

VICODI

Minutes of Kickoff-Meeting

Michael Jelden

October 24, 2002

Abstract

The intention of the kick-off meeting for the VICODI project, held in the offices of SYSTRAN in Soisy-sous-Montmorency on October 1, 2002, was to:

- Allow the partners to get acquainted
- Review the general guidelines for administrative and financial issues, proceedings and reporting
- Start mutual discussion on all issues of the project in the presence of all
- Give participants an idea of the interdisciplinary issues that will be encountered in VICODI, which combines the field of history with computer science

Most all of the important issues were discussed and all participants of VICODI agreed on the next steps that are needed so that this project can be successfully managed.

List of Participants

European Commission (EC):	Pierre-Paul Sondag
RIDEMO:	Edvins Snore
	Juris Zubkans
	Viesturs Sutko
Salzburg Research (SRFG):	Bob Mulrenin
	Dietmar Glachs
FZI Forschungszentrum Informatik an der Universität Karlsruhe:	Gábor Nagypál
University of Newcastle (UNEW):	Tim B. Kirk
University of East Anglia (UEA):	Richard Deswarte
National Library of Czech Republic (NKP):	Zdenek Uhlir
SYSTRAN:	Pierre-Yves Foucou
	Alex Tallas
	Michael Jelden

Presentation of Participants

The participants presented themselves briefly , and described their specific role within the VICODI project as follows:

Bob Murenin, Salzburg Research: Has experience with XML based history/cultural heritage databases (as in WP4 and 6). Will take part in ontology building, database setup, contextualisation tool, and contribute to the ontology editor and the annotation tool.

Edvins Snore, RIDEMO: Technical Coordinator , will deliver SVG engine and will coordinate all of the required technical work described in the workpackages.

Viesturs Sutko, RIDEMO: Managing Director

Juris Zubkans, RIDEMO: Technical expert

Richard Deswarte, the University of East Anglia: Professor of History . (Key) content provider, content identification (weblinks etc.).

Tim Kirk, the University of Newcastle: Professor of History . (Key) content provider.

Zdenke Uhlir, the Czech National Library, Prague: Historical researcher. Content provider.

Gábor Nagypál, FZI Forschungszentrum Informatik, Karlsruhe: Ontology experts. Will build ontology, MSKS server and contribute to the ontology editor and authoring tool.

Piere-Paul Sondag, the European Commission: Project officer of the EC, responsible for VICODI. He is contacted by SYSTRAN in case of questions.

Alex Tallas, for SYSTRAN: Director of Corporate Sales and responsible administrative project coordinator for EC projects.

Michael Jelden, for SYSTRAN: Vicodi project manager, computational linguist. Responsible for the Multilingual aspects of VICODI, machine translation and term substitution.

Pierre-Yves Foucou, Chief Technology Officer of SYSTRAN, (joined the meeting after the lunch-break to get a first-hand idea of the scope of the project)

Introductory Presentation

Pierre-Paul Sondag gave an introductory presentation on the principles of IST (Information Society) project management and on the duties of each participant of EC funded projects in particular.

Topics covered were:

- Funding: Only costs particular and necessary to the project can be made eligible. Costs must occur during the project.

- Discounts (that the participant may receive from third parties) will be treated according to historical costs in company accounting. No marketing/distribution costs will be reimbursed.

- EC's standard financial reporting forms must be used.

- Time spent on the project by staff of the respective companies has to be proven by the company.

- Names, function, category of personnel and worked hours have to be indicated and reported.

- Only depreciation of durable equipment (i. e. computers) can be reimbursed.

- Subcontracting should not exceed 20% of total amount of work.
 - Research tasks contracted out are direct costs, funding can be claimed under the cost category "subcontract"

 - Administrative tasks contracted out are indirect costs, no funding can be claimed. These costs are structural costs as are buildings, heating, stationery, telecommunications etc.

- Overheads rates are calculated by dividing these structural costs by the number of researchers using the structure. Hence administrative tasks contracted out can enter in the overheads (rates calculation)
An exception is the admin. coordinator who can claim administrative tasks contracted out as direct costs.

- Miscellaneous costs: Prior agreement is necessary in case of any costs not included in Annex 1 of the contract.

- Costs statements have to be submitted every 6 (six) months. They have to be submitted to SYSTRAN in hardcopy form properly signed and stamped by the

authorized persons of each organization, who will forward them to the EC. Delays in submission of more than two months entitles the EC to refuse immediate payment, and payment of the amount can be differed for the next period.

- Travel costs statements must include details of the trip.
- Non Euro-zone partners have to use the contractual exchange rate as published on the ECs webpage.
(europa.eu.int/comm/budget/inforeuro/fr/index.htm)
- Up-to-date forms have to be used
(as found on www.cordis.lu/ist/cpfclaim.htm)
- All details on costs statements requested by the EC must be provided by the respective contractors.
- Before the end of the project not more than 85% of the entire eligible costs will be paid by the EC to the parties. The remaining 15% will be paid after the project is finished.
- The administrative coordinator has to inform the EC on payments made to the participants.
- EC authorization is needed in case of transfer of budget between cost categories of the same contractor. Prior authorization, through an addendum to the contract, is requested if the cumulated transferred amounts exceed 20%.

Authorization from the EC has to be requested if the amount exceeds 20%.
The same is valid in case of transfer of costs between contractors.
- Projects are assessed once a year by external experts, in Luxembourg.
- Financial audits of the participants of VICODI are possible up to 5 years after the end of the contract, and also during the project.
- In case of further exploitation of project results (patents, commercial exploitation etc.), a mutual agreement between consortium members should be made. It should be made as early as possible, during the project. .
- Reporting is requested as follows:
 - General project management**: One report each 3 months
 - Project report (by all participants)**: Each 6 months
 - Technology implementation plan, by all**: End of project
 - Final report**, : End of project
- A project website has to be set up. .
- ``Concertation days``: Held in Luxembourg, must be attended by all members of (different, not only VICODI) EC funded projects
(www.cordis.lu/ist/cpt/2000clust.htm).

VICODI Technical Issues

Edvins Snore (RIDEMO) made a presentation based upon a demo web-page of VICODI.

The following points were discussed in the break-out sessions and were subject to further discussion:

- How can the relations between topics, author of article, (geographic) origin be established and what will its graphic representation look like?
- Does the ontology have to be translated by humans or is automatic translation feasible, i. e. is the ontology static or dynamic?
- How can individual permissions for the administrator (main editor), subordinated contributors/editors (possibly volunteers) be set up in order to enable them to annotate/categorize documents, but still keep control of their work?
- How can scope and limits of historic contents be defined?
- How and from which sources should the thesauri used in VICODI be build?
- What is a context of a document? The context of the historical material, the context of the document author or other? How many LATCH contexts are allowed? Only one, or more? Is there context only for the whole document, or also for document parts, paragraphs?
- Ontology building: The ontology should not contain ' ' real-world' ' items, but only concepts. ' ' Real-world' ' things will then be attributed to the ontology concepts by linking them in the VICODI database. This database will then be enriched by the contextualisation engine, i. e. the contextualisation engine attributes concepts from the ontology to the articles.
- How would the final architecture of VICODI be designed? The technical workflow from article to search result should be designed around a central database, containing the historical items and, attributed to that field, tags from the ontology to mark the context. All other elements such as SVG engine for the graphical representation, the machine translation, contextualization engine and annotation tool are all fed by or feeding that central element.

A sketch of the general architecture, as shown by Edvons Snore in his presentation is attached to that report.

These issues are to be resolved until the final technical specification can be submitted (project month 6, i. e. by the end of March 2003)

Presentation about Ontologies

Gábor Nagypál (FZI) gave an introduction on ontologies in general, stressing the differences between thesauri and ontologies.

It was agreed upon that ontologies made out of abstract concepts, that are arranged hierarchically and related to each other. "Real world items", i. e. in the case of VICODI historical content, are then attributed to one or more instances of these abstract concepts in order to enable the search engine find related items etc. As a consequence of this, the ontology is a relatively static structure, whereas the items categorized by elements from the ontology can be dynamic content.

A search for related items would search for similarly tagged historical items in the relative database field and return them.

Multilingual Issues, Machine Translation

The ontology will be automatically translated by SYSTRAN and post-edited by humans at the editor's site.

The page with the search results will provide a link next to every document that matched the query. This link leads to another html page on which the user can choose to translate that document into one of VICODI's languages (into/from English, French, German, and English-Latvian term substitution).

Open Issues

- What will the search engine be like? Will we use a real search algorithm (who' d develop it? Does someone have one?), or simply search for the words contained in the search phrase and enrich/filter these results statistically or by semantic analysis?
- Who clarifies copyright questions regarding linked documents from the web (at runtime of the project results)? SYSTRAN can answer the general questions regarding copyright, but if varying URLs are dynamically linked from the internet this issue must be clarified by the editors case by case.
- No copyright has to be clarified if documents are not cached locally. So maybe documents from the internet should not be cached locally at all? Store the link and attribute it to ontological tags would be enough. Spiders can analyze the remote documents.
- A VICODI project webpage has to be set up. This should be done by the technical coordinator.

These issues are a prerequisite for further action and need due attention by all participants.

Conclusions and Action Points

As a conclusion of the general discussions and the break-out sessions the following was mutually agreed upon, by all participants, as next steps to drive VICODI forward:

- 1.The source(s) of the historical documents has to be defined: Proprietary documents, remote URLs, external thesauri, documents with third-party copyright holders, university resources...?) (to be done by historians, SYSTRAN can clarify general copyright issues)
- 2.Ontology: Has to be developed by UNEW and UEA in collaboration with FZI and SR, build with the contents of the corpus in mind (by FZI, UEA, UNEW, SR, and RIDEMO): The universities will provide FZI and RIDEMO with a small collection of sample documents (20-50 documents, on varying topics). FZI in collaboration with RIDEMO will develop a draft ontology for the historians to work on, i. e. to give them something to start with.
- 3.Building of Contextualisation Tool/Engine (CE): The CE tries to

automatically guess the domain of a document from the ontology, and returns also the precision/probability of results (to lead into the following point) (by SR, FZI)

4.Annotation tool: If the CE returns ambiguous/imprecise results, a human editor will edit the context/ontology/domain of the article in the Annotation Tool (by SR, FZI)

5.The contents of the corpus, built in action point No. 1 has to be tagged according to ontology. This work has to be done by the contextualisation engine and the annotation tool in collaboration with and under supervision by the historians.

6. SYSTRAN will provide partners with a model Consortium Agreement for them to review and comment on.