

Testbed – a walk-through

Digital Preservation Planning: Principles, Examples and the Future with Planets, July 2008

Matthew Barr

HATII at the University of Glasgow



Humanities
Advanced Technology
& Information Institute



Contents

- Definitions and goals
- Achievements so far
- **Walkthrough**
- Testbed architecture and core technologies
- Future work



Definition

“A controlled environment for experimentation and evaluation, with metrics and benchmark content that allow comparison of tools and strategies”



Why do we need a Testbed for Digital Preservation?

- Perform scientific research in digital preservation
 - Evaluate preservation approaches in diverse “real life” settings
 - Avoid duplication of work
 - Maximise use of invested resources
- Need for a **dedicated research environment**
- Systematic execution of experiments by different institutions with
 - automated evaluation
 - reproducibility
 - documentation and accessibility



Contribution of the Testbed to Planets

- To preserve digital objects, we need:
 - Preservation tools:
 - Characterisation tools
 - Preservation action tools (migration / emulation)
 - A preservation plan
- The Planets Testbed will give us:
 - Information on the usability of preservation tools and services in various conditions / on various types of data, e.g.:
 - Degree of perservation of certain characteristics of data
 - Speed of tool
 - Institution independent results (focusing on tools and data only)
 - (Aggregated) information will feed into the Planets Tools Registry
 - Information can be used in a preservation plan



The Planets Testbed

- Foundation for objective evaluation of preservation tools and strategies:
 - Load content
 - Perform experiments: select data and tools, run the experiment, evaluate results, compare outcomes
 - Provide support for validation of preservation strategies
 - Benchmark tools and services
- Consists of:
 - Data storage, hardware, Planets software, Testbed application
 - Benchmark content



Achievements so far

- First Testbed prototype release in September 2007
- Planets-wide first release of Testbed in March 2008
- Release to be used by Planets partners for digital preservation experiments in July 2008
- Planning for corpora established
- Planning for participation of external institutions established



First Prototype (Sept. 2007)

- Mainly intended for internal feedback on user interface
- No integration with Planets Interoperability Framework (e.g. single sign-on, user management, data registry)
- Main focus on the experiment design and the experimenter user type
- Execution still a mock-up



First Release (March 2008)

- Experiments can be fully executed
- Custom versions of certain Interoperability Framework elements, such as the Service and Data Registries
- Integration of the IF single sign-on service
- Limited number of Services available
- Services can be registered through the Service Registration Wizard
- Arbitrary migration services can be registered and used in an experiment



New in current release (V.0.6, July 2008)

- Experiment approval and scheduling system
- Expanded experiment search functionality
- Data browser for browsing and selecting experiment data
- Planets FTP area for supplying sample experiment data
- WebDAV interface for uploading and managing data available
- Significant increase in the number of available services
- Various updates to the GUI

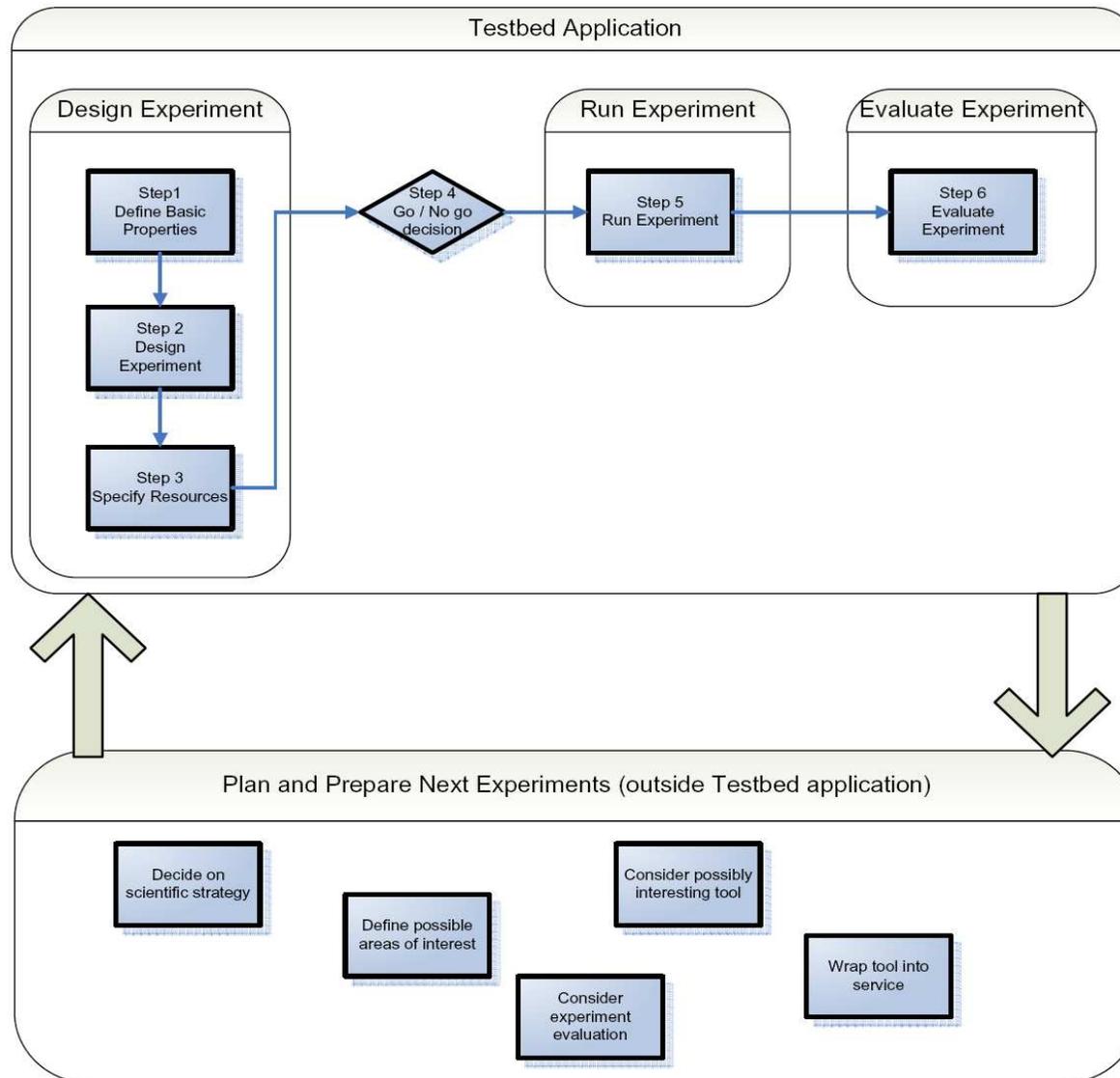


Services currently available

- PNG to TIFF Converter
- JPG To TIFF Converter
- JPG to PNG Converter
- TIFF to PNG Converter
- Xena DOC to ODF and ODF to PDF
- PS to PDF Basic Migration
- OpenXML Migration
- Simple Characterisation
- Image Identification Service



The Testbed Experiment Process



Experiment methodology

- Built on work undertaken by the Dutch Digital Preservation Testbed and DELOS Testbed, but
 - Integration with Planets Interoperability Framework
 - Focus on workflow design and on automation
- Experiment Process: based on that of the Dutch and DELOS testbeds, but
 - different end-user typology → simplification (from 12 steps to 6)



The Testbed Central Instance and the Testbed software

- The Testbed Central Instance is hosted by HATII at the University of Glasgow
 - Available to all Planets partners at <http://testbed.hatii.arts.gla.ac.uk:8080/testbed/>
 - Experimenters are encouraged to use this central instance to ensure the seamless aggregation of experiment results
- A downloadable instance of the Testbed software is available for local installation at <http://gforge.planets-project.eu/gf/project/ptb>



Testbed Architecture: Key Principles

1. Designed to be platform independent, robust and scalable
 - Java Enterprise Edition
2. Designed to execute experiments on a wide array of preservation tools and services
 - Web Service approach: tools wrapped as web services can be accessed by the Testbed application by means of a platform-independent URI
3. Designed to be a part of the overall Planets software suite
 - Sharing of common components across the entire project
 - Testbed development can focus primarily on the components that are unique to the TB



Web Service Approach

- All preservation tools required for Testbed experiments are deployed and accessed as Web Services
- All tools must be wrapped as WSI-compliant web services so that
 - Services can be registered with the testbed
 - Service templates can be created
 - Experimenters can then access these templates to simulate the specific usage of a tool
 - The Testbed is **extensible**. Any WSI tool can be made available
- Steps involved in registering and configuring a service are handled by the Testbed administrator



Next steps

- Next releases
- Planning for corpora
- Planning of experiments
- Opening to external institutions



Planning for Corpora

- Digital preservation corpus: annotated collection of digital objects
 - Annotations should contain the criteria against which given algorithms will be evaluated
- By October 2008 a detailed plan for the building of the corpora will be in place
- Corpora will ensure that a sufficient knowledge base is available for each experiment
 - will avoid the duplication of effort for experimenters



Planning for Experiments

Next steps (summer 2008):

1. Decide on good evaluation criteria
 - Make use of PLATO case studies
2. Decide on test strategy
 - (in depth testing of small sets / broad testing / tools for most common object types / tests on as many different object types as possible / etc.)
3. Set up a test schema
 - Get partners involved & monitor progress
4. Evaluate and continue
 - Use evaluation information to reconsider steps and continue
 - Broaden the automated evaluation concepts



Opening to external institutions

- Current areas of work:
 - Work out Testbed benefits for external organisations
 - Define criteria for filtering organisations
 - Establish key contacts in the organisations
 - Devise concrete testing tasks the organisations to be involved in
 - Assign timescales and resources for testing
- Agreement with external institutions in place by November 2008
- Opening of Testbed for external experiments by April 2009



Testbed – a walk-through

Thanks!

Matthew Barr

HATII at the University of Glasgow



Humanities
Advanced Technology
& Information Institute

