Astro-WISE: Tracing and Using Lineage for Scientific Data Processing

Johnson Mwebaze, Danny Boxhoorn and Edwin Valentijn

Kapteyn Astronomical Institute
University of Groningen
The Netherlands

NBIS 2009 / 19-21 August / Indianapolis, Indiana, USA

Outline

Introduction

Scientific data doubles every year

- caused by successive generations of inexpensive sensors + exponentially faster computing

Changes the nature of scientific computing

It becomes increasingly harder to extract knowledge

The Virtual Observatory

the Sloan Digital Sky Survey (SDSS)[1],

the 2 Micron All Sky Survey (2MASS)[2]

the NRAO VLA Sky Survey (NVSS)[3].

the Low Frequency Array (LOFAR)[4], now under construction,

the soon to-be commissioned ESO VLT Survey Telescope (VST) in northern Chile, equipped with the large field of view CDD camera OmegaCAM[5],

and the upcoming space mission GAIA.

[1] Sloan Digital Sky Survey, http://www.sdss.org

[2] 2MASS project, http://pegasus.phast.umass.edu/

[3] The NRAO VLA Sky Survey,http://www.cv.nrao.edu/nvss/

[4] The LOFAR (Low Frequency Array) project

[5] Wide Field Imager for the VST, http://www.astro.rug.nl/ omegacam

The core of the system exploits three properties in a database environment.

Firstly, we apply the principle of inheritance using Object Oriented Programming (Python), where all Astro-WISE objects inherit key properties for database access, such as persistency of attributes.

Secondly, the linking (associations or references) between instances of objects in the database is completely maintained.

Thirdly, continuous growth of the database through the addition of new information or improvements made to existing information.

Database Architecture and Persistence Objects

core of Astro-WISE pipeline processing and data lineage

Processing is done through invocation of methods on these objects. Since these objects are persistent, all operations and attributes of these objects are automatically saved into a database.

All the I/O of the pipeline processes makes use of a database

The database stores all persistent objects, attributes and raw data either as fully integrated objects or as descriptors

Data can therefore only be manipulated through interaction with the database allows extending and modifying object attributes and method definitions through inheritance and polymorphism

Users can modify functionality in modules, insert them into the system, or add a module on top of what currently exists, as long as these modules obey the standard data model.

Version Control

Preservation Management

Tracing Lineage in Astro-WISE

Tracing Lineage in Astro-WISE

Tracing Lineage in Astro-WISE

Tracing Lineage in Astro-WISE

Tracing Lineage in Astro-WISE

Experiments [Case Studies]

Conclusions and prospects

???? (cont)

Conclusion and Prospects

?