The Planets Digital Preservation Project

Birte Christensen-Dalsgaard

Planets

- Four-year EU-Funded (FP6) Digital Preservation research and technology development project.
- Increase Europe's ability to ensure long-term access to its cultural and scientific heritage
  - Improve decision-making about long term preservation
  - Ensure long-term access to valued digital content
  - Control the costs of preservation actions through increased automation, scaleable infrastructure
  - Ensure wide adoption across the user community and establish market place for preservation services and tools
- Build practical solutions
  - Integrate existing expertise, designs and tools
  - Share and build

Planets

- Brings together Archives, Libraries, researchers and technology companies
  - Builds on strong digital archiving and preservation programmes
  - Addresses core challenges
  - Focuses on needs of Libraries and Archives
- Will provide an interoperable framework to enable
  - Third-parties to provide tools and services
  - Vendors to integrate preservation services
  - Content owners to ensure long-term access to their digital content
- Will use an empirical approach to gather evidence
- Outreach shows potential to make a difference

Planets partners I

- The British Library
  National Library,
  Netherlands
  Austrian National Library
  State and University Library,
  Denmark
  Royal Library, Denmark

- National Archives, UK
  Swiss Federal Archives
  National Archives,
  Netherlands

Planets partners II

- Tessella Plc
- IBM Netherlands
- Microsoft Research, Cambridge
- Austrian Research Centers
- Hatii at University of Glasgow
- University of Freiburg
- Technical University of Vienna
- University at Cologne
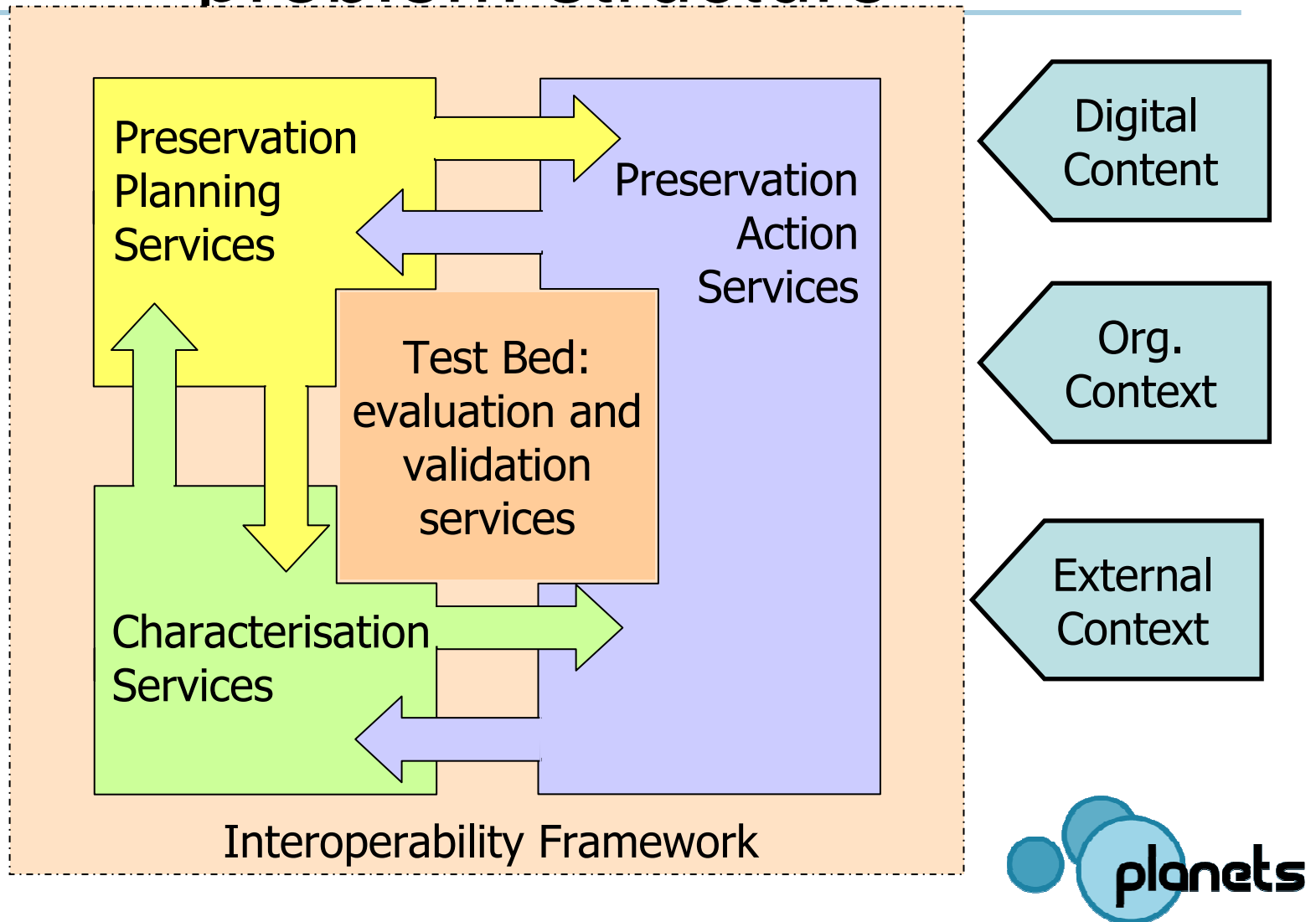
The Team

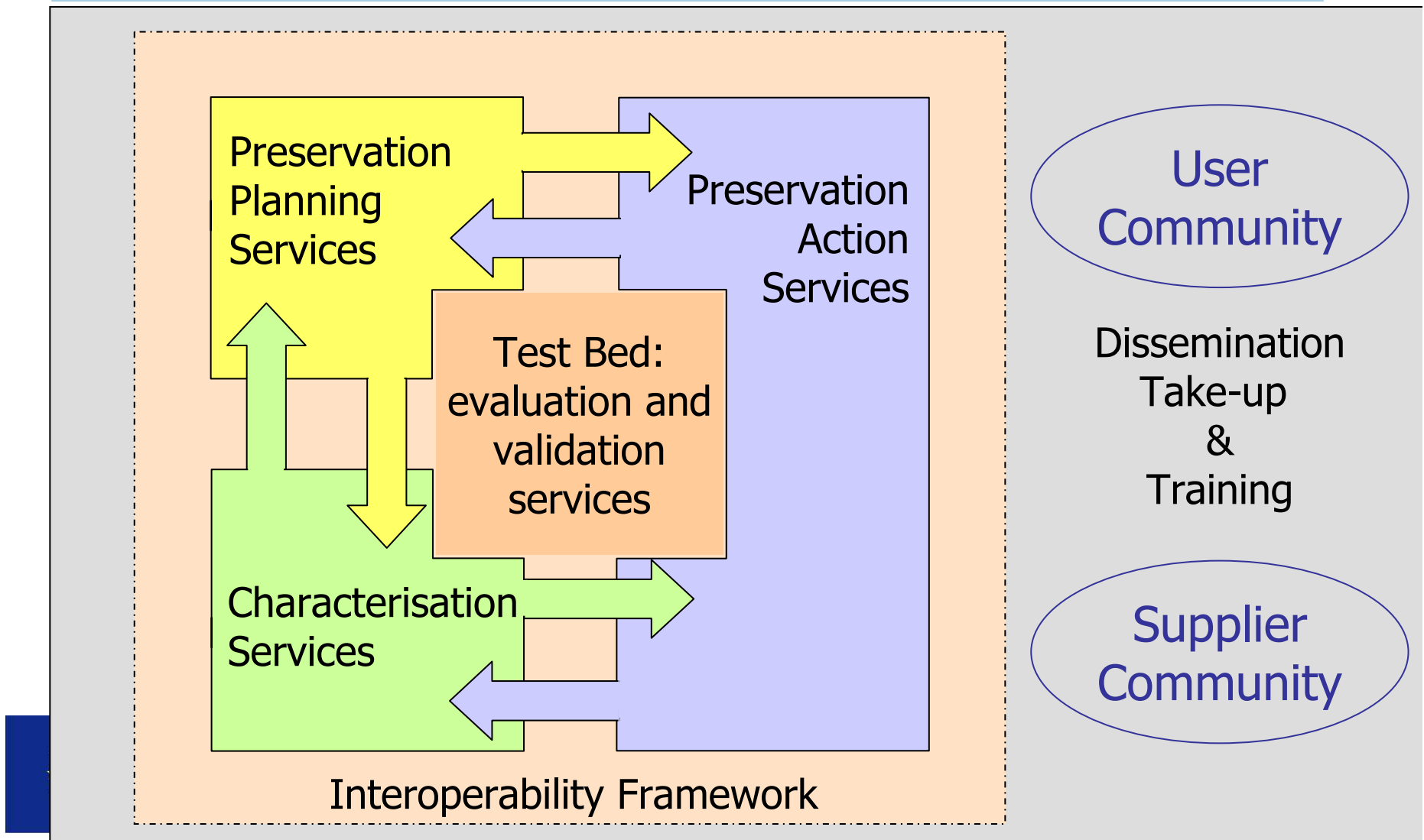**All Staff Meeting, Feb 2007**

Planets approach

- **Planning services** that empower organisations to define, evaluate, and execute preservation plans
- Methodologies, tools and **services for Characterisation** of digital objects
- Innovative solutions for **Preservation Actions**
- An **Interoperability Framework** provides services distributed services
- A **Testbed** enables objective evaluation of protocols, tools, services and plans
- **Outreach**, workshops and training to engage the user and vendor communities

Project architecture reflects problem structure

Project Architecture Reflects Problem Structure

Preservation Planning Services

Preservation Action Services

Test Bed: evaluation and validation services

Characterisation Services

Interoperability Framework

User Community

Dissemination Take-up & Training

Supplier Community

Interoperability Framework

The glue that holds the Planets tools and services together

- Provide service registries
  - Characterisation services
  - Preservation action services
- Provide shared services
  - Security, authentication, authorisation
  - Monitoring, logging, auditing
  - Intermediate data, repository, file system space
  - Execute and manage workflows
- Enable third-parties to provide tools and services
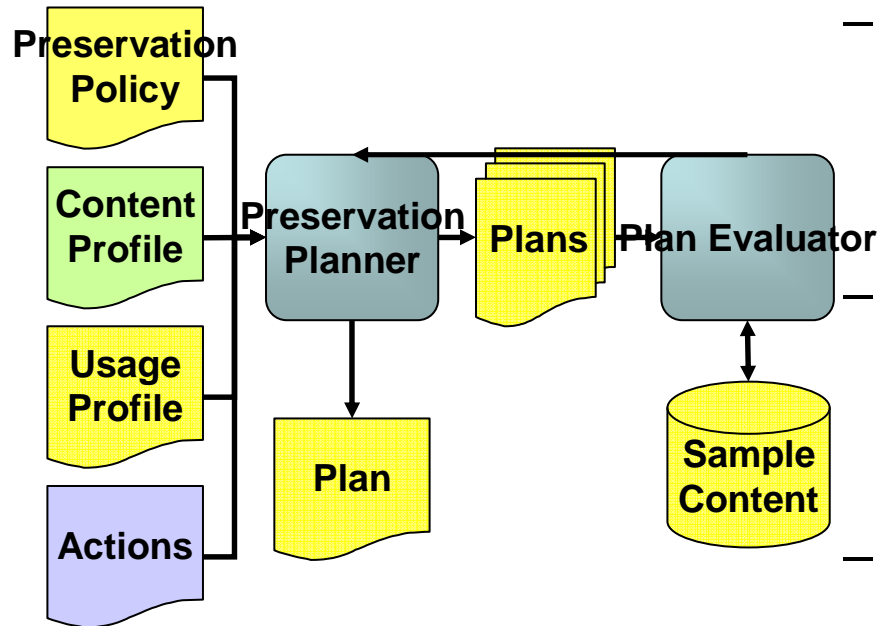- Enable vendors to integrate preservation services

Interoperability Framework
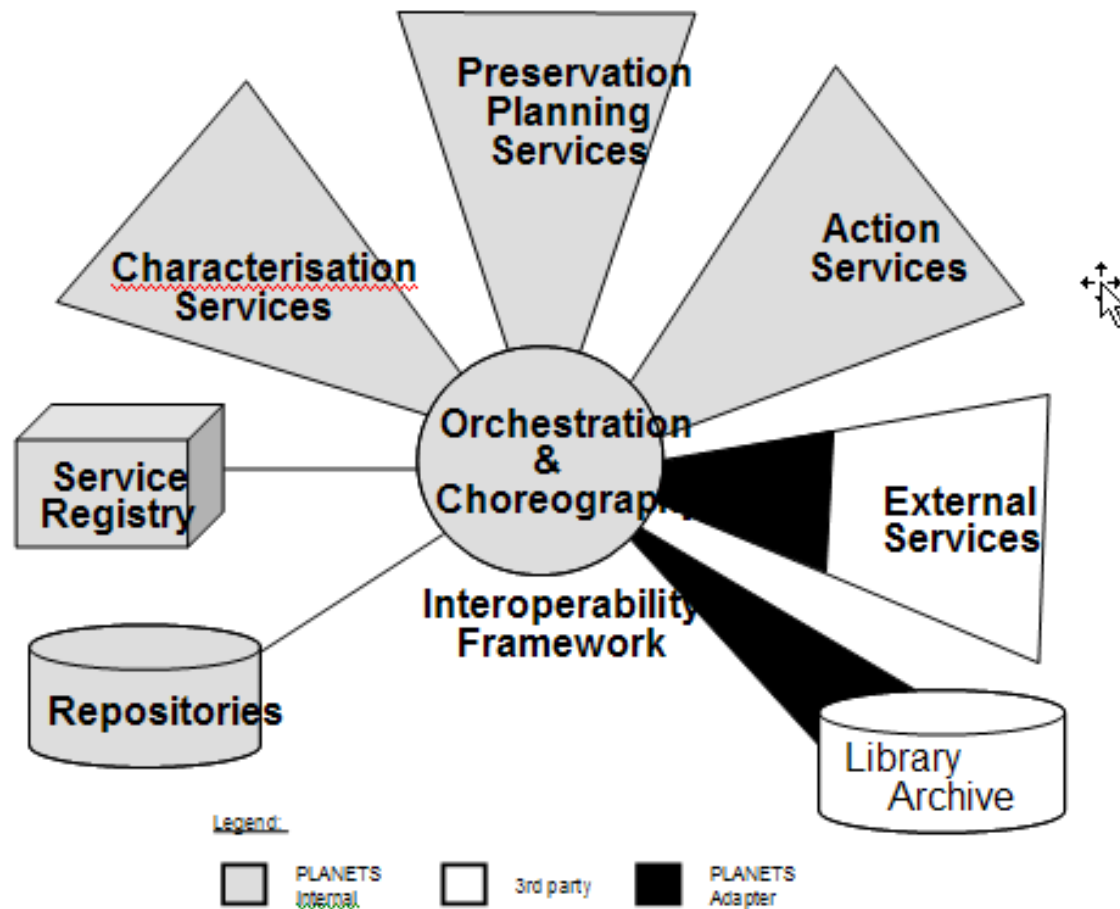
- Interoperable distributed services
- Service registries and shared data-stores
- Encapsulate tools as services
- Orchestration capability to combine services

Preservation Planning

- Input:
  - Preservation policy
  - Collection and community profile
- Feedback
  - Plans can be executed on sample content and evaluated
- Execution
  - Plans can impact a repository, ingest workflow, delivery workflow
- Validation
  - Services will be evaluated in real organisational contexts

Content characterisation

- **Characterise content to support preservation**
  - Reduce up-front metadata costs
  - E.g., Harvard segmented images based on tool parameters
- **Build on TNAs PRONOM for file-format identification**
  - Define a characterisation language (XCDL)
  - Define an extraction language (XCEL)
  - Define an pluggable interpreter
- **Extend to measure loss due to actions**
  - All transformations cause loss
- **Leverage understanding to improve file formats**
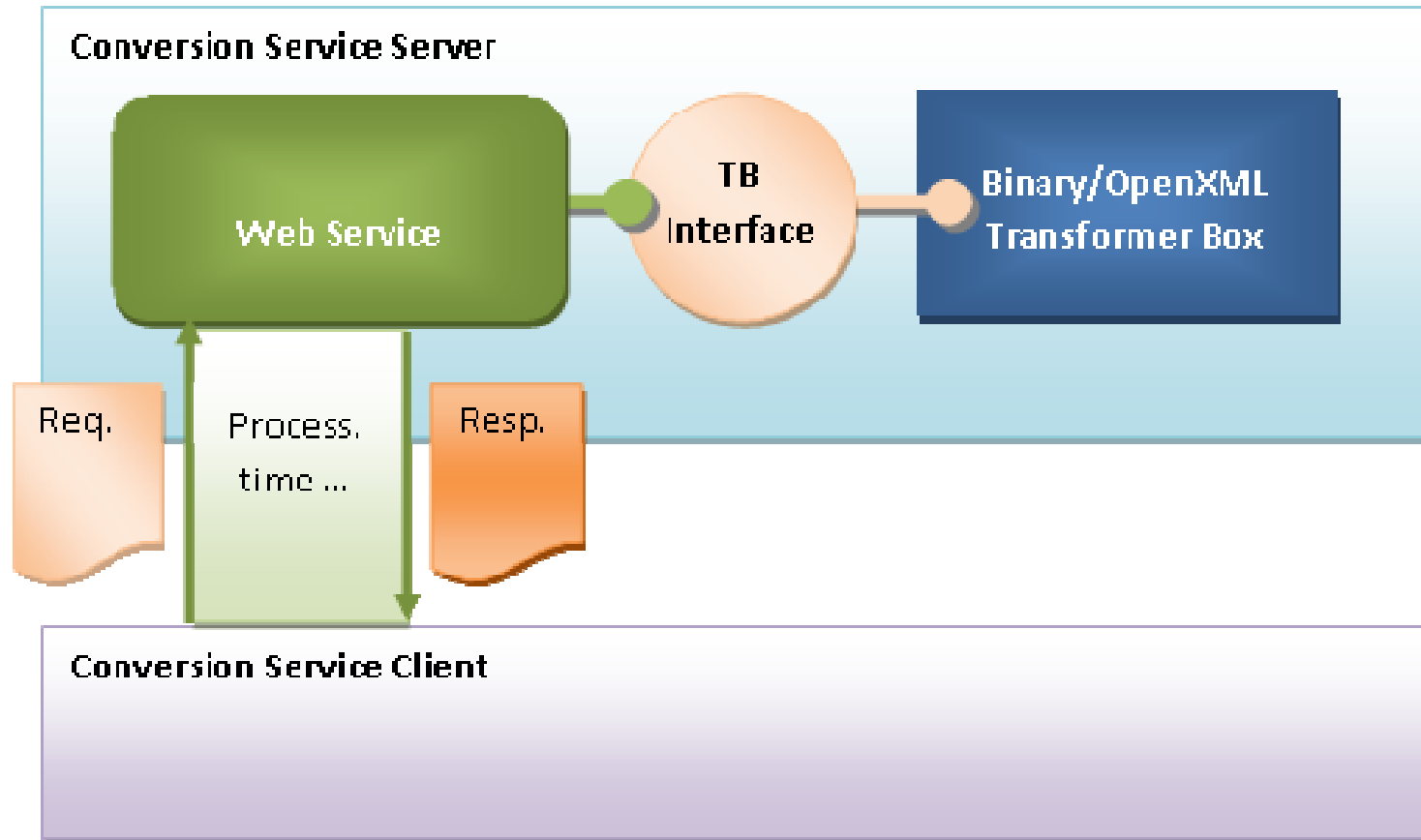 - Address a root cause of digital obsolescence

Preservation actions
Transform content

- Wrap third-party transformation tools
- Fill gaps with new tools
- Preserve relational databases
  - Build on Swiss Archive work
- Preserve Office content
  - Build on MSFT tools

Illustration of the use of service oriented architecture

Preservation actions
Transform environments

- Modular emulation of the full hardware/software environment
  - Provides full look-and feel
  - Superb for highly dynamic content
- Leverage Virtual Machine technology
- Layered durable emulation
  - Build on IBM Universal Virtual Computer (UVC)
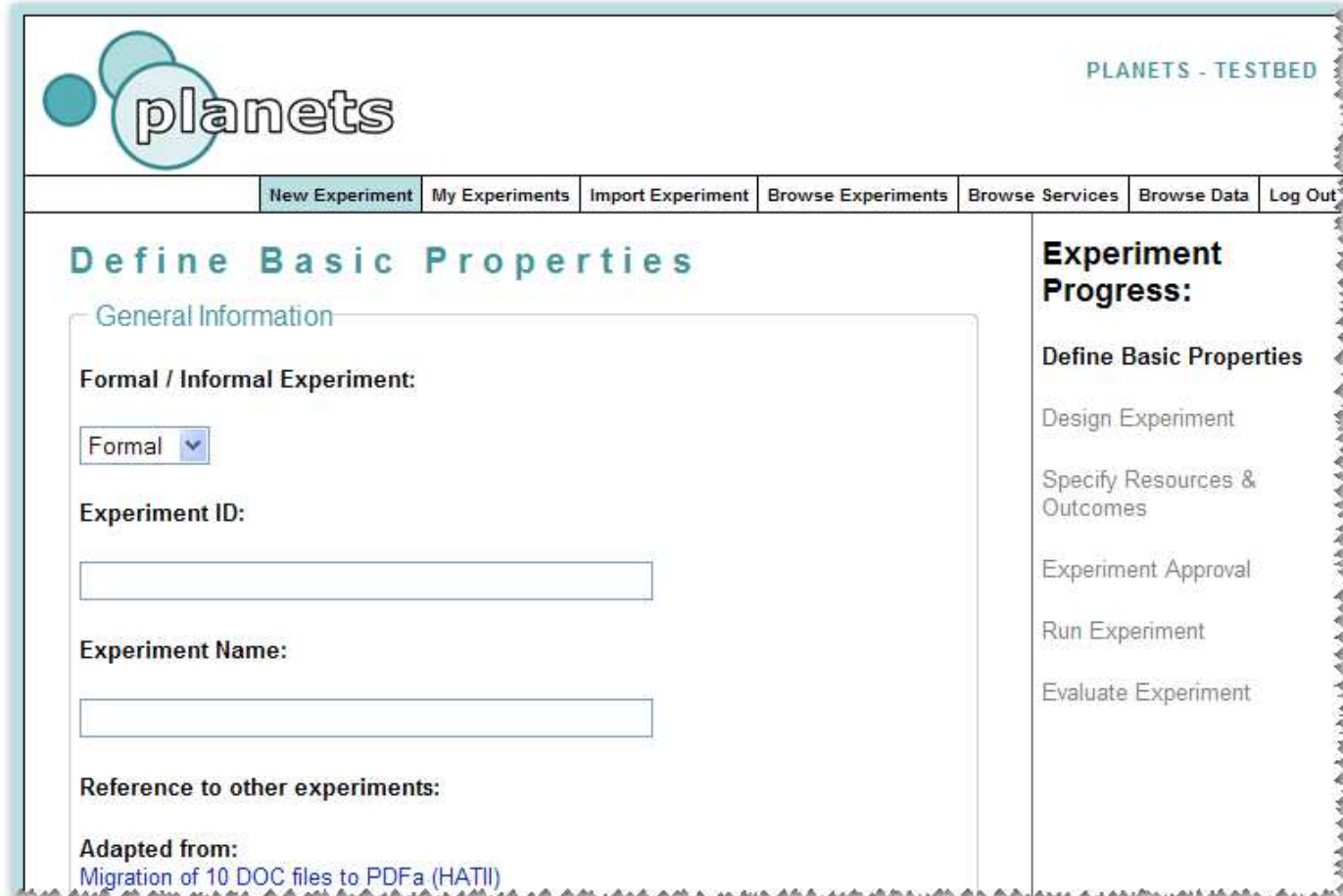 - Establish abstract device drivers

Testbed

- **Provides a foundation for objective evaluation**
  - Load content
  - Experiment: collect data, evaluate results, compare outcomes
  - Validate preservation plans
  - Benchmark tools and services
- **Consists of**
  - Data storage, hardware, Planets software, testbed software
  - Benchmark and other content
- **Provides resources for**
  - The project partners
  - The preservation community
  - External organisations
  - Tool and service certification

Testbed – Screen Shot

Design Experiment

Status

- **Fall 06**
  - Built the team
  - Gathered initial requirements
  - Conducted workshops and surveys
- **Winter 07**
  - Built specifications
  - Evaluated component technologies
- **Spring 07**
  - Finalised many technical and implementation decisions
  - Started to build tools and services
- **Summer 07**
  - Initial prototypes completed
  - First experiments conducted

Conclusion

- Planets Is a major EU co-funded digital preservation project
- Addresses the needs of Libraries and Archives
- Has made substantial progress towards a service-oriented preservation infrastructure
- Looks forward to working with the international Digital Library community to test, evaluate, refine, improve