



## MIXED Consultation Workshop

September 10, 11, 12 2009

### MEMO 03 September 4th 2009

#### Contents:

1. Themes for the MIXED workshop
2. The workshop experts
3. Directions to the workshop hotel
4. Reimbursement of the costs

#### 1. Themes for the MIXED workshop

The previous memo explained what MIXED is, where the project stands at the moment, and what we expect from it. In this memo I will put the question as to the future of MIXED in a perspective, a perspective that is ... your perspective. Hopefully.

The playground of MIXED has multiple facets, and, looking at current activities all over Europe and way beyond, we have identified these four facets: (i) digital preservation, (ii) research archives, (iii) research infrastructures, (iv) open source software communities. Here are some questions concerning MIXED and these facets, on which we hope you can shine your light during the workshop, individually and collectively.

##### (i) digital preservation

MIXED represents a variant of the migration strategy for digital preservation. We call it "smart migration" because the method breaks loose from the wheel of eternal upgrades of vendor formats. Instead it converts material into an 'enlightened' format, where it may rest for ever. When the material is used, a dissemination copy of it may easily be converted, on the fly, to a current vendor format. Question that arise are:

- how do you evaluate "smart migration", in the presence of plain migration and emulation?
- migration to the XML data format works best for those aspects of content that can be expressed in XML in standard ways. Presentation and action are aspects for which few canonical formalisms exist. What impact does this have on the usability of smart migration?

##### (ii) research archives

A digital preservation measure such as smart migration must be implemented in repository context, i.e. an archive or repository with digital research data. Questions here are:

- what is the best interface of MIXED in view of being used in a repository workflow?
- where in the archival process (ingest, administration, management, dissemination) would you want to use the software?
- what is needed to trust the quality of the MIXED conversions?
- is MIXED a contribution to being/becoming a Trusted Digital Repository in the sense of TRAC, DRAMBORA or the Data Seal of Approval?

##### (iii) research infrastructures

The task of digital preservation is too daunting to be tackled by individual repositories. There are tasks that do not even make sense at that level, such as standardization of data formats and metadata. In order to exploit the benefits of digital data for scholarly purposes, we need research

infrastructures. These are not only about physical data connections, but much more about mutual understandings about how to refer to data, how to shape metadata, how to deal with rights. The vision is that research data in individual repositories will become transparently referable, accessible, usable and reliable for any researcher with access to the infrastructure as a whole.

Questions for MIXED are:

- what is needed to deploy MIXED not in one repository, but on a network of repositories?
- how can MIXED collect the best practices in the area of file format conversions, so that networks of repository can employ these practices?
- how should MIXED be governed? An international board? A foundation with contributing members?

#### (iv) open source software communities

The MIXED software will be published as open source software. That will guarantee the legal usability of the software for whoever needs it. But we want more: contributions of new conversions, improvements of existing conversions, better interfaces and integrations with repository systems, such as FEDORA. The main questions are:

- how do we foster an open source MIXED community?
- what platform do we choose? SourceForge, Apache?
- who are likely to contribute, and in what way? Individuals? Repositories with budget for digital preservation measures? Software vendors? Government ministries?
- will this help to keep the MIXED software sustainable over time?

These are questions. We hope that in the process of pondering them, some scenarios will crop up by which we can put MIXED firmly on the road.

## **2. The workshop experts**

The goal of the MIXED consultation workshop is to come up with scenarios for a sound future of the MIXED framework. The organisers would like to have input from a number of fields of expertise: digital preservation, scientific data archives, research infrastructures and open source software development. In the last weeks we had some changes in the list of participants. Below you will find the most recent overview.

Which hat do you wear?

- Alison Heatherington, services manager for digital preservation at the National Archives (UK).
- Amir Bernstein, researcher responsible for the archiving of relational databases. Project manager and product owner of SIARD. Swiss Federal Archives.
- Barbara Sierman, manager digital preservation at the KB (national library of the Netherlands). Involved in EU projects Planets and Driver.
- Dirk Roorda, information expert. Leader of the MIXED project.
- Ellen Kraffmiller, senior software engineer and project manager at Harvard University. Involved in the Dataverse network project.
- Henk Harmsen, deputy director DANS.
- Jan van Mansum, developer in the MIXED project (DANS).
- Jeroen Rombouts, projectleader 3TU Datacentre, Delft University of technology.
- John Doove, project coordinator at SURFfoundation, initiator of innovation in higher education and research.
- Marc Kemps-Snijders, coordinator user applications at Max Planck Institute for Psycholinguistics, Nijmegen.
- Nathan Adams, assistant director of software development at University of Michigan, ICPSR.
- Peter Doorn, director of DANS.
- Rainer Schimdt, workpackage leader and member of the technical coordination committee in the Planets project. AIT Austrian Institute of Technology.

- René van Horik, program manager at DANS. Involved in the dissemination workpackage of the MIXED project.
- Rob Grim, research data specialist. Tilburg University. Member management team Open Data Foundation.
- Rutger Kramer, coordinator software development DANS.
- Sebastian Rahtz, information manager at University of Oxford computing services. Expert in technical aspects of the Text Encoding Initiative.
- Steven Krauwer, researcher and project manager in language and speech technology. Coordinator of CLARIN (Common Language Resources and Technology Infrastructure)
- Vladislav Makarenko, scientific programmer at Max Planck Digital Library (developer on eSciDoc project).

### 3. Directions to the workshop hotel

The hotel reservation for the participants is done by the workshop organisation. This is for Thursday and Friday night (September 10<sup>th</sup> and 11<sup>th</sup>) and for some participants (depending on the travel scheme) also for Wednesday September 9<sup>th</sup>. The MIXED consultation workshop will be held at the Carlton Beach Hotel at Scheveningen, near Den Haag (The Hague). The address of the hotel is:

Carlton Beach Hotel  
 Gevers Deynootweg 201  
 2586 HZ Scheveningen/The Hague  
 Tel. +31 (0)70 - 354 14 14  
 Fax +31 (0)70 - 352 00 20  
 info@beach.carlton.nl

The hotel is about 1 hour travel by public transport from Schiphol airport. Take the train from Schiphol airport to either Den Haag CS (Centraal) or Den Haag HS (Hollands Spoor). From The Hague 'Hollands Spoor' train station, take tram #1 or #9 and stay on until the last stop "Zwarte Pad". From The Hague Central Station take tram #9 and stay on until the last stop "Zwarte Pad".

Websites:

Carlton Beach Hotel: <<http://www.carlton.nl/beach/default-en.html>>

Journey planner for all Dutch public transport: <<http://journeyplanner.9292.nl/>>

NS: Dutch Railways: <<http://www.ns.nl/cs/Satellite/travellers>>

Public Transport in The Hague (tram): <<http://www.htm.nl/Pages/DEF/245.html>>

### 4. Reimbursements of the costs

The costs for travel, accommodation and meals are covered by the workshop. The hotel reservation is made by the organisation and directly paid by the organisation. For the reimbursement of the other costs we would like to ask you to fill in a spreadsheet form, accompanied by (copies) of tickets etc. The spreadsheet is emailed to the participants. The printed signed form has to be sent to:

DANS  
 PO Box 93067  
 Den Haag  
 The Netherlands

For more information on the MIXED project, contact

[Dirk.roorda@dans.knaw.nl](mailto:Dirk.roorda@dans.knaw.nl)

For information concerning the consultation workshop, contact

[Rene.van.horik@dans.knaw.nl](mailto:Rene.van.horik@dans.knaw.nl), tel. +31-6-23297389