CASPAR - Preservation Methodology Validation (1) testbed demos (2) Audit and Certification and (3) Provenance and Authenticity

Stephen Rankin (STFC), Matt Dunckley (STFC), Brian Mcilwrath (STFC), Esther Conway (STFC), David Giaretta (STFC), CASPAR & DCC

wePreserve

Outline

- RepInfo Network Example (Again).
- Validation with RepInfo.

Information Society

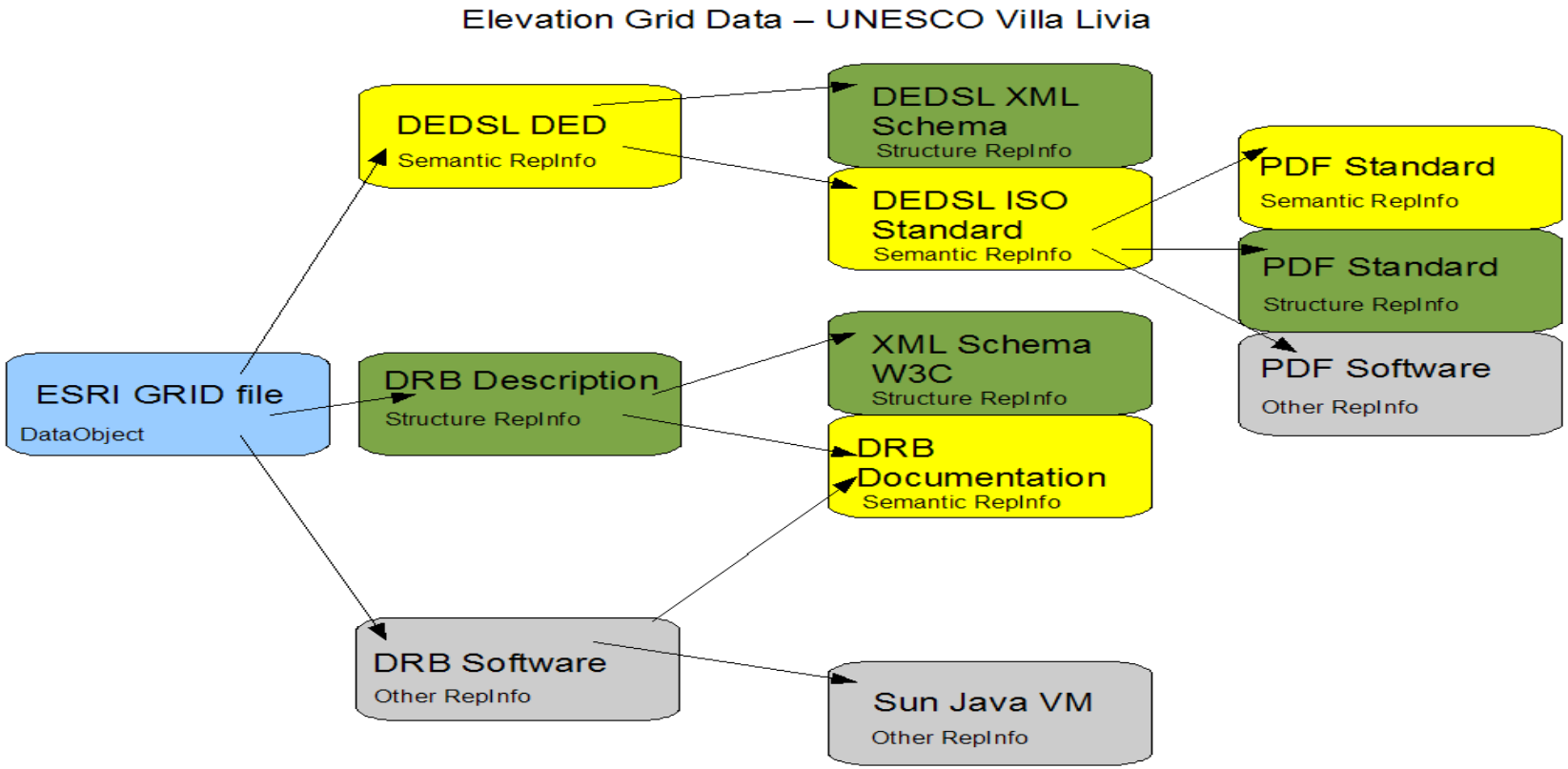
digital preservation Europe

planets

CASPAR

wePreserve

RepInfo Network Example



Information Society

digital preservation europe

planets

CASPAR

Validation with RepInfo

- Can a future user use the RepInfo and RepInfo network to reuse the data?
- Documents, software etc as RepInfo difficult to validate against the data. Quality of RepInfo.

What do we mean by reuse the data?

wePreserve

What do we mean by reuse the data?

- Simplify the problem!
- Go from the data bits to a simple Information Object
- This means reading the data with the Structure RepInfo and applying the Semantics RepInfo
- Feed Information Object into NEW software.

Information Society

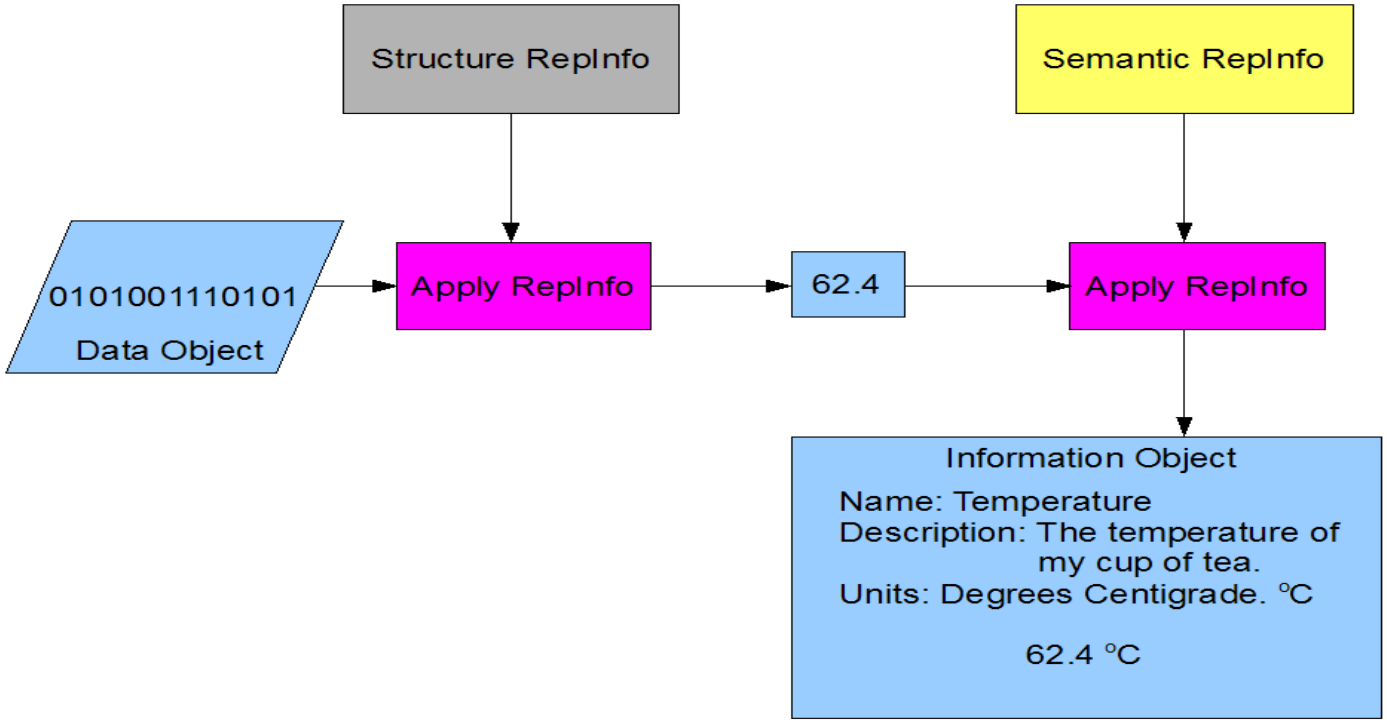
digital preservation Europe

planets

CASPAR

wePreserve

RepInfo for a Hot Cup of Tea



Information Society

digital preservation Europe

planets

CASPAR

Creation of new Formal Representation Information

- Curating existing RepInfo is OK and necessary, but we can do more…
- Typically documents that describe the structure and semantics do not allow easy validation.
- Information in documents typically not machine readable.
- Semantics is usually poorly linked to the data and data values.
- The types of simple data objects (data types) are not explicitly defined.
- Data will have to be read into future software if it is going to be reused.

wePreserve

Formal Descriptions of Structure and Semantics

- Digital data is composed of bits which somehow have to be mapped to primitive software data types (numbers, strings etc)
- Numbers etc need to be assigned simple semantics, Name, Description, Unites etc.

Information Society

digital preservation Europe

planets

CASPAR

wePreserve

Formal Descriptions of Structure

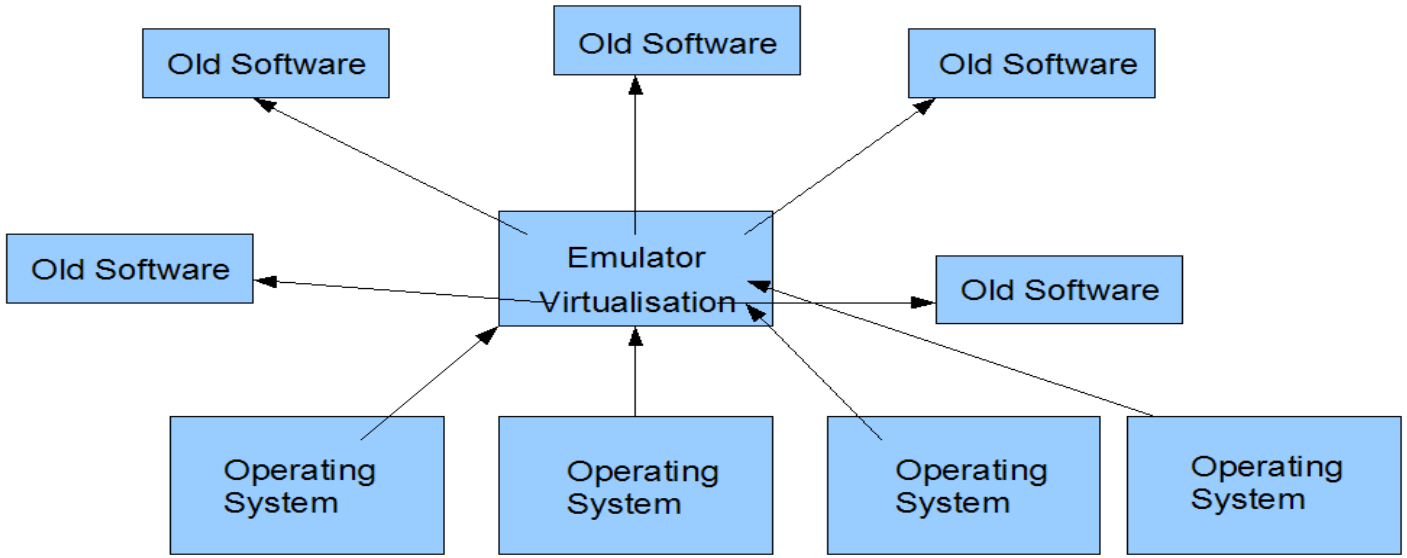
- CNES EAST tools (http://east.cnes.f), OASIS, EAST C Library (reference implementation).
- Also DEBAT (BEST Tools) http://debat.c-s.fr/
- Data Request Broker (DRB) - http://www.gael.fr/drb/site/
- JNI Wrapper for EAST C Library in our SVN repository (jnieast).
- Interfaces for a more general data description language structure and semantics API in our SVN (DSSIL). http://developers.casparpreserves.eu:8080/

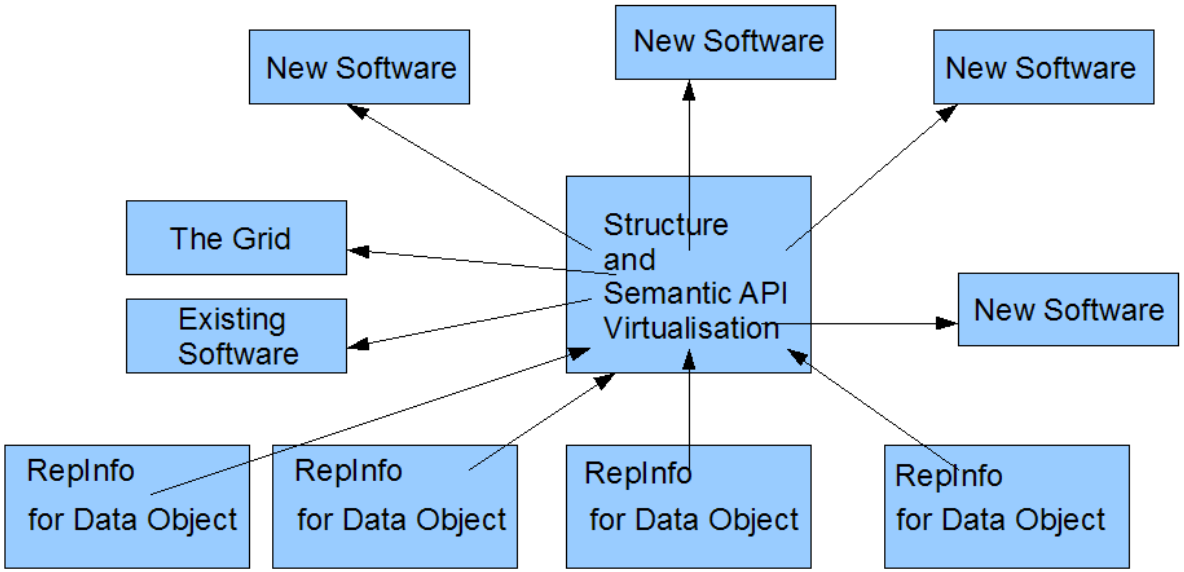
Information Society

digital preservation europe

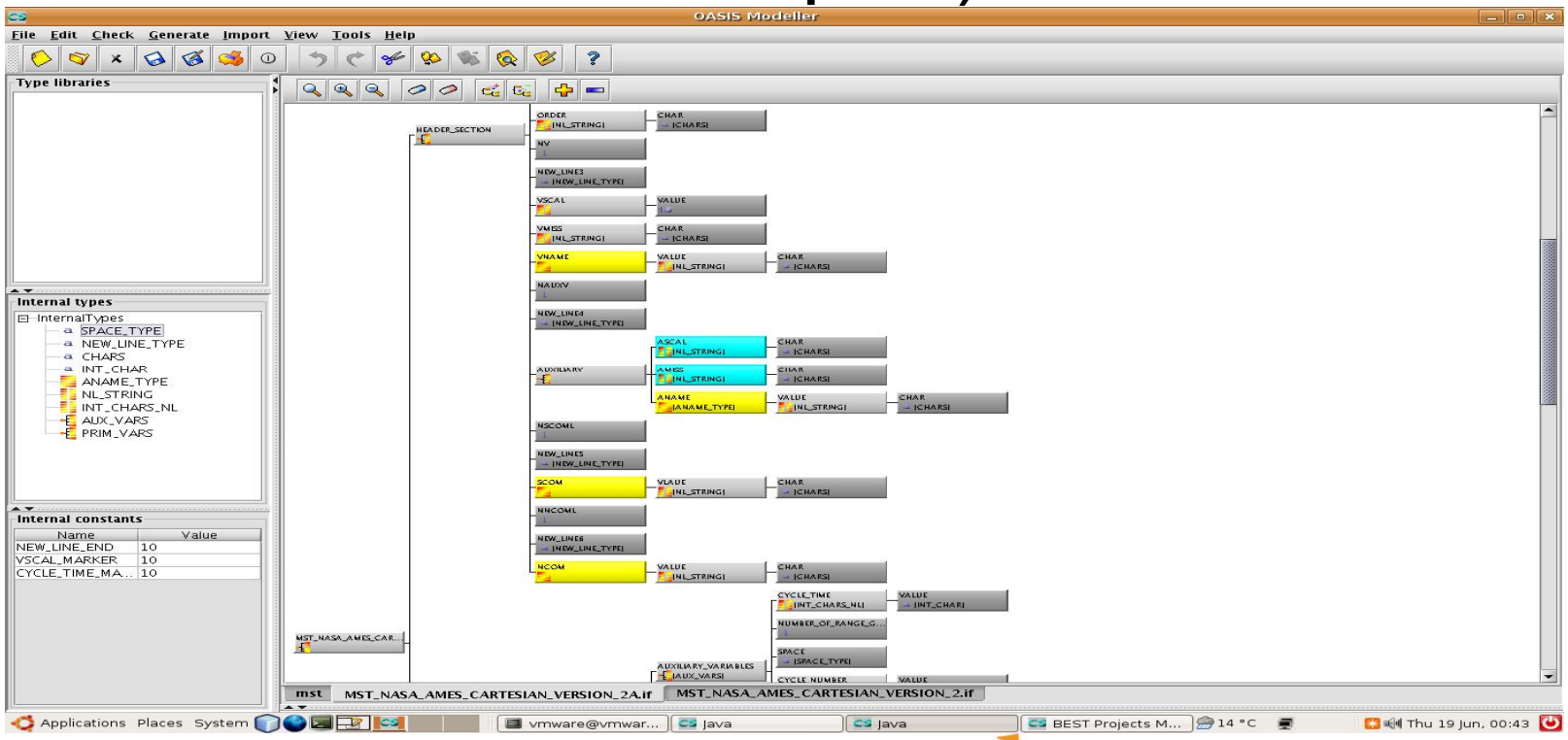
planets

CASPAR

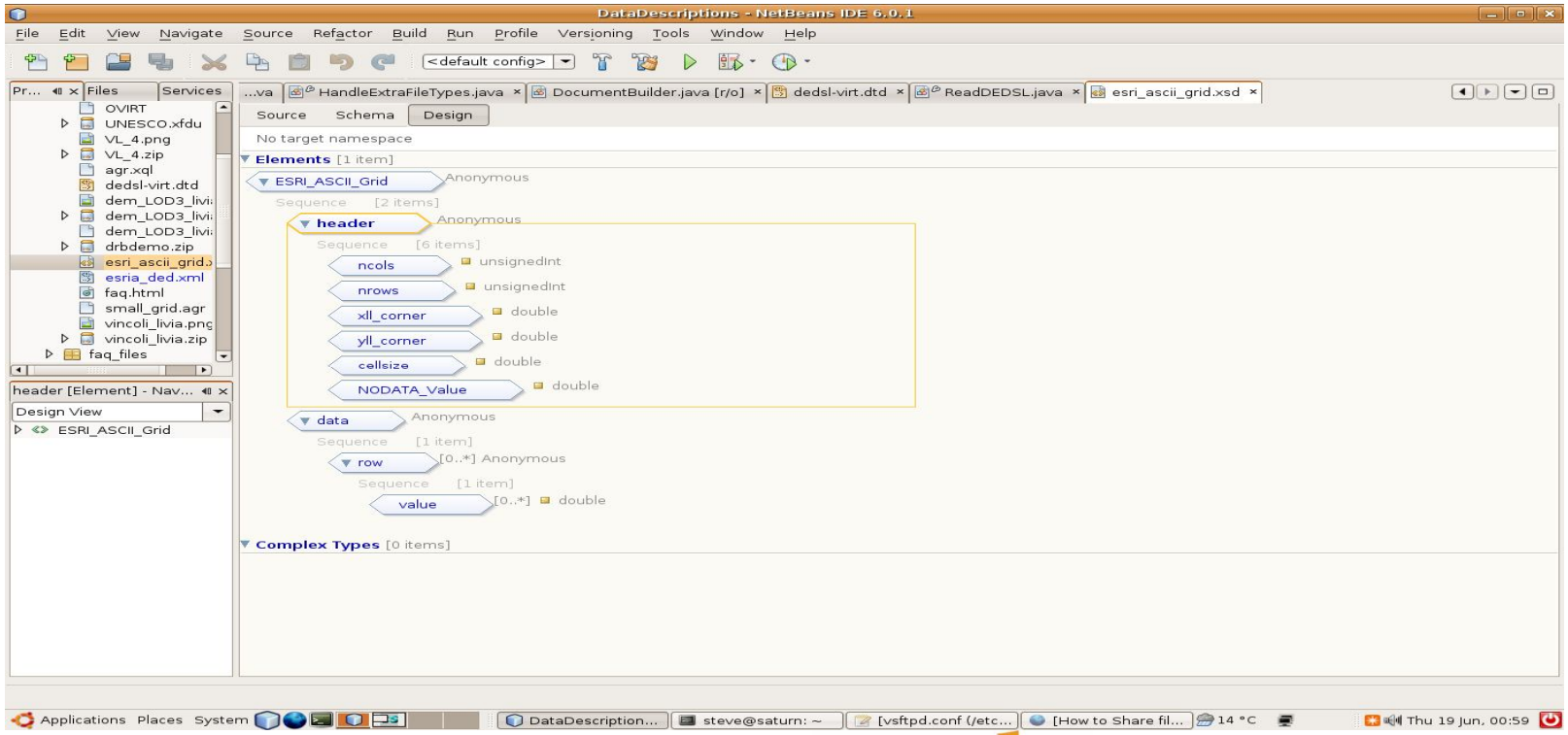

Old Software
Old Software
Old Software
Old Software
Emulator
Virtualisation
Old Software
Operating System
Operating System
Operating System
Operating System


New Software
New Software
New Software
The Grid
Structure and Semantic API Virtualisation
Existing Software
New Software
RepInfo for Data Object
RepInfo for Data Object
RepInfo for Data Object
RepInfo for Data Object

Formal Descriptions of Structure (Tool Examples)



Formal Descriptions of Structure (Examples)



Formal Descriptions of Semantics

DEDSL Abstract, PVL, and XML(DTD) syntax for defining some simple data semantics.

- Only a small number of required attributes for a given data structure, NAME, DEFINITION, UNITS (conditional), *ENTITY*_TYPE (conditional), ENUMERATION_VALUES (conditional), TEXT_SIZE (conditional).
- You can define your own attributes.
- You can reuse definitions from other dictionaries.

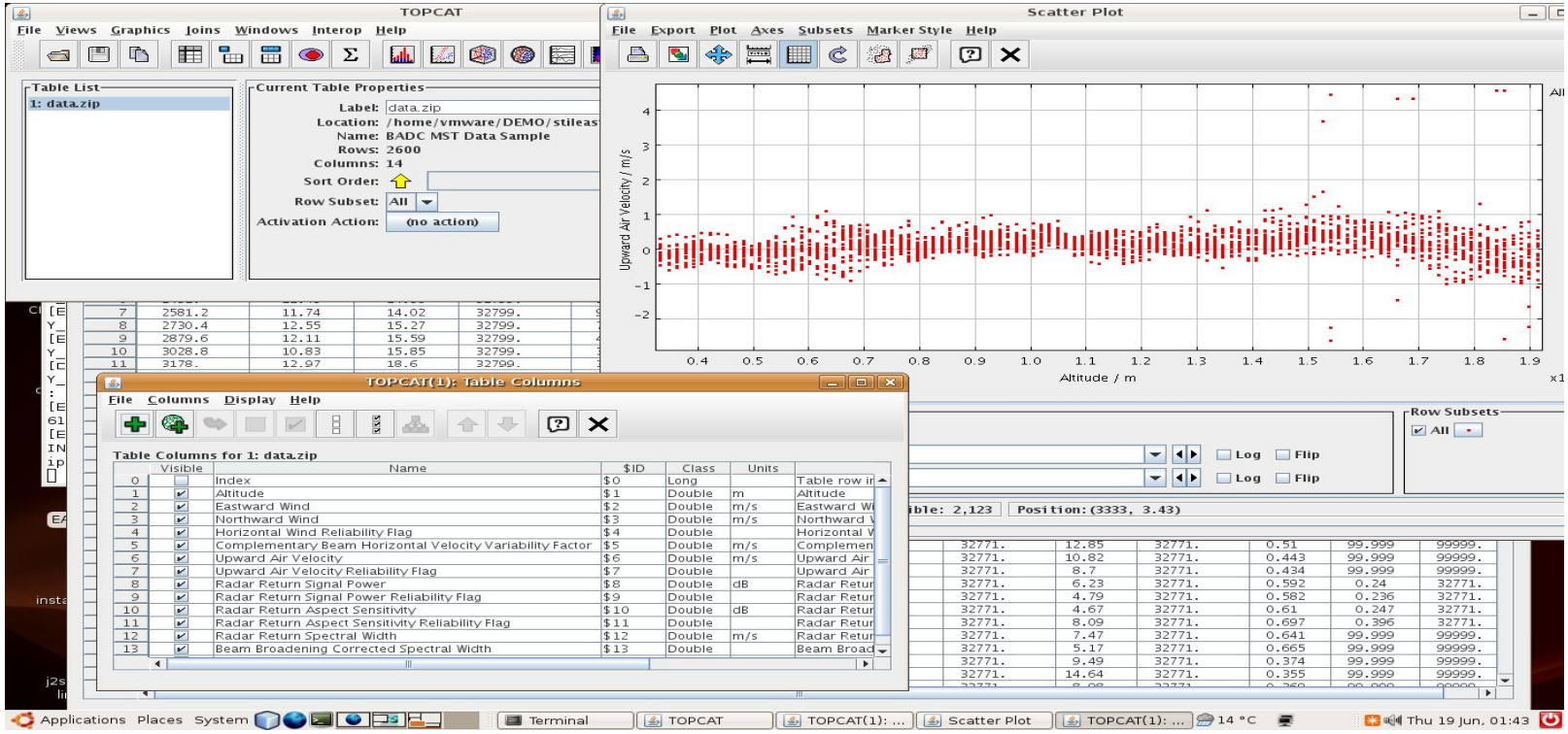
Link the data structures to the semantics via the EAST access path or an XPATH, i.e. define a new attribute – EAST_PATH (OASIS tool does this).

Data Object Virtualisation

- Just talked about simple semantics for data values.
- But data is typically a set of objects.
- Table is composed of columns and columns are composed of values.
- Semantics associated with the columns and with the table.
- You may also have collections of tables that are all related, and the relationships also have semantics.
- Current working is on a collections interface.

wePreserve

Table Data Example



Information Society

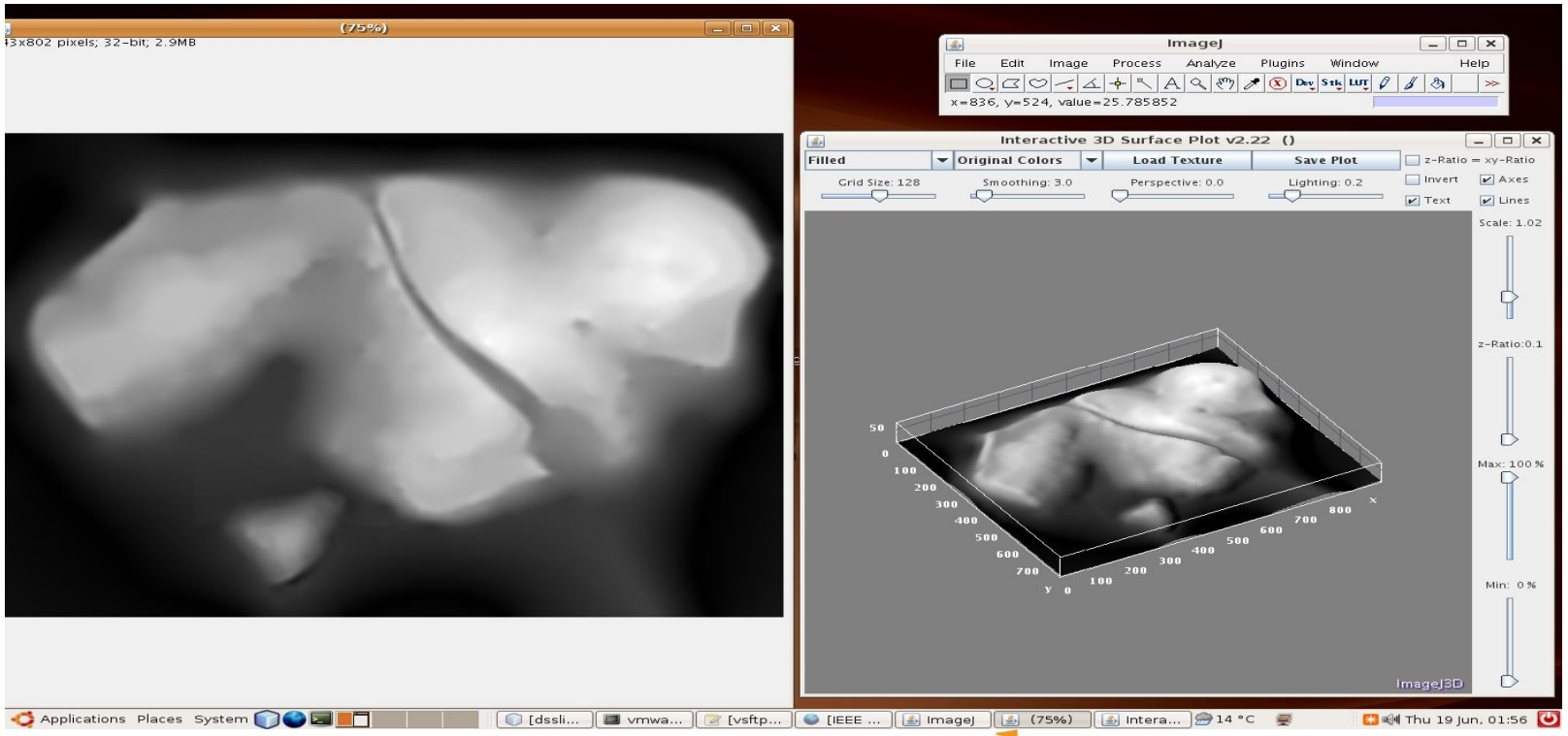
digital preservation europe

planets

CASPAR

wePreserve

Image (GRID) Data Example



digital preservation europe

planets

CASPAR

Information Society

CASPAR/DCC RepInfo Tools Training Material

- Video tutorials of using the RepInfo GUI tool
- Documentation tutorial for using the RepInfo GUI tool.
- Brain Training Game for RepInfo
- Video tutorial of using RepInfo creation tools (creating structure and semantics).
- Packaging tutorial.
- More under construction – Programming guides etc.
- All very much under development but a preview for the material is available at: http://twiki.dcc.rl.ac.uk/bin/view/Main/TheStoryPlanAndPresentation
- Or we can come to you or you can come to us for a tutorial on our tools. You just need to ask!