (such as sex, nudity, violence etc.) can be collected directly by examining the existing work (e.g., the ICRA's work) in the illicit Internet domain without accomplishing a long and fastidious in-depth examination of a corpus of specific documents. It is also possible that some 'ontologies' already exist in this domain. Bottom-up means that, given also the general approach of NKRL when contrasted, e.g., with the reduced number of ICRA categories, an at least partial examination of a corpus is nevertheless necessary to complete the top-down approach. The partners express a consensus about this position.

The meeting ends at noon of Tuesday.