

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

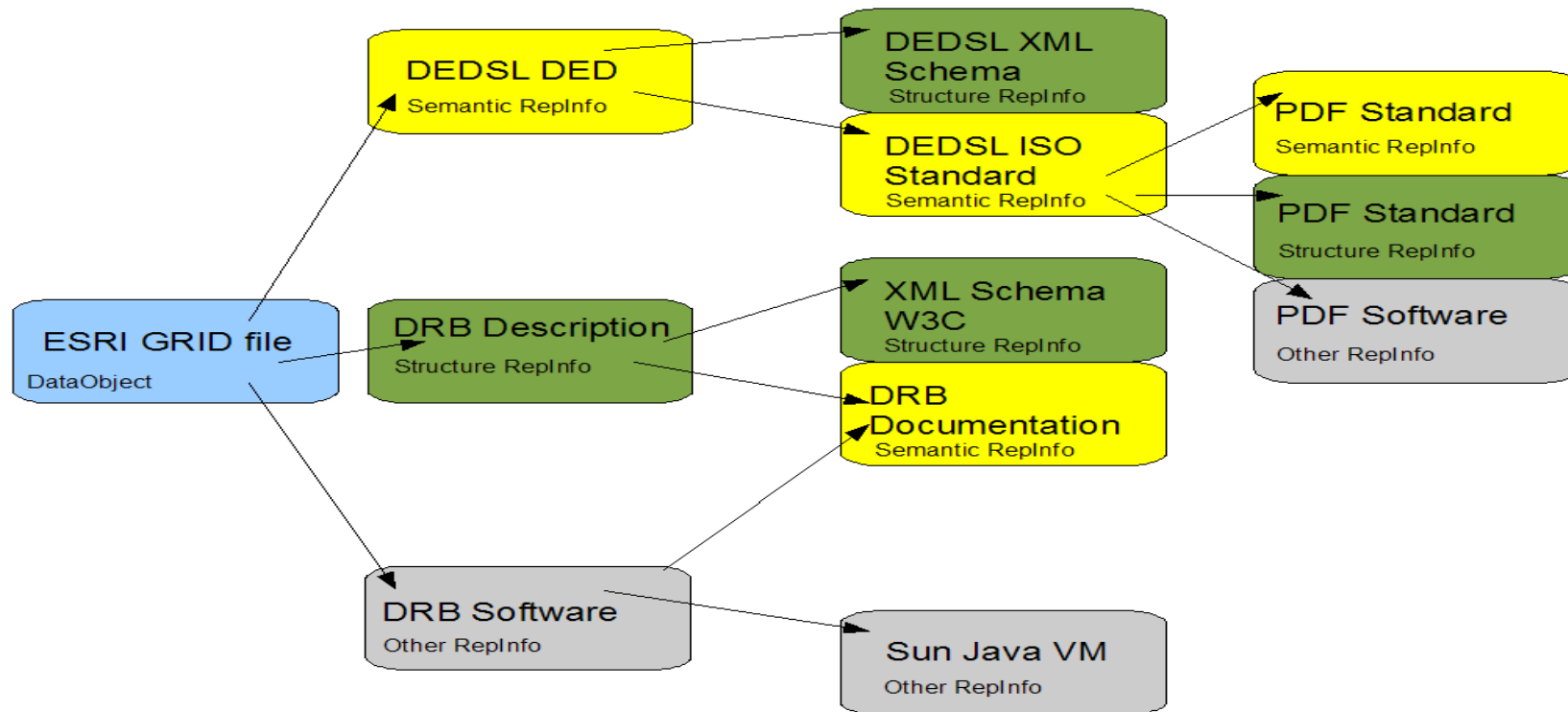
Stephen Rankin (STFC), Matt Dunckley
(STFC), Brian Mcilwrath (STFC), Esther
Conway (STFC), David Giaretta (STFC),
CASPAR & DCC

Outline

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

RepInfo Network Example

Elevation Grid Data – UNESCO Villa Livia



Validation with RepInfo

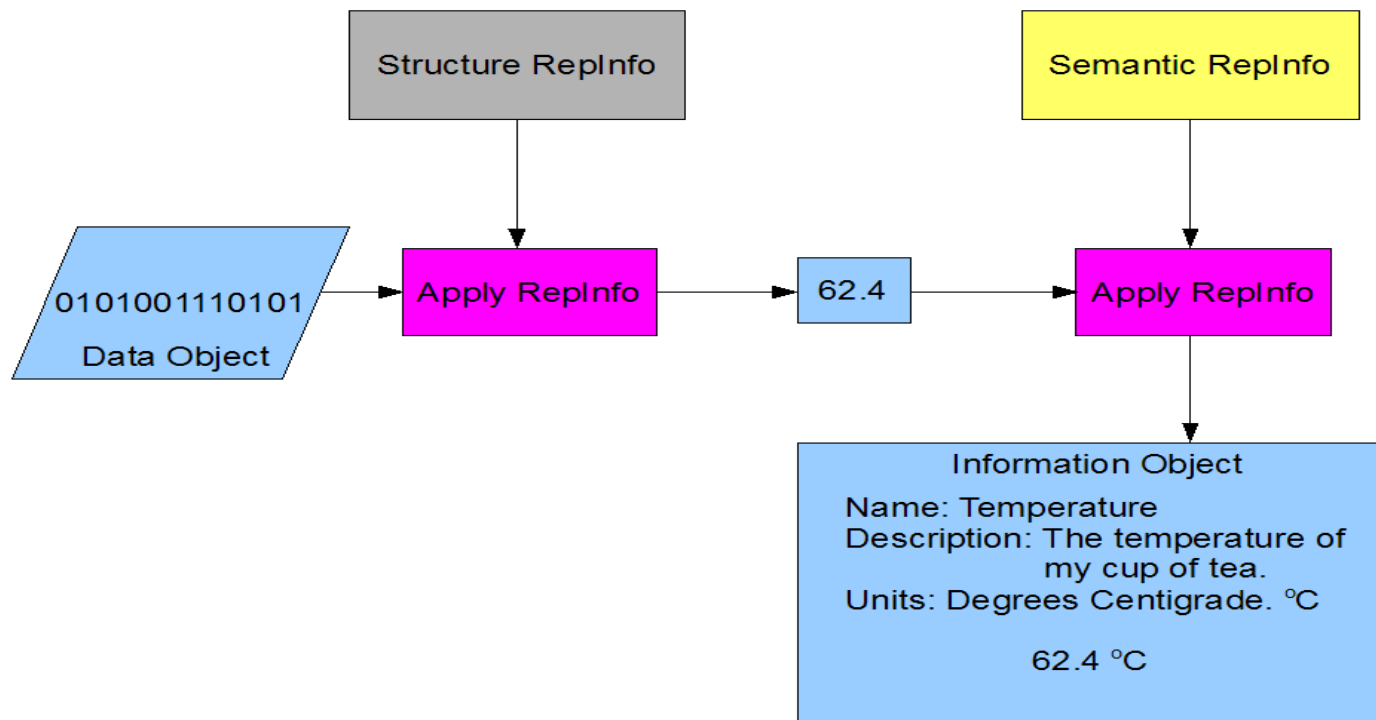
- Can a future user use the RepInfo and RepInfo network to reuse the data?
- Documents, software etc as RepInfo difficult to validate against the data.
Quality of RepInfo.

What do we mean by reuse the data?

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.

RepInfo for a Hot Cup of Tea



Creation of new Formal Representation Information

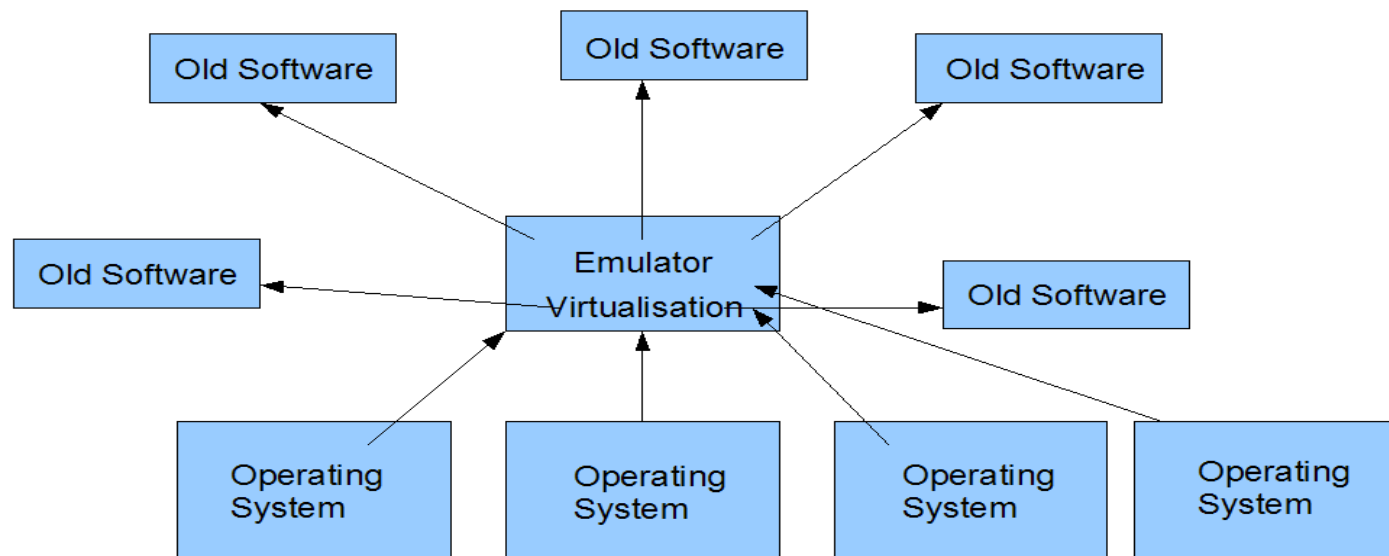
- Curating existing RepInfo is OK and necessary, but we can do more...
- Typically documents that describe the structure and semantics do not allow easy validation.
- Information in documents typically not machine readable.
- Semantics is usually poorly linked to the data and data values.
- The types 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.

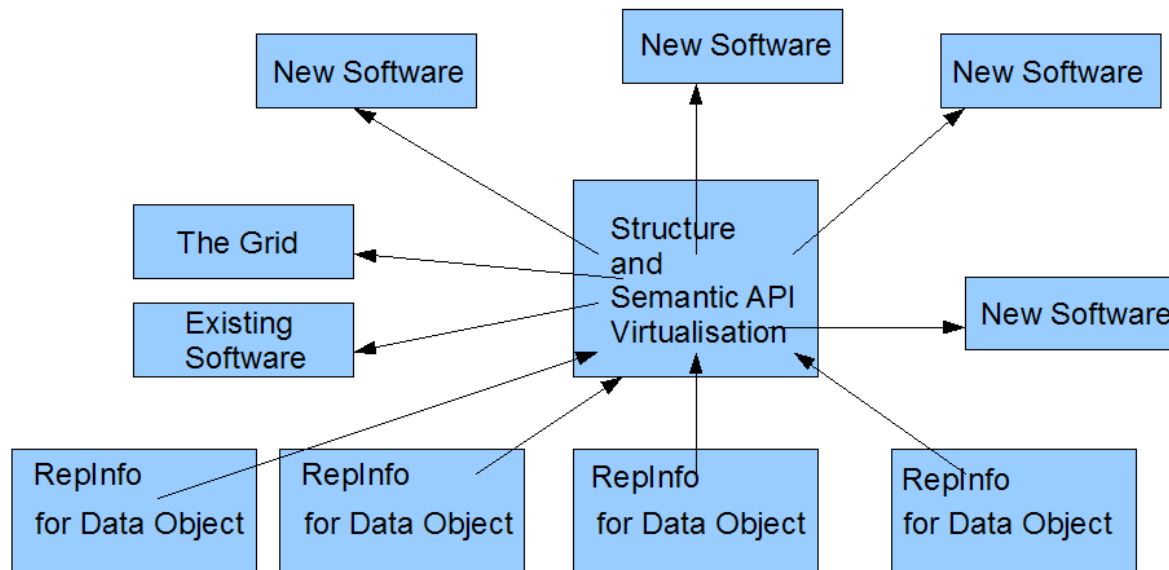
Formal Descriptions of Structure and Semantics

- 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.

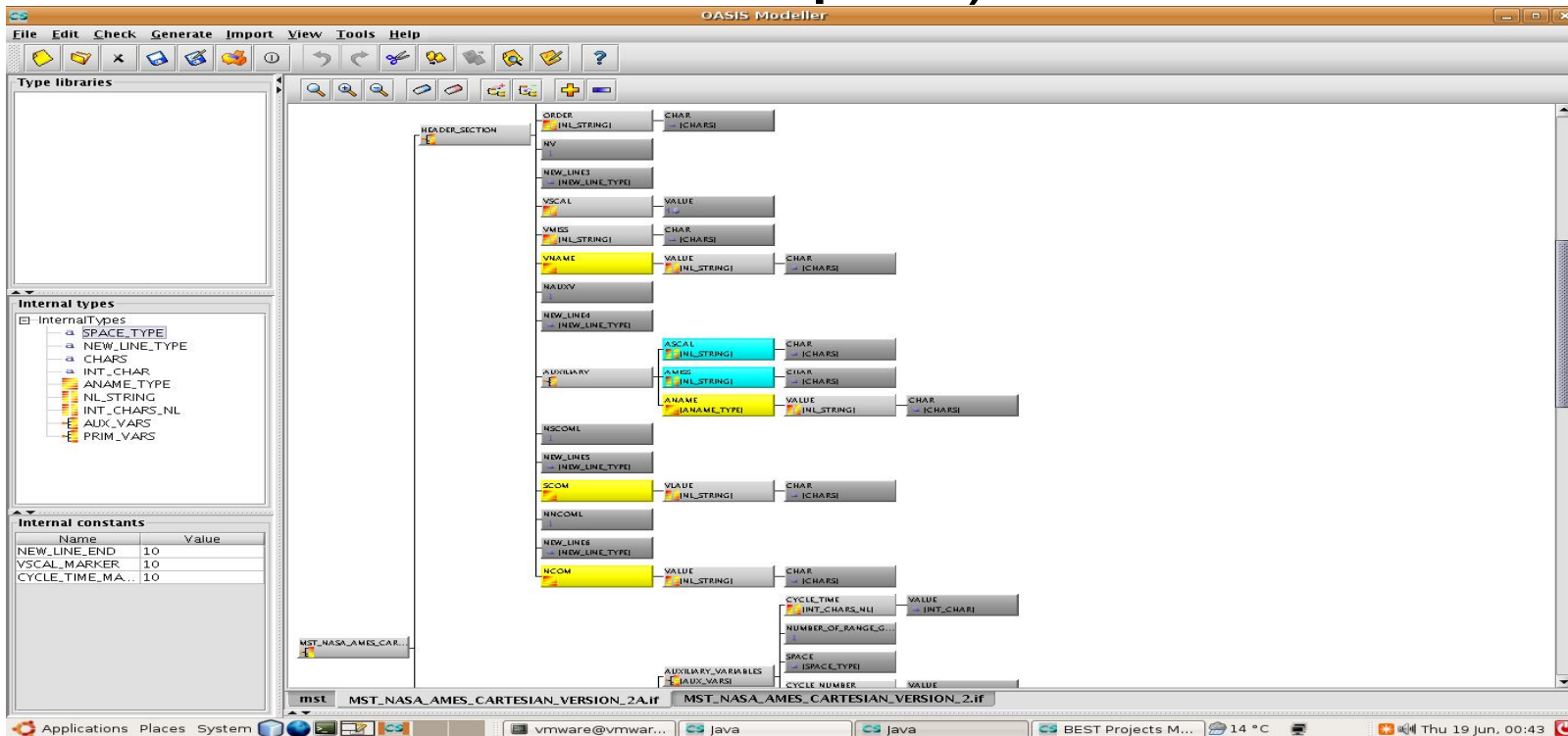
Formal Descriptions of Structure

- 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/>



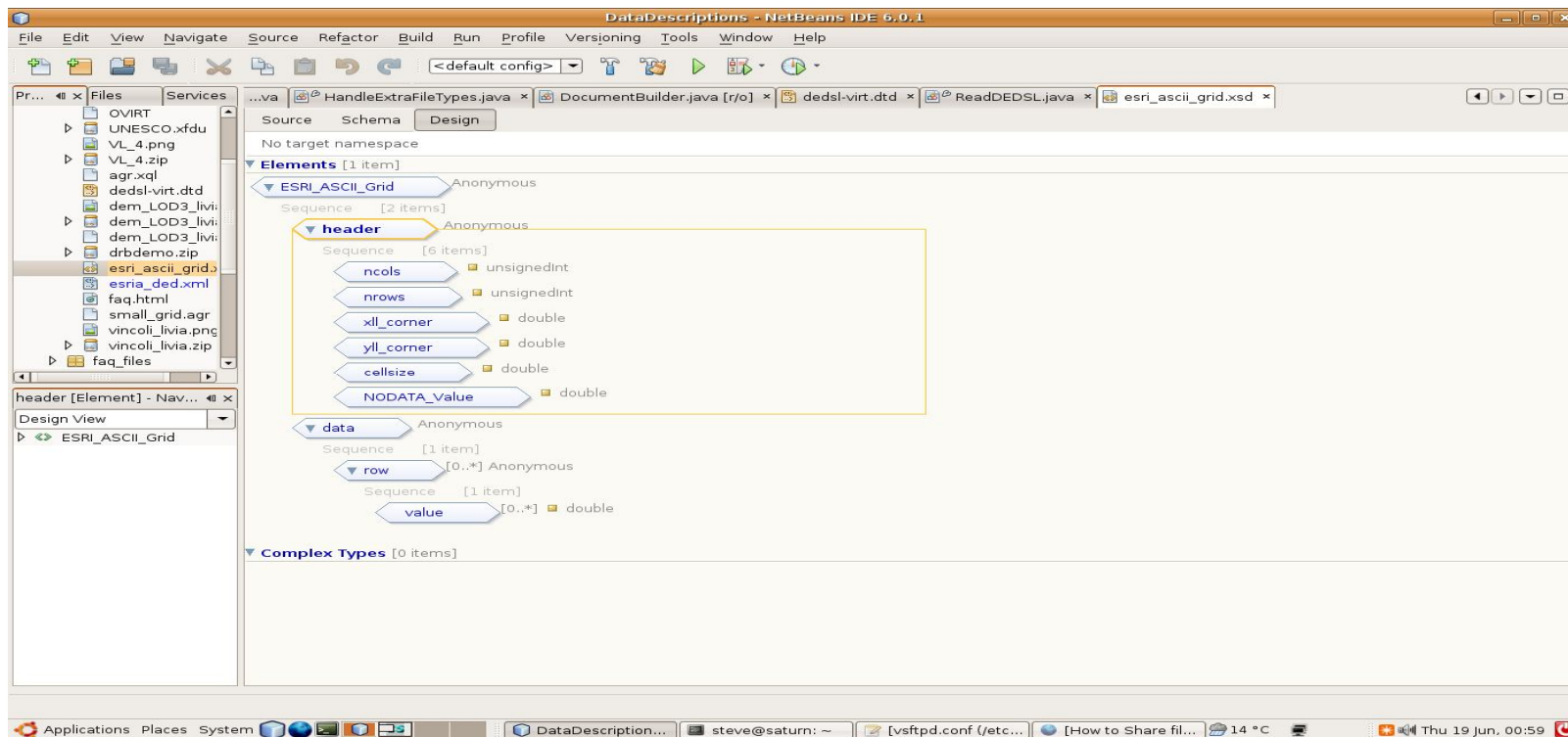


Formal Descriptions of Structure (Tool Examples)



The screenshot displays the OASIS Modeller software interface. The main window shows a hierarchical tree structure of formal descriptions. The tree includes nodes such as HEAD_SECTION, NV, NDV_LINES, VSCAL, VNAME, NADSV, NDV_LINES, AUXILIARY, NSCOML, NSCOM, NNCOML, NCOM, CYCLE_TIME, NUMBER_OF_RANGE_C, SPACE, and CYCLE_NUMBER. Each node is associated with specific data types and values, such as INL_STRING, CHAR, VALUE, and INT_CHARS_NL. The interface also features a menu bar (File, Edit, Check, Generate, Import, View, Tools, Help) and a toolbar with various icons. On the left side, there are panels for 'Type libraries', 'Internal types', and 'Internal constants'. The 'Internal types' panel lists various types like SPACE_TYPE, NEW_LINE_TYPE, CHARS, INT_CHAR, ANAME_TYPE, NL_STRING, INT_CHARS_NL, AUX_VARS, and PRIM_VARS. The 'Internal constants' panel shows a table with columns for Name and Value, listing constants like NEW_LINE_END, VSCAL_MARKER, and CYCLE_TIME_MA... with values of 1.0. The bottom status bar shows the current file being edited: MST_NASA_AMES_CARTESIAN_VERSION_2A.if and MST_NASA_AMES_CARTESIAN_VERSION_2.if.

Formal Descriptions of Structure (Examples)



The screenshot displays the NetBeans IDE interface for editing an XML Schema (XSD) file named 'esri_ascii_grid.xsd'. The Design View shows the following structure:

- Root Element: **ESRI_ASCII_Grid** (Anonymous)
- Sequence: **header** (Anonymous, 2 items)
 - Sequence: **header** (Anonymous, 6 items)
 - Element: **ncols** (unsignedint)
 - Element: **nrows** (unsignedint)
 - Element: **xll_corner** (double)
 - Element: **yll_corner** (double)
 - Element: **cellsize** (double)
 - Element: **NODATA_Value** (double)
 - Element: **data** (Anonymous)
- Sequence: **data** (Anonymous, 1 item)
 - Element: **row** (Anonymous, [0..*])
 - Sequence: **row** (Anonymous, 1 item)
 - Element: **value** (double, [0..*])

The 'Complex Types' section is currently empty.

Formal Descriptions of Semantics

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

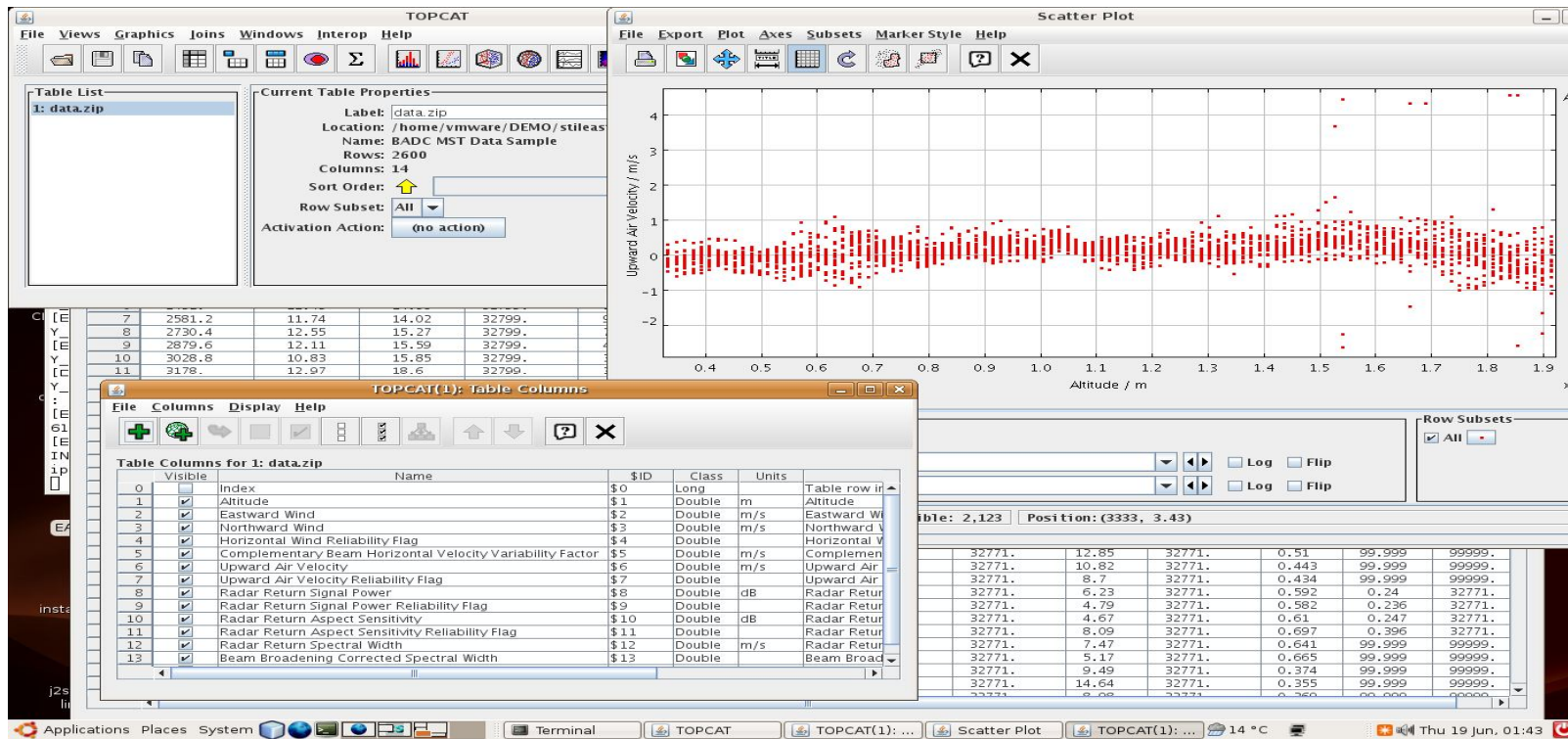
- Only a small number of required attributes for a given data structure, NAME, DEFINITION, UNITS (conditional), *ENTITY_TYPE* (conditional), ENUMERATION_VALUES (conditional), TEXT_SIZE (conditional).
- You can define your own attributes.
- You can reuse definitions from other dictionaries.

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

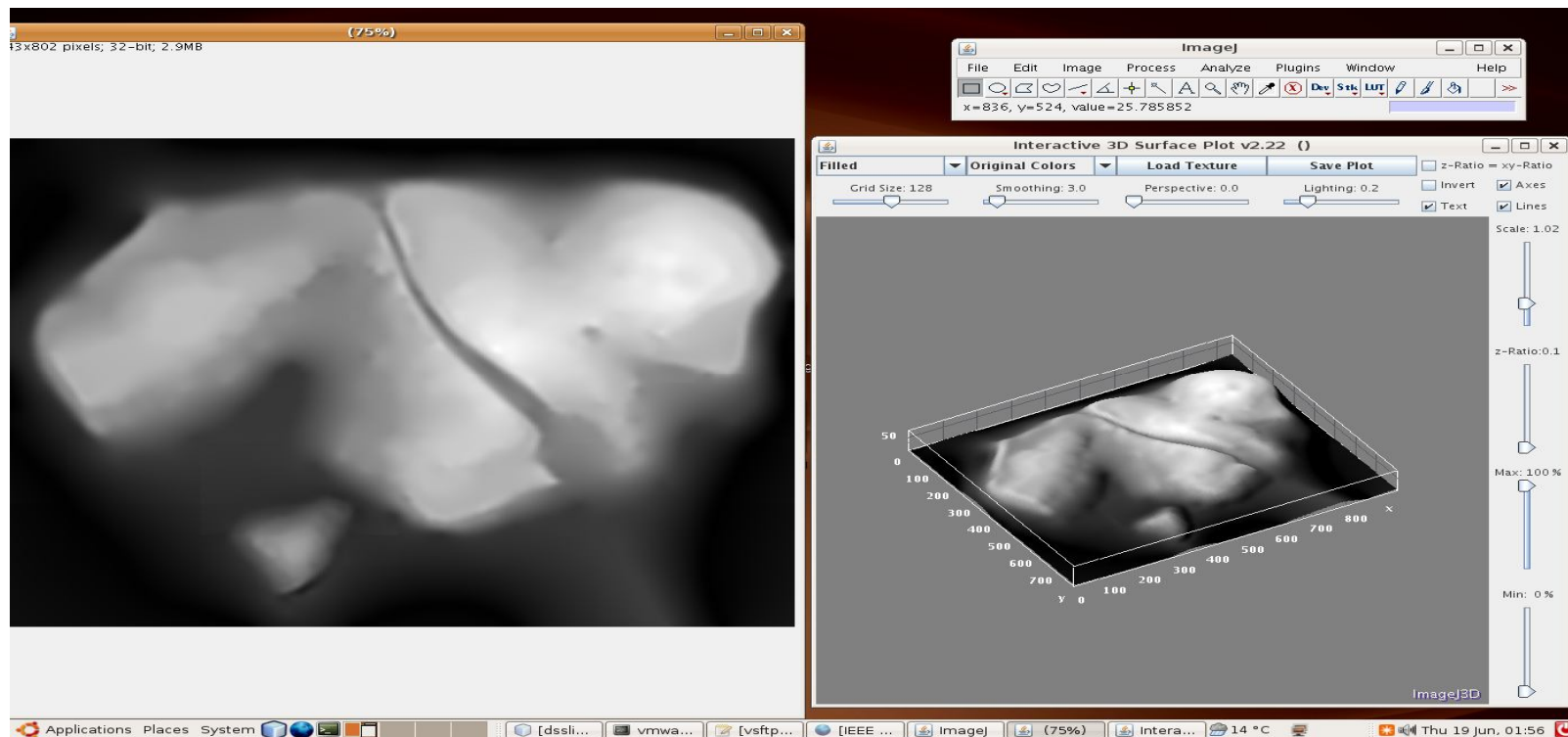
- Just talked about simple semantics for data values.
- But data is typically a set of objects.
- Table is composed of columns and columns are composed of values.
- Semantics associated with the columns and with the table.
- You may also have collections of tables that are all related, and the relationships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

Table Data Example



The screenshot displays the TOPCAT software interface. The main window shows a table with columns for Index, Altitude, Eastward Wind, Northward Wind, Horizontal Wind Reliability Flag, Complementary Beam Horizontal Velocity Variability Factor, Upward Air Velocity, Upward Air Velocity Reliability Flag, Radar Return Signal Power, Radar Return Signal Power Reliability Flag, Radar Return Aspect Sensitivity, Radar Return Aspect Sensitivity Reliability Flag, Radar Return Spectral Width, and Beam Broadening Corrected Spectral Width. A scatter plot titled 'Scatter Plot' shows 'Upward Air Velocity / m/s' on the y-axis (ranging from -2 to 4) and 'Altitude / m' on the x-axis (ranging from 0.4 to 1.9). The plot displays a dense cluster of red data points, indicating a strong correlation between altitude and upward air velocity. A 'Table Columns' dialog box is open, showing the list of columns and their properties. The system tray at the bottom indicates the date and time as 'Thu 19 Jun, 01:43'.

Image (GRID) Data Example



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!