

A Laboratory Notebook System

EuroPython 2012 (05.07.2012, Florence, Italy)

Andreas Schreiber <Andreas.Schreiber@dlr.de>
German Aerospace Center (DLR)



Knowledge for Tomorrow



Overview

- Background
 - Good Laboratory Practice
 - Scientific Workflows
 - Laboratory Notebooks
 - DataFinder

- DataFinder-based Laboratory Notebook
 - Data model
 - Process documentation
 - Evidential preservation
 - Signing data

- Future Work



Background



Knowledge for Tomorrow



Background

Good Laboratory Practice

*The principles of Good Laboratory Practice (GLP) have been developed to **promote the quality and validity** of test data used for determining the safety of chemicals and chemicals products.*

OECD Principles on Good Laboratory Practice (as revised in 1997)

*[The recommendations] are designed to provide **a framework for the deliberations and measures** which each institution will have to conduct for itself according to its constitution and its mission*

Deutsche Forschungsgemeinschaft:

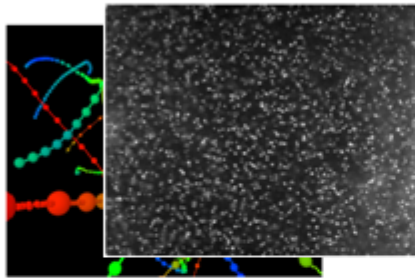
Sicherung guter wissenschaftlicher Praxis (Safeguarding good scientific practice) 1998 (p.50).



Background

Scientific Workflow

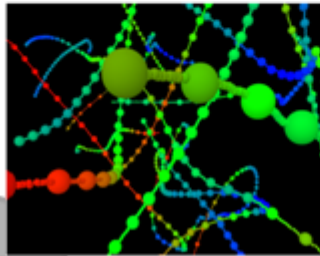
5 **Archiving:** subsequent use, display of research results



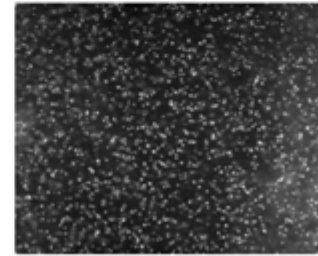
1 **Planning, Design:** Literature, Data of other projects



4 **Interpretation, Publication**



2 **Execution:** Apparatus, Computer



3 **Evaluation:** Computer, (self developed) Software

Picture adapted from: www.belab-forschung.de

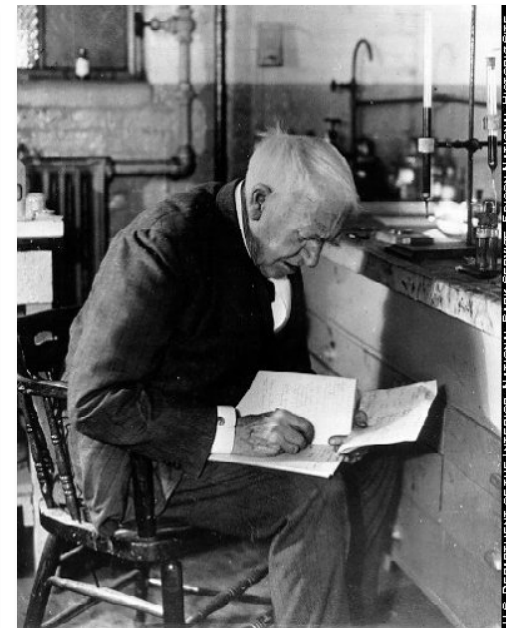
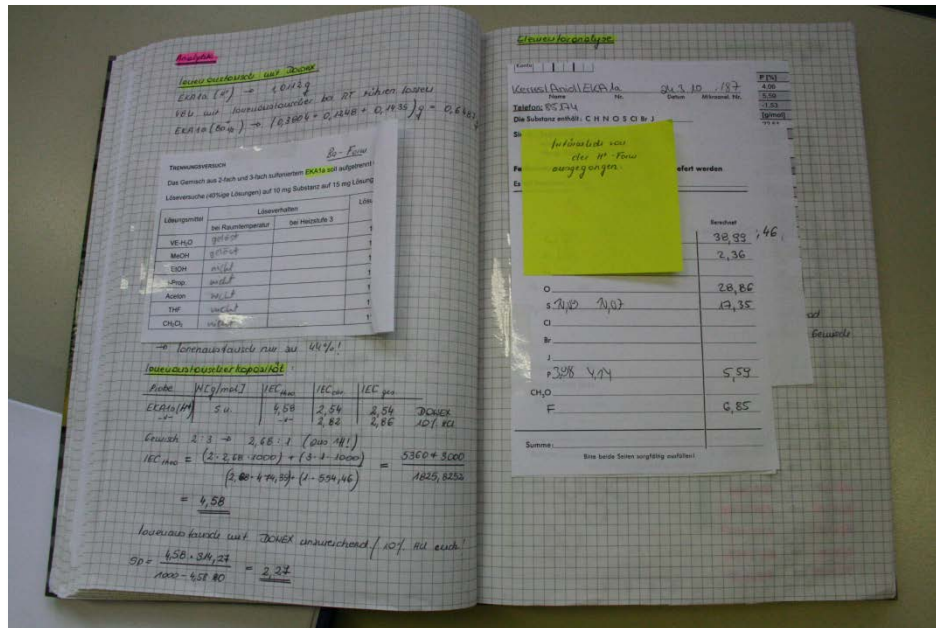


Background

Laboratory Notebooks

“The laboratory notebook is the diary of the experimenting scientist“

(Schreiben und Publizieren in den Naturwissenschaften
 Von Hans F. Ebel, Claus Bliefert, Walter Greulich; chapter 1.3 - page 16)





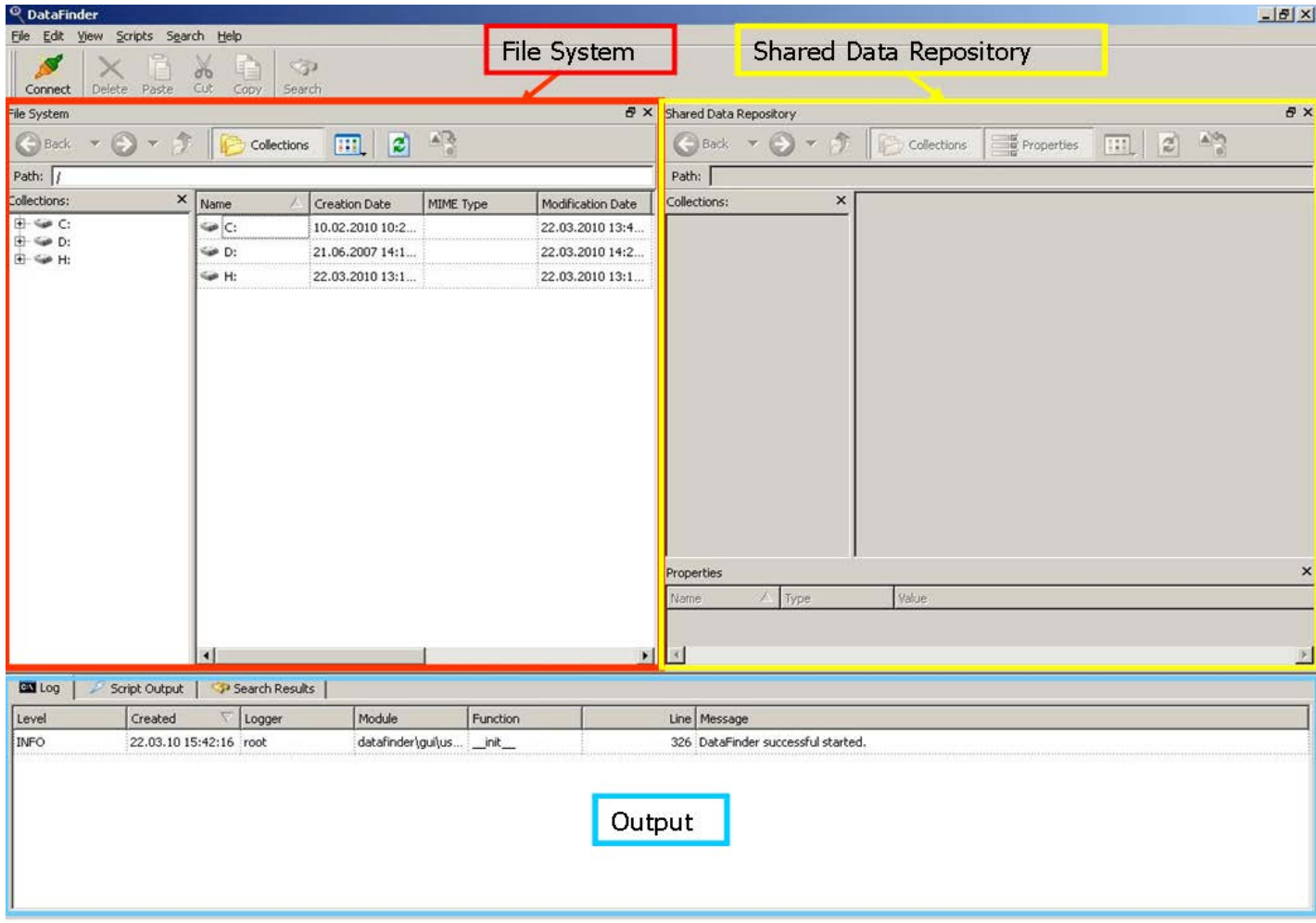
Background

DataFinder

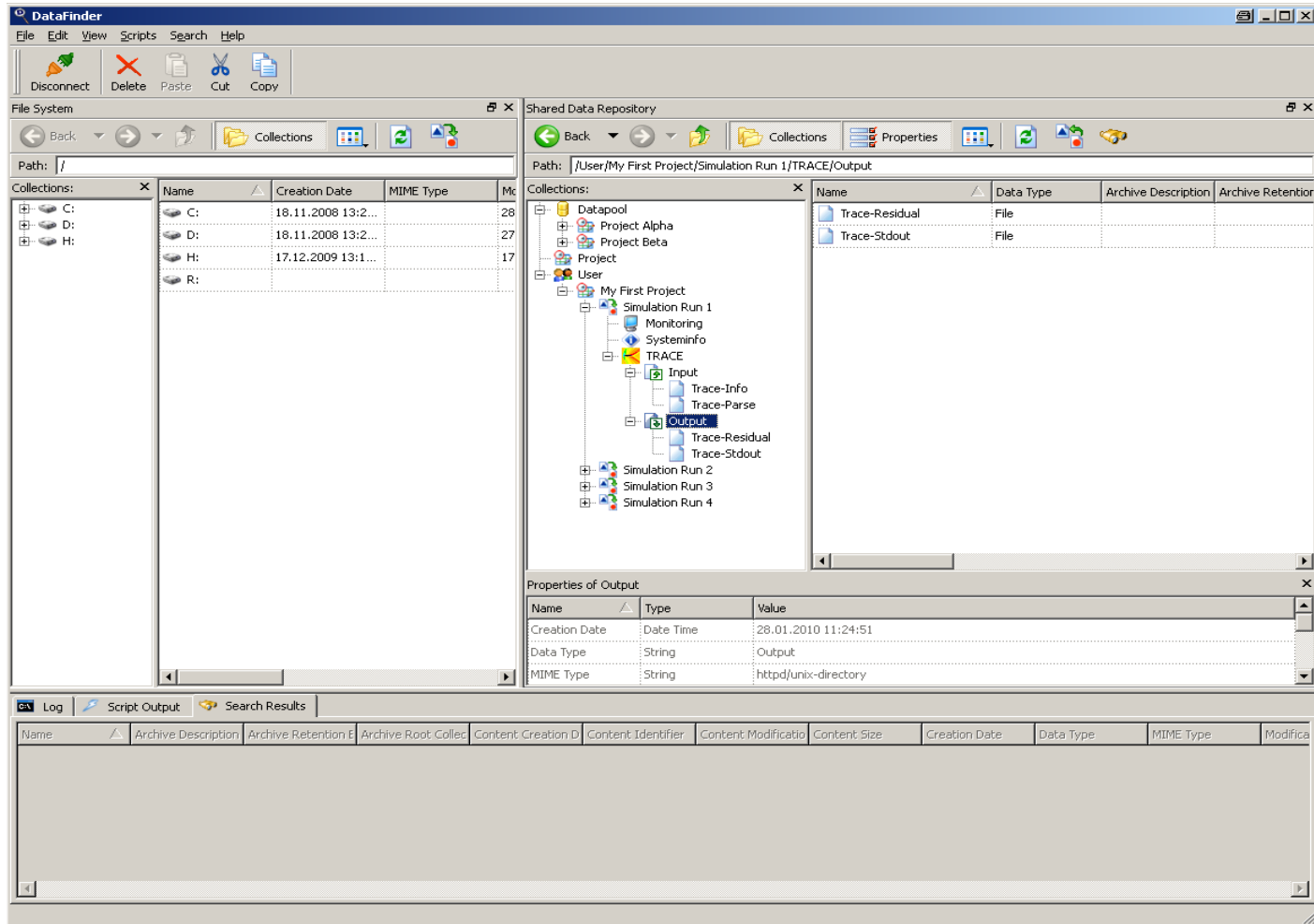
- Data management system: DataFinder
 - Developed by DLR
 - Open Source Project (BSD License)
 - Implemented in Python
 - Data management and work flow management
 - Supports meta data handling



DataFinder User Interface

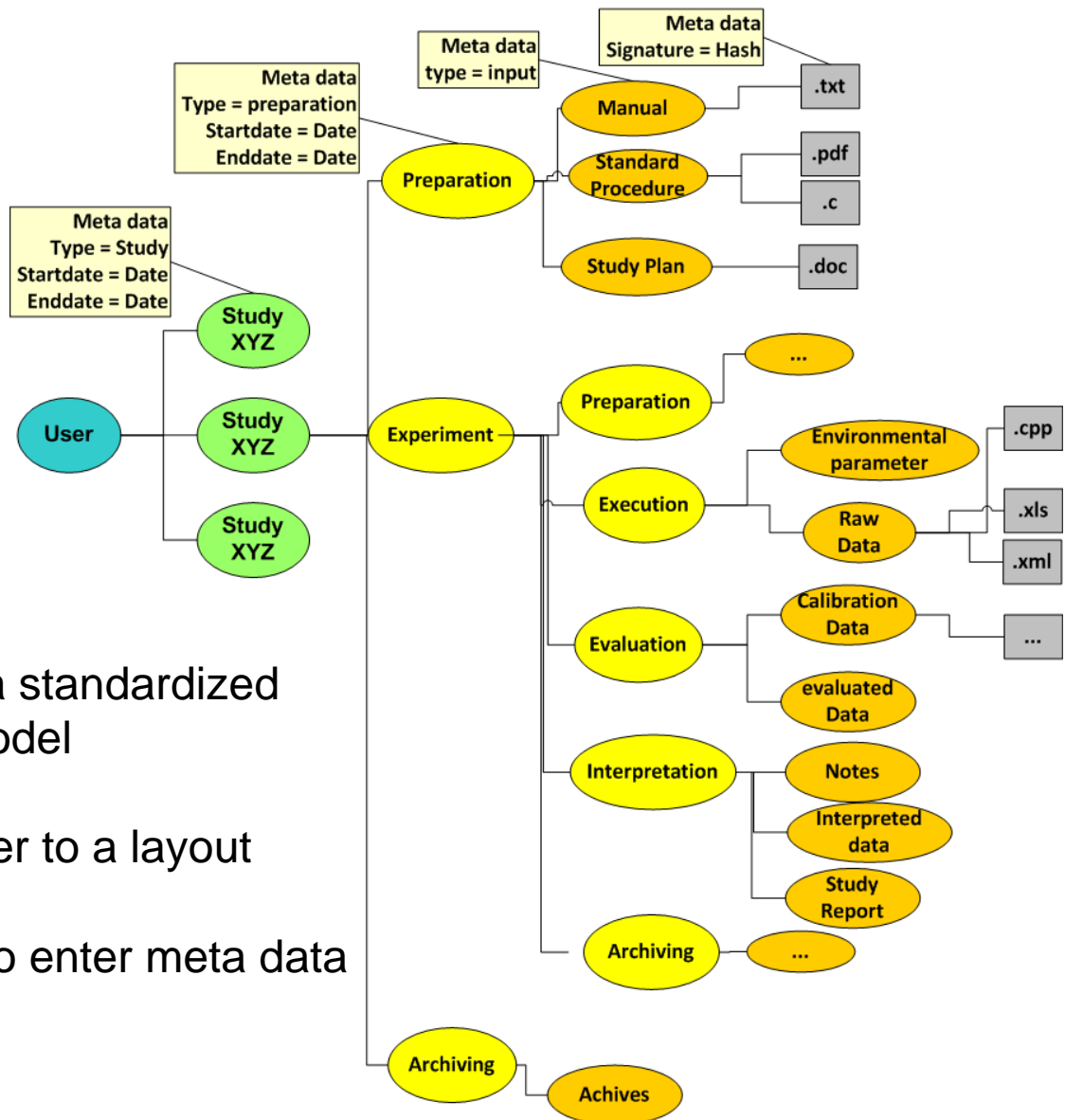


DataFinder – Connected to Repository



DataFinder

Structuring Data



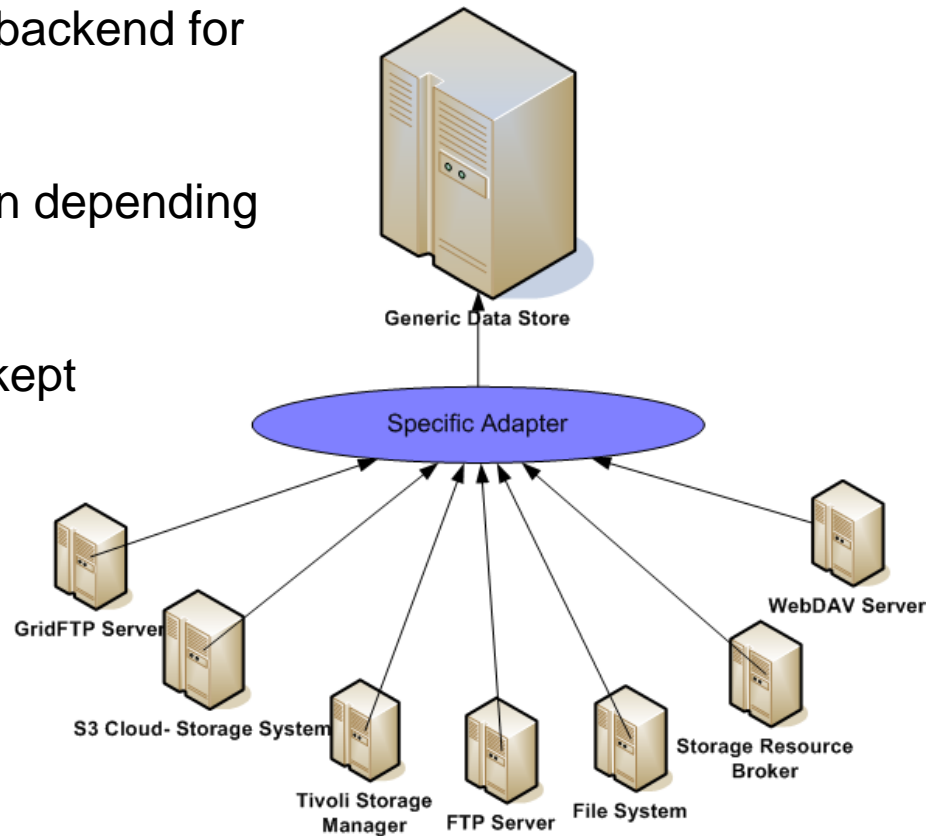
- Structuring of data in a standardized way through a data model
 - Restricting the user to a layout
 - Forcing the user to enter meta data



DataFinder

Heterogeneous Storage Resources

- Using heterogeneous storage backend for data
 - Best fitting storage solution depending on data
 - Existing solutions can be kept
 - Using offline storage is possible



DataFinder

Script Extensions

- DataFinder is extendable by Python scripts
 - Integration with existing environment
 - Automation of data processing steps



DataFinder-based Laboratory Notebook

Knowledge for Tomorrow



Laboratory Notebook

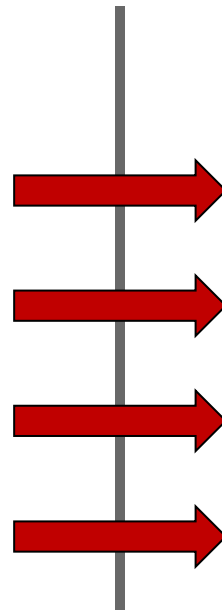
Requirements for Good Scientific Documentation

Requirements:

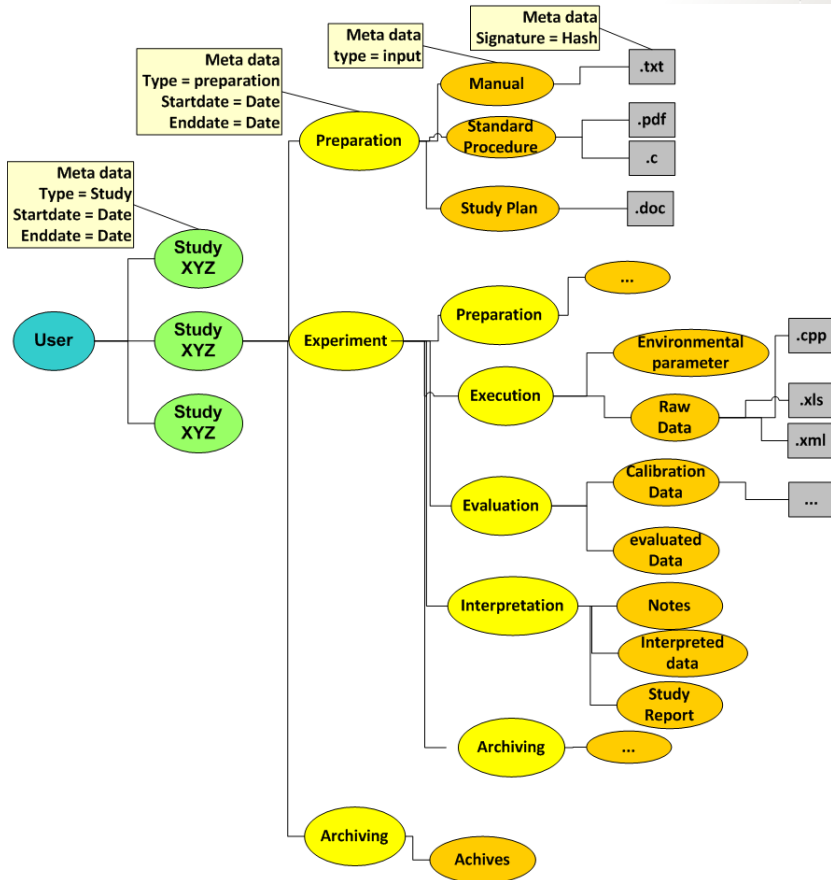
- Data structure
- Traceability
- Durability
- Credibility

Realization:

- Data model
- Process documentation
- Evidential preservation
- Signing data



Realization Data Model



Shared Data Repository

Path: /Study_DLR/Experiment_Config_1/Report_First_Config/Proceedings_IEEE_2011

Collections:

- Department general information
 - camera systems
 - windtunnel
- Study_DLR
 - Archives
 - Experiment_Config_1
 - Modell_Drawings
 - Report_First_Config /Study_DLR/Experiment_Config_1
 - Proceedings_IEEE_2011
 - Summarizing
 - Results_March_2011
 - Test_02_03_2011
 - Test_14_03_2011
 - Project information
 - project plan

Name	Type	Value
Creation Date	Date Time	08.03.2011 13:39:18
Data Type	String	Study Report
MIME Type	String	httpd/unix-directory
Modification Date	Date Time	08.03.2011 13:45:35
Owner	String	
Size	Number	0 KB

Function	Line	Message
%(core)... load	76	Loading icons...
%(core)... load	71	Loading scripts...
%(qtuius)..._init_	357	DataFinder successful started.



Realization

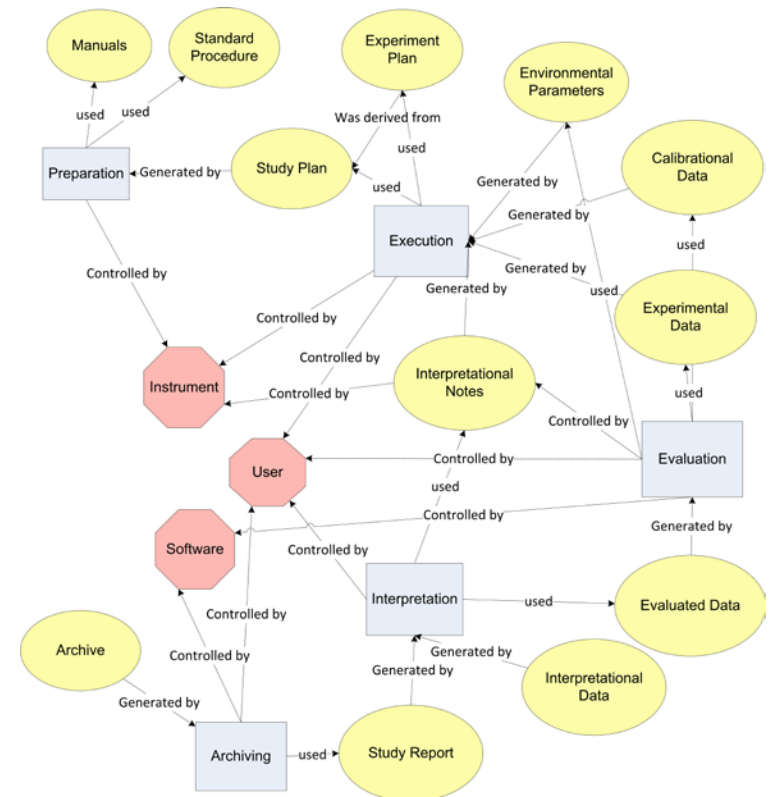
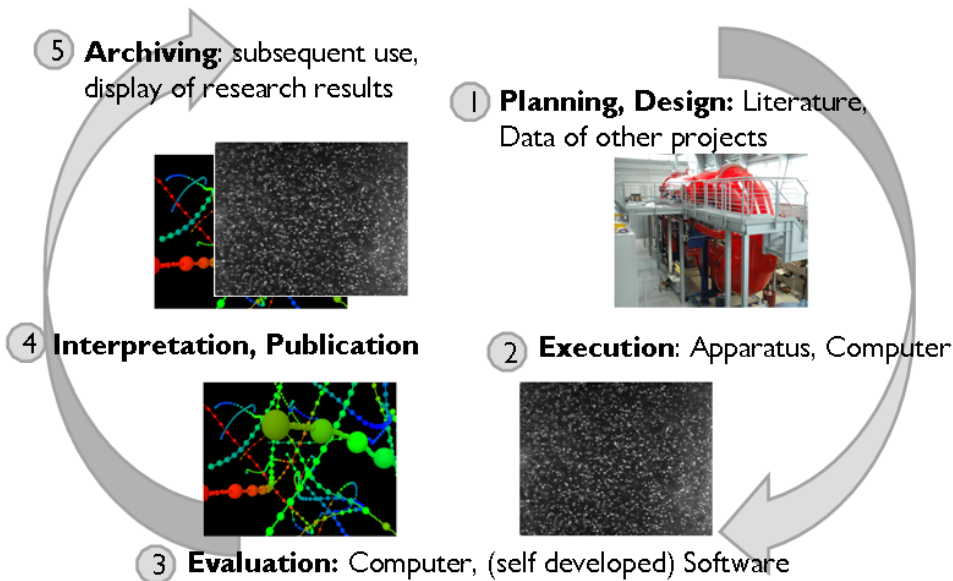
Process Documentation

- Process documentation: Recording the **Provenance** of that process
- Provenance (lat. provenire = to come from): origin of data, source
- Provenance of process gives **traceability** and **credibility**
- Steps to add Provenance recording to software (i.e., DataFinder)
 1. Developing a provenance model for the „Good Laboratory Practice“
 2. Provide Provenance storing system
 3. Integration into DataFinder

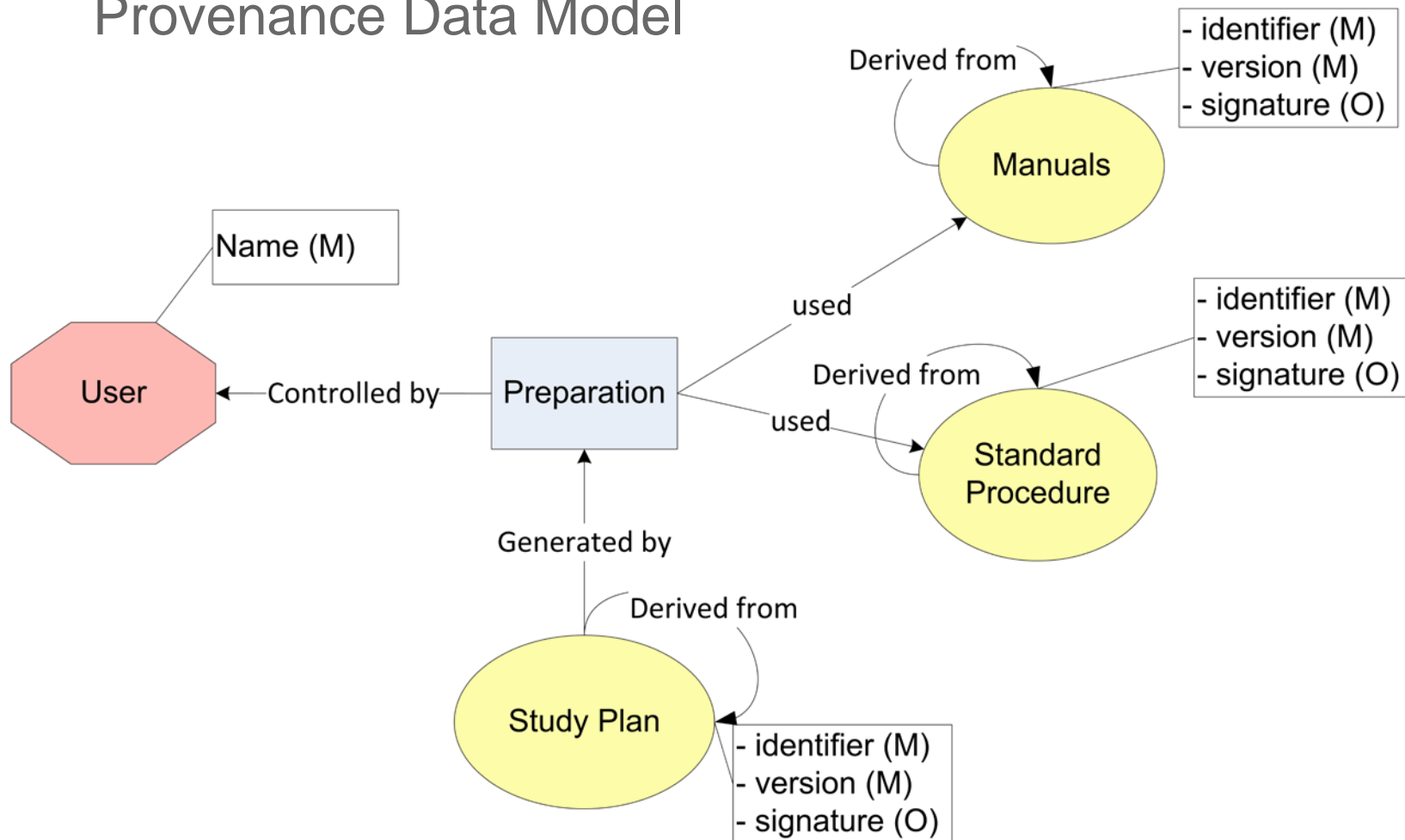


Process Documentation Provenance Data Model

- Apply methodology to define a Provenance model
- Representation of the real world's process



Process Documentation Provenance Data Model



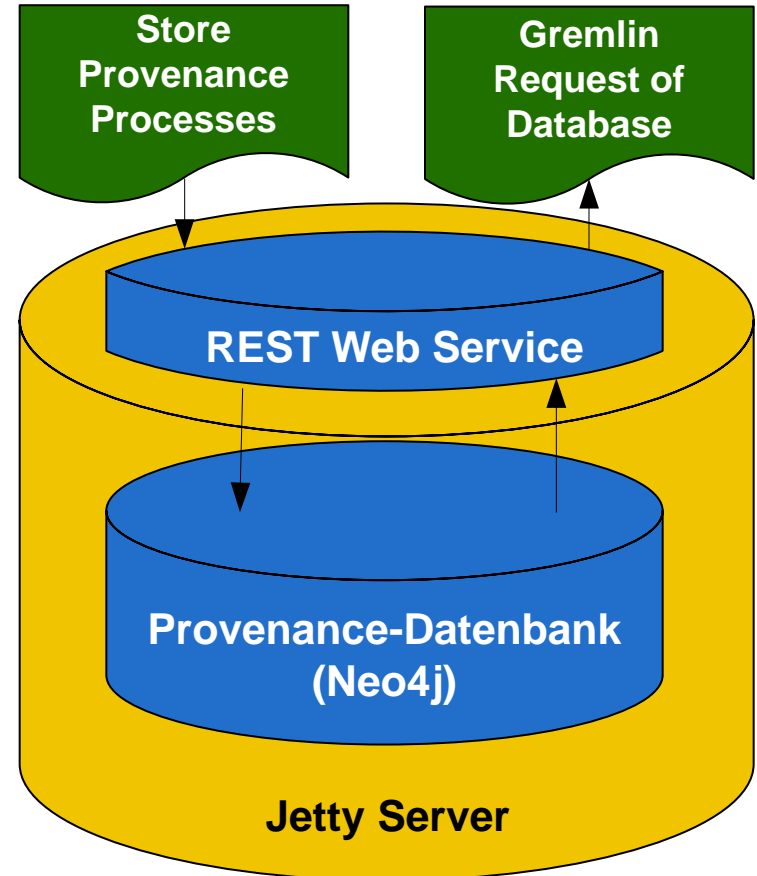
Process Documentation Provenance Storing System

Provenance Store *prOost*

- Java Implementation
- Server: Jetty
- Graph Database: Neo4j

- Interfaces
 - Storing Provenance (REST)
 - Extracting Provenance (REST)
 - Extracting Provenance (Servlet)

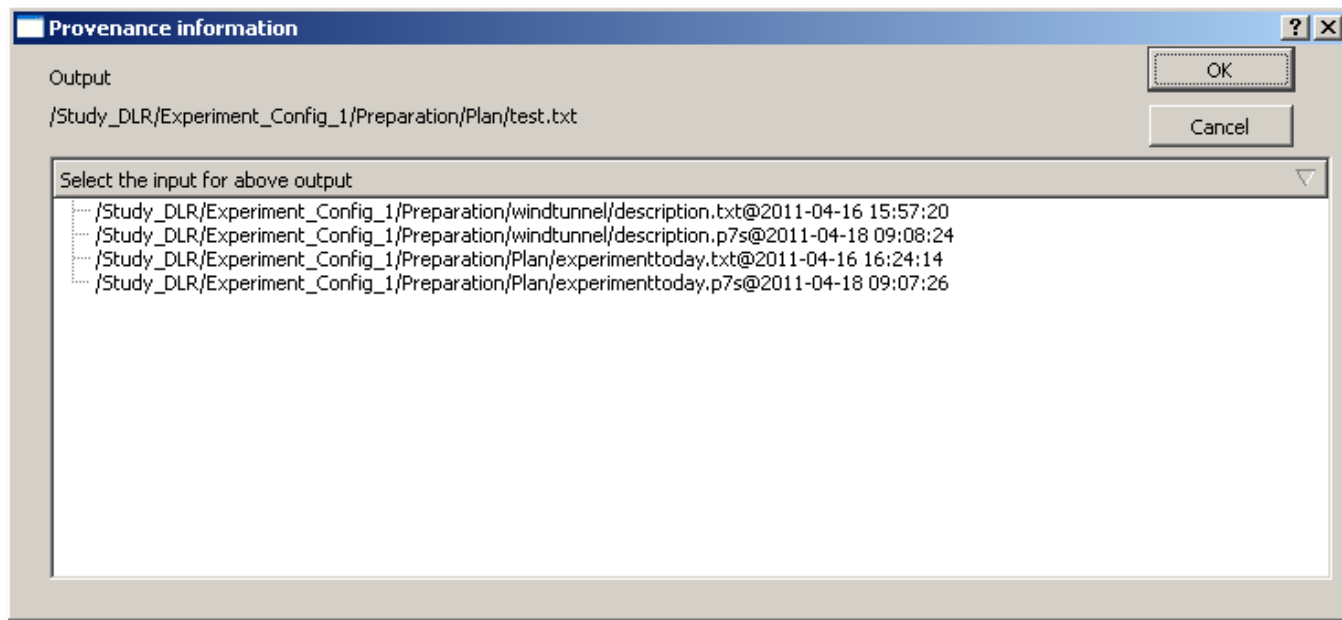
- Open Source (Apache License 2.0)
 - <https://proost.sourceforge.net>



Process Documentation

Integration Into DataFinder

- User actions on files are recorded in the provenance store
- Dialog for asking additional questions



Realization

Evidential Preservation

„Recommendation 7: Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.“

Deutsche Forschungsgemeinschaft:

Sicherung guter wissenschaftlicher Praxis (Safeguarding good scientific practice) 1998 (p.55).

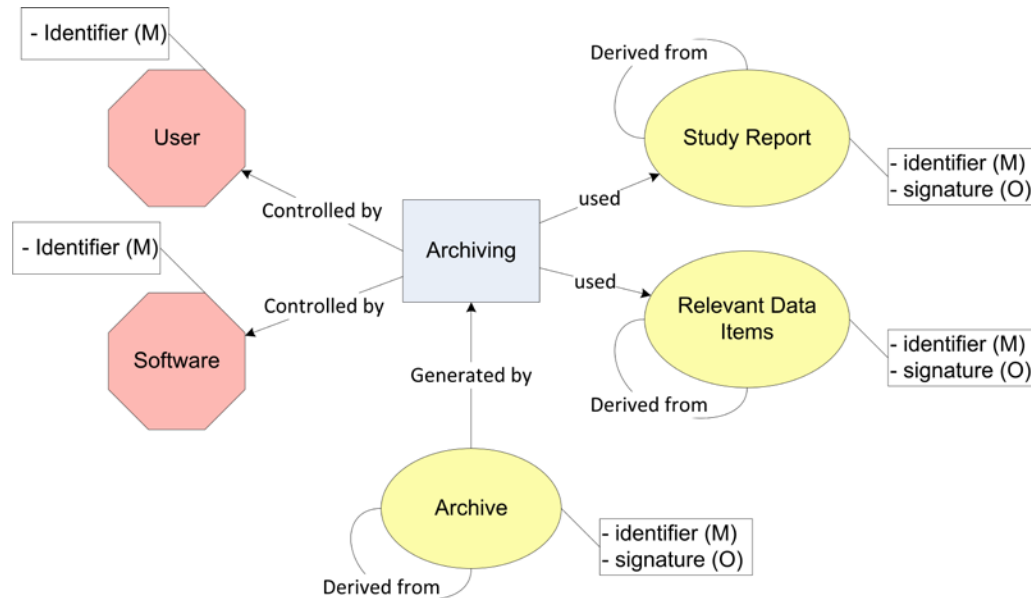
- Steps to add evidential preservation to software (i.e., DataFinder)
 1. Create an archive with all relevant data (e.g., for a publication)
 2. Integration of a preservation service



Evidential Preservation

Create an Archive With All Relevant Data

Extraction of data relevant for the preservation process

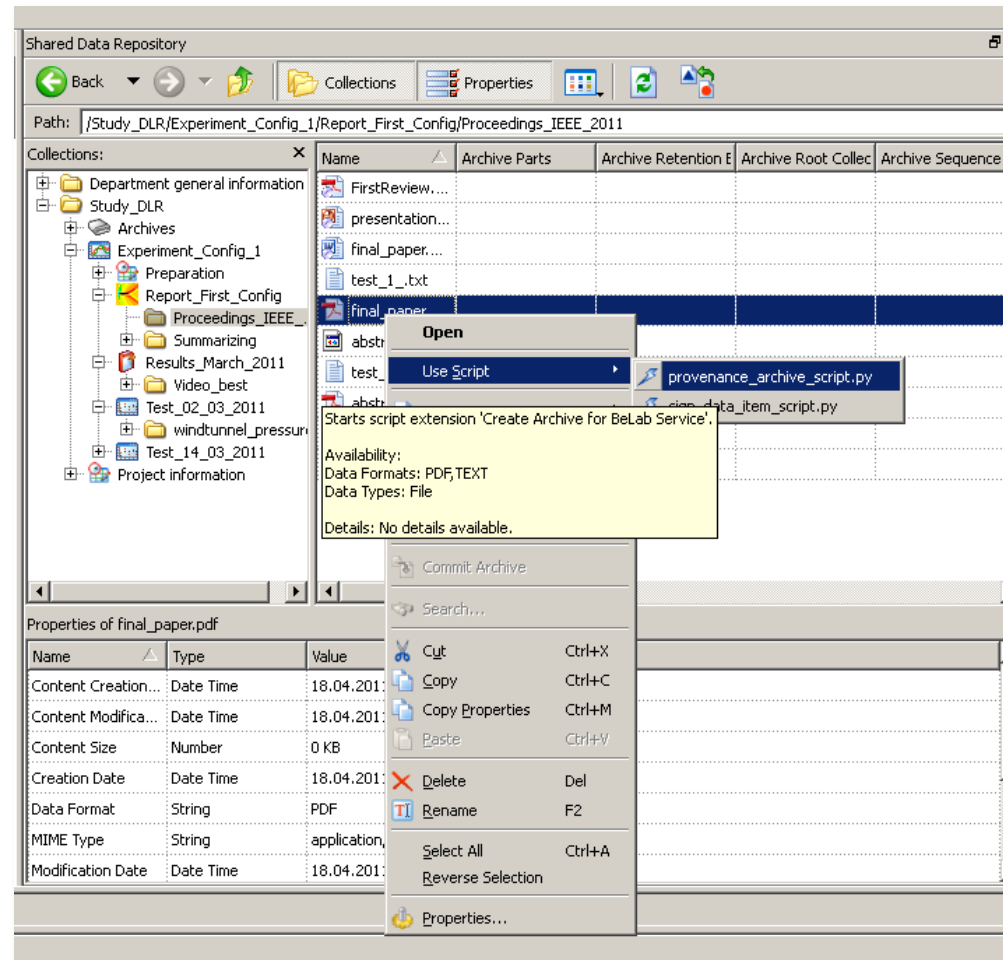


Evidential Preservation

Create an Archive With All Relevant Data

In DataFinder

- User chooses report (publication etc.)
- Python script queries relevant files from the Provenance store
- Relevant files are added to an archive
- Archive is stored in DataFinder



Evidential Preservation

Integration of a Preservation Service

We use the BeLab service (Beweissicheres Laborbuch Project)

- DFG Project (<http://www.belab-forschung.de>):
 - Physikalisch Technische Bundesanstalt Braunschweig
 - Karlsruher Institute of Technology
 - Universität Kassel
- The BeLab service
 - characterizes the preservation time of an item
 - characterizes the legal trustworthiness of an item
 - stores the archive securely



Evidential Preservation

Integration of a Preservation Service

In DataFinder

- User chooses an archive and activates script
- Script sends the archive to BeLab service via WS-Secure
- The service processes the archive
- Service returns preservation information, which is stored



Realization

Signing Data

- Authenticity in general
- Attesting authentication
- Steps to add data signing to software (i.e., DataFinder)
 1. Concept:
 - Signing files: signature stored as meta meta item
 - Meta data: Extraction as XML file, then signed
 2. Integration into DataFinder



Signing Data

Integration Into DataFinder

Signature of the data (files) as separate file

- User chooses a file and executes script
- A signature file is generated (PKCS #7)
- Signature file is stored in the DataFinder

The screenshot shows the DataFinder application interface. At the top, a file tree displays a hierarchy of folders including 'Department general information', 'Study_DLR', 'Archives', 'Experiment_Config_1', 'Preparation', 'Plan', 'windtunnel', 'Report_First_Config', 'Proceedings_IEEE_2', 'Summarizing', 'Results_March_2011', 'Video_best', 'Test_02_03_2011', 'Test_14_03_2011', and 'Project information'. A file named 'minuteOne.txt' is selected in the tree.

Below the tree, the 'Properties of minuteOne.txt' window is open, showing the following details:

Name	Type	Value
Content Creation...	Date Time	08.02.2011 19:14:04
Content Modifica...	Date Time	08.02.2011 19:14:20
Content Size	Number	0 KB
Creation Date	Date Time	16.04.2011 15:58:00
Data Format	String	TEXT
MIME Type	String	text/plain
Modification Date	Date Time	16.04.2011 15:58:09
Owner	String	

At the bottom, the 'Log' window displays a table of system messages:

Level	Created	Logger	Module	Function	Line	Message
INFO	18.04.11 09:08:30	root	datafinder.core.c...	load	102	Loading icons...
INFO	18.04.11 09:08:30	root	datafinder.core.c...	load	97	Loading scripts...
INFO	18.04.11 09:08:24	webdavLogger	lib.site-packages...	_blockCopySocket	227	Transferred 1415 bytes.
INFO	18.04.11 09:08:24	script	signdata.sign_da...	handleSigning	98	signature successfully stored

ScreenCast-O-Matic.com

Future Work



Knowledge for Tomorrow



Future Work

Enhanced User Interface

- User interface for taking notes
 - Annotation of data
- Doing calculations and data analysis (similar to MATLAB or Mathematica Notebooks)
 - Integration of The Larch Environment
 - Integration of NumPy/IPython
- Exploring Provenance data
 - Insights and understanding of processes
- Tablet version
 - Entering data
 - Synchronization for offline use



Questions?

Summary

- DataFinder-based Electronic Lab Notebook
- Traceability, Durability, and Credibility for data
- Documentation, evidential preservation, and data signing

Andreas Schreiber
Andreas.Schreiber@dlr.de
<http://www.dlr.de/sc>

