



1. Activity Area: Preservation Working Group

Leader(s):	Clément Oury, BnF & Gina Jones, LC		
Current status	A		
Previous status	G		

1.1 Individual project or activity:

Name:	Overall working group management
Start date:	2008
End date:	NA
Summary of objectives:	Ensure that the goals of the PWG are shared by all members; ensure that the work packages are well conducted; ensure that information is well distributed between members

1.1.1 Summary of active tasks and milestones due in reporting period

Active tasks and / or Milestones	Date of completion	RA G last period	RA G this period	Commentary / Notes, such as progress made and explaining changes in plan or dates
Scheduled phone meetings	19 Nov 2010	G	G	Scheduled phone calls with Preservation Working Group members.
Release PWG Notes	25 Jan 2011	G	G	Final notes to be sent electronically

1.1.2 Issues, findings and actions arising

Of note, BnF submitted WARC to Pronom. On November 5, 2010, WARC was added to the Pronom DB and has the PUID fmt/289:

<http://www.nationalarchives.gov.uk/PRONOM/fmt/289>

1.1.3 Summary of active tasks and milestones planned for the next period

Active tasks and / or Milestones	Planned date of completion	Commentary / Notes
Phonecall / all	Feb 2011	To demonstrate Risks DB for testing by members.

1.2 Individual project or activity:

Name:	Risk assessment (WP8)
Start date:	2009



End date:	2011
Summary of objectives:	Provide an update of the assessed risks to web archives Create a foundation for performing periodic risk assessments for web archives

1.2.1 Summary of active tasks and milestones due in reporting period

WP 8 Risk Assessment	11/30/10	G	G	Risks section tested and launched in DB for review.
	11/30/10	G	G	Bibliography implemented in DB, phase 1.
	12/16/10	G	G	Meeting on risks and bibliography section.
	12/27/10	G	G	Institutional risk assessment section tested and launched for review.

1.2.2 Issues, findings and actions arising

None

1.2.3 Summary of active tasks and milestones planned for the next period

Risk assessment (WP8)	01/15/10	Finalize and launch risk controls section of db.		
	02/1/11	Finalize initial implementation of the bibliography data input section.		
	02/25/11 – May 2011	Preservation working group members perform institutional risk analysis on risks identified in DB.		

1.3 Individual project or activity:

Name:	Viruses (WP6)
Start date:	2009
End date:	2011
Summary of objectives:	Assess the risks of keeping viruses in web archives Provide scenarios to identify and discard viruses. If recognized necessary, set up a project to encourage one or several AV tools to manage WARC files.

1.3.1 Summary of active tasks and milestones due in reporting period

Viruses (WP6)	10/15/10	G	G	First draft of viruses taxonomy.
Viruses		G	A	Review of list of virus types.



(WP6)				
-------	--	--	--	--

1.3.2 Issues, findings and actions arising

None

1.3.3 Summary of active tasks and milestones planned for the next period

Viruses (WP6)	03/01/10	Review of list of virus types.
---------------	----------	--------------------------------

1.4 Individual project or activity:

Name:	Information packages (WP7)
Start date:	2009
End date:	2011
Summary of objectives:	analyze information packages designed by different institution to ingest their web archives into repositories share documentation on this topic (e.g. METS profiles) produce a best practice report

1.4.1 Summary of active tasks and milestones due in reporting period

Information packages (WP7)	09/25/10	G	G	A general chart to analyze information packages has been produced and reviewed by the members.
Information packages (WP7)	12/01/10	G	G	A draft questionnaire has been sent to members for review.
Information packages (WP7)		G	A	Each concerned institution should fill the chart and share their information packages.

1.4.2 Issues, findings and actions arising

None

1.4.3 Summary of active tasks and milestones planned for the next period

Information packages (WP7)	02/15/10	Each concerned institution should fill the chart and share their information packages.
----------------------------	----------	--

1.5 Individual project or activity:

Name:	Environment database (WP3)
Start date:	2009
End date:	Release of the DB: 2011



Summary of objectives:	Set up a database where the main components of the web technical environment (browsers, plug-ins, formats, crawlers...) would be described over time
------------------------	--

1.5.1 Summary of active tasks and milestones due in reporting period

Environment database (WP3)	11/8/10	G	G	Teleconference with LC and Ross Spencer, Pronom on refining and improving the data model and its relationship to Pronom.
Environment database (WP3)	12/31/10	G	G	Draft implementation of the software module
Environment database (WP3)		G	A	Launch the Software module of the DB for testing
Environment database (WP3)		G	A	Start describing environments.

1.5.2 Issues, findings and actions arising

Work for the database during this quarter has been primarily focused on the WP 8 (risk assessment and bibliographic section) of the PWG DB. A significant in-depth study of the software and format part of the data model has been conducted. The Software module of the DB and the Environments Work Package has been moved to amber because of a significant change to the data model, migrating to embrace the Pronom model. This will ensure compatibility of data exchange between the PWG DB and Pronom and hopefully GDFR/UDFR.

1.5.3 Summary of active tasks and milestones planned for the next period

Environment database (WP3)	3/15/2011	Launch the Software module of the DB for testing
Environment database (WP3)	3/15/2011	Start describing environments.
Environment database (WP3)	4/15/2011	Produce a statement explaining the scope of the goals of this DB, communicate on it

1.6 Individual project or activity:

Name:	Format identification for web archives (WP1)
Start date:	2009
End date:	
Summary of	There are two complementary projects to develop Jhove 2



objectives:	modules for web archives. The first one (funded by BnF) will notably provide ARC and GZIP modules, parallelism management (to increase performance) and integration of the “file” identification tool. The second one (proposed by CDL, to be voted by SC) is intended to provide modules for WARC, HTML, JPEG, GIF and PNG, and to integrate virus analyses.
-------------	---

1.6.1 Summary of active tasks and milestones due in reporting period

Jhove2 (link with WP1)	Sept 2010	G	G	The web archive file formats” WP (WP1) has identified the major formats that are considered as a priority. The Jhove2 development projects have been designed according to the results of the WP1 study.
Jhove2 (link with WP1)		G	A	Promote the deliverance to the SC of a funding proposal for additional format modules for JHOVE2.

1.6.2 Issues, findings and actions arising

None.

1.6.3 Summary of active tasks and milestones planned for the next period

Active tasks and / or Milestones	Planned date of completion	Commentary / Notes
JHOVE2 Proposal to SC	12/30/10	Promote the deliverance to the SC of a funding proposal for additional format modules for JHOVE2. Work of PWG members on specifications and tests will start afterwards, in case of a positive vote.

1.7 Individual project or activity:

Name:	Plato for web archives (link with WP2)
Start date:	2009
End date:	2011
Summary of objectives:	Test Plato (preservation planning tool) for web archives. Focus for example on a web-specific format (such as Flash video)

1.7.1 Summary of active tasks and milestones due in reporting period

Project decided in Vienna



1.7.2 Issues, findings and actions arising

None

1.7.3 Summary of active tasks and milestones planned for the next period

Plato for web archives (link with WP2)	02/31/10	Write the project description.
---	----------	--------------------------------