



Curating and Preserving
Collaborative Digital Experiments

Jose Enrique Ruiz
IAA-CSIC

May 19th 2011
2011 IVOA Spring Interop Meeting - Naples



Wf4Ever team



1. Intelligent Software Components (ISOCO, Spain)
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (PSNC, Poland)
5. University of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)



The University
of Manchester



Astronomy research is entirely digital Time has come to go "Beyond the PDF"

- Preserved experiments
- Methodology "in action"
- All data exposed
- Reproducible
- Repeatable
- Reusable
- Repurposeable
- Participatory
- Collaborative
- Formative



Wf4Ever goals

All components related to the research lifecycle should be available.

Preserved and easily retrievables

- Proposals
- Data
- Processes
- Workflows
- Publications

RESEARCH OBJECT



Research Objects in Astronomy

- Metadata (Author, Instrument, Research group, etc.)
- Description of the experiment (Strategy, Expected results, etc.)
- Observing proposal
- Auxiliary and raw data
- Reduced science-ready data
- Digital environment needed
- Scripting and software used
- Web services
- **Scientific workflow**
- Final data products
- Standard publication



Scientific Workflows

- Automation
- Repeatable
- Reproducible
- Encourage best practices
- Modular nature allows
 - Reuse
 - Repurpose
- Exposes the scientific method
- Formative
- Scientist friendly



Scientific Workflows

- Automation vs. The intrinsic exploratory nature of Science
- Documented vs. Hidden knowledge
- Web services vs. Local software
- On-line data vs. Local data
- Modular vs. Unstructured
- Open Science vs. Proprietary
- Preserved
- Classified and indexed
- Referenced and retrievable



Workflow preservation is complex

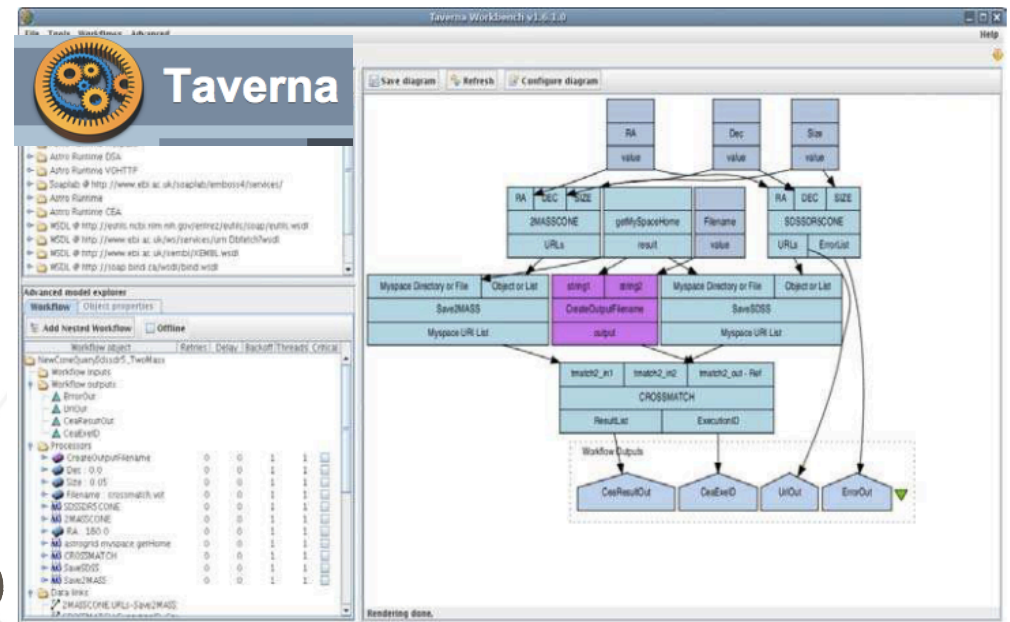
- Interpreted through their execution
- Complex models are required to describe them
- Provenance is a complex issue in a cloud of services
- Need of Web Semantics, Ontologies, Linked Data, etc..
- Resources are often beyond control of scientists

The oven

A workflow enactment and management system
University of Manchester

- AstroTaverna (AstroGrid)
 - SOAP
 - AstroRuntime

- Reflex (ESO)
- AladinJLOW Plugin (CDS)



The recipes store Oxford university

- Find workflows
- Share workflows and files
- Find people
- Build communities
- Publish packages
- Tag workflows
- Score and rate workflows
- Comment on workflows
- Write reviews



The screenshot shows the 'myexperiment' website interface. At the top, there are navigation links for 'Home', 'Users', 'Groups', 'Workflows', 'Files', and 'Packs'. A search bar is present with a dropdown menu set to 'All'. The main content area is titled 'Workflows' and displays a list of workflow entries. The first entry is 'Mapping OligoNucleotides to an assembly (v7)' by Wassinki, created on 13/02/09. It includes a description, a license (Creative Commons Attribution-Share Alike 3.0 Unported License), and a version history section. The second entry is 'Add Mesh String to Biological Process (v2)' by Paul Fisher, created on 03/10/07. It also includes a description and a license. On the right side, there is a 'New/Upload' section with a 'Workflow' dropdown and a 'GO' button. Below that is a 'Log in / Register' section with fields for 'Username or Email', 'Password', and 'Remember me'. There is also a 'Popular Tags' section with 25 tags, including 'benchmarks', 'bioinformatics', 'BLAST', 'cheminformatics', 'data integration', 'ebi', 'example', 'gene', 'graph', 'kagg', 'Kegg Pathways', 'localworker', 'microarray', 'mygrid', 'ondex', 'pathway', 'pathways', 'phenotype', 'protein', 'pubmed', 'sequence', 'taverna', and 'text mining'.

Living Working Research Objects

- Ubiquitous storing and computing
 - Data archives and local data
 - Web services and scripts
 - Python based community
 - VO standards
-
- Modular to reuse individual parts
 - Access rights at different levels of granularity
 - VOspaces



Published Research Objects

- Archival
- Classification
- Indexing
- Retrieval
- Versioning

- Community reuse
- Rating, scoring and annotations
- Scalable in semantic repositories

- Permanent URIs, Linked Data, Semantics, etc.
- Interlink with catalogs/digital libraries



Users roles

Collaborator

Dealing with Living Working Research Objects in a research group.

Reader

Skims titles and abstracts of Published Research Objects.

Comparator

Looking for similar Research Objects to those she/he is working with.

Re-user

Extract modules from workflows and use them for his own purpose.

Publisher

Wants her/his work to be known.

Evaluator

He evaluates, rates, comments and recommend a specific Research Object.

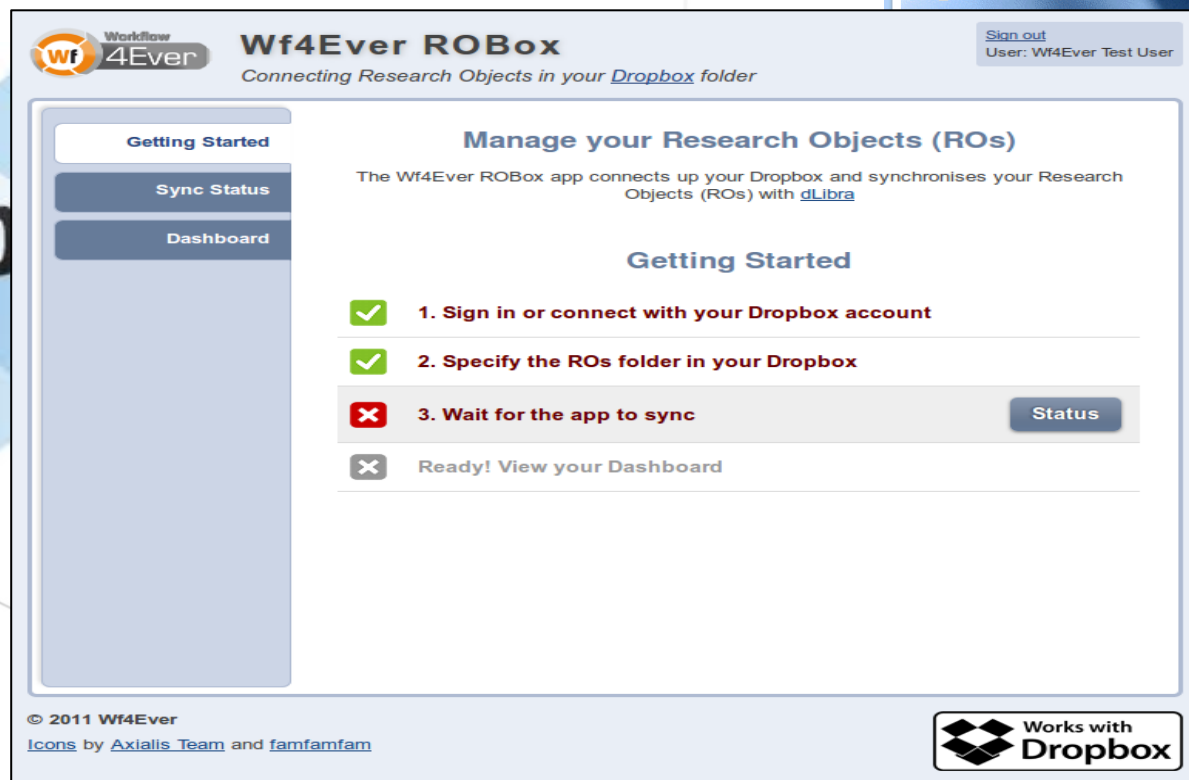
Most of them are **active roles** run the workflows with (different) data

ROBox: the basket

Seamless contribution to a collaborative platform

A shared folder in Dropbox becomes a Working Research Object

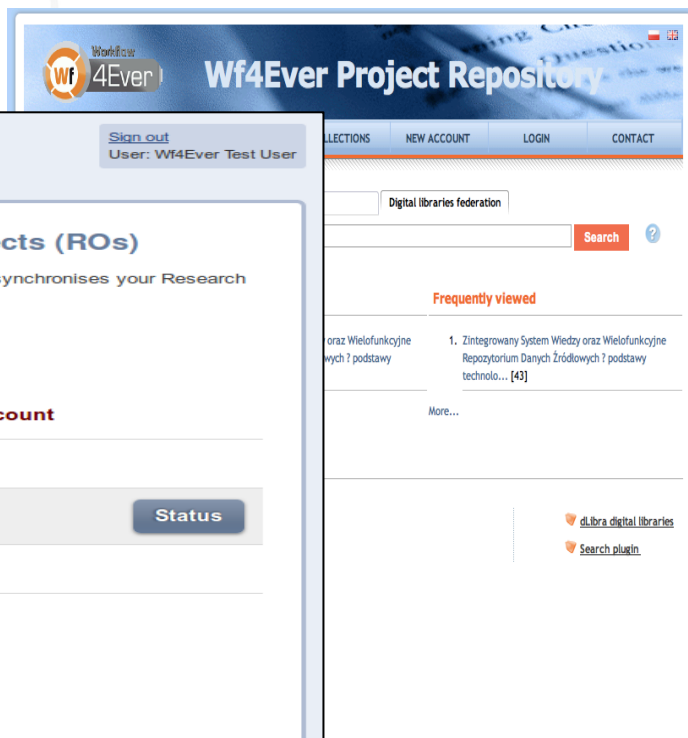
Automatic generation of metadata



The screenshot shows the Wf4Ever ROBox application interface. The header includes the Workflow 4Ever logo and the text "Wf4Ever ROBox" and "Connecting Research Objects in your [Dropbox](#) folder". A sidebar on the left contains navigation links: "Getting Started", "Sync Status", and "Dashboard". The main content area is titled "Manage your Research Objects (ROs)" and includes a sub-header "Getting Started". Below this, there is a list of steps:

- ✓ 1. Sign in or connect with your Dropbox account
- ✓ 2. Specify the ROs folder in your Dropbox
- ✗ 3. Wait for the app to sync [Status](#)
- ✗ Ready! View your Dashboard

At the bottom of the interface, there is a copyright notice: "© 2011 Wf4Ever" and "Icons by [Axialis Team](#) and [famfamfam](#)". A "Works with Dropbox" logo is also present.



The screenshot shows the Wf4Ever Project Repository website. The header includes the Workflow 4Ever logo and the text "Wf4Ever Project Repository". A navigation bar contains links: "COLLECTIONS", "NEW ACCOUNT", "LOGIN", and "CONTACT". Below the navigation bar, there is a search bar with the text "Digital libraries federation" and a "Search" button. A "Frequently viewed" section is visible, listing items such as "1. Zintegrowany System Wiedzy oraz Wielofunkcyjne Repozytorium Danych Źródłowych ? podstawy technolo... [43]". At the bottom right, there are links for "dl:libra digital libraries" and "Search plugin".

We are moving into a world where *computing and storage are cheap* and *data movement is death*.

In a Cloud of services and data, *web services should benefit of the same privileges acquired by Data*.

- Curation and preservation (identifiers)
- Discovery (semantics) of web services (linked "services"?)
- Characterization: input, outputs, functionality, etc.
- Copies (authenticity) or similar web services used as alternates
- Permissions, licenses, platform, costs, etc.
- Metrics for quality: popularity, use stats, logs uptime, etc.
- Versioning and authoring (referenced and acknowledged)