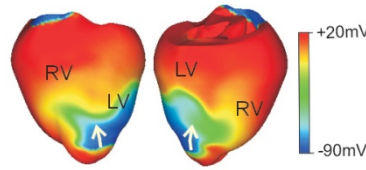
An Approach to Software Preservation

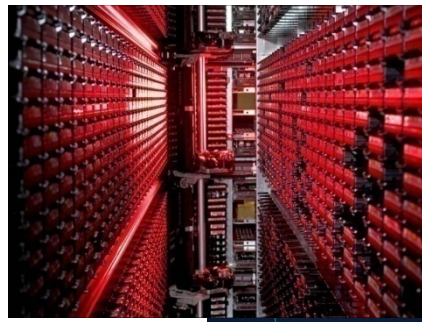
PV 2009, Madrid

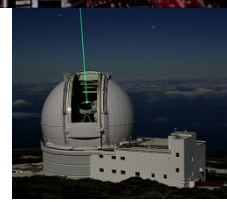
Arif Shaon, Brian Matthews, Juan Bicarregui, Catherine Jones (STFC), Jim Woodcock (Univ of York)

1 December, 2009

Science and Technology Facilities Council







- Provide large-scale scientific facilities for UK Science
  - particularly in physics and astronomy
- E-Science Centre – at RAL and DL
  - Provides advanced IT development and services to the STFC Science Programme
  - Strong interest in Digital Curation of our science data
  - Keep the results alive and available
  - R&D Programme: DCC, CASPAR

Long-term Preservation of Software

- JISC funded work: Tools & Guidelines for the preservation of software as a research output
  - Used the JISC funded: Significant Properties of Software Report
- Software very large topic
  - Diversity in: *application of software* and *software architecture* and *scale of software* and *provenance* and *user interaction*
- Project needed to limit scope
  - Scientific and mathematical software
  - Limited commercial consideration
  - Limit consideration of user interaction
- Finding information
  - Literature, Standards (e.g. the OAIS Reference Model)
  - Case Studies: Talking to developers of products and software repositories
- Developing a framework for software preservation.

Software Preservation

- What is software preservation?
  - Storing a copy of a software product
  - Enabling its retrieval in the future
  - Enabling its reconstruction in the future
  - Enabling its execution in the future

Not what most software developers and maintainers do.

Why Preserve Software ?

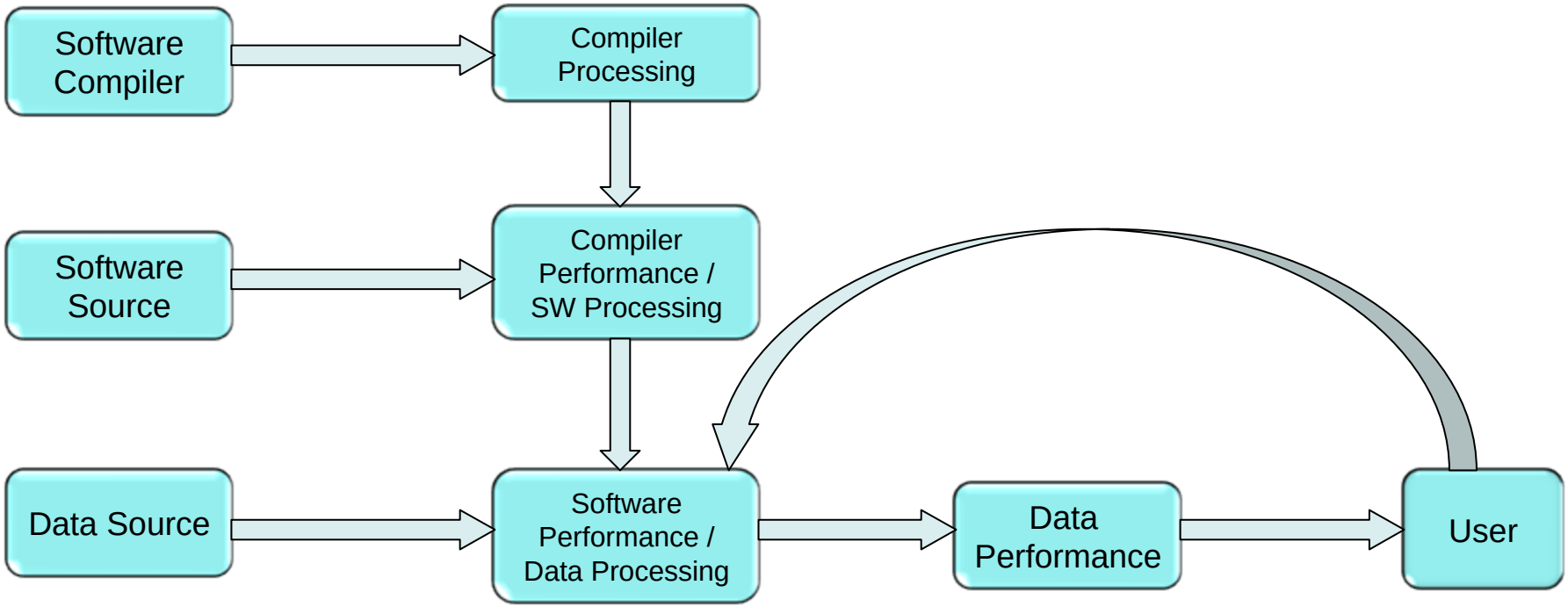
- Preserving the Data
  - Preserving the software is necessary to preserve other data
  - Keep the data live and reusable
  - Prime motivation for STFC
- Preserving the work
  - E.g. research work in Computing Science
  - Reproducible
- Handling Legacy
  - Specialised code from the past which still needs to be used
  - Usually seen as a problem!
- Museums and archives:
  - Either supporting Hardware
    - E.g. Bletchley Park, Science Museum,
  - Or in its own right
    - Chilton Computing, Multics History Project

A Conceptual Framework for Long-term Software Preservation

Three aspects to the framework:

- A Performance Model for software
  - Determine what it means to preserve s/w
  - Adequacy of performance of s/w
  - Based on the NAA performance model for digital preservation
- Model for describing s/w artefacts
  - As complex digital objects.
  - Versions and variants
- Properties for preservation
  - For retrieval, reconstruction, replay

Performance Model for Software



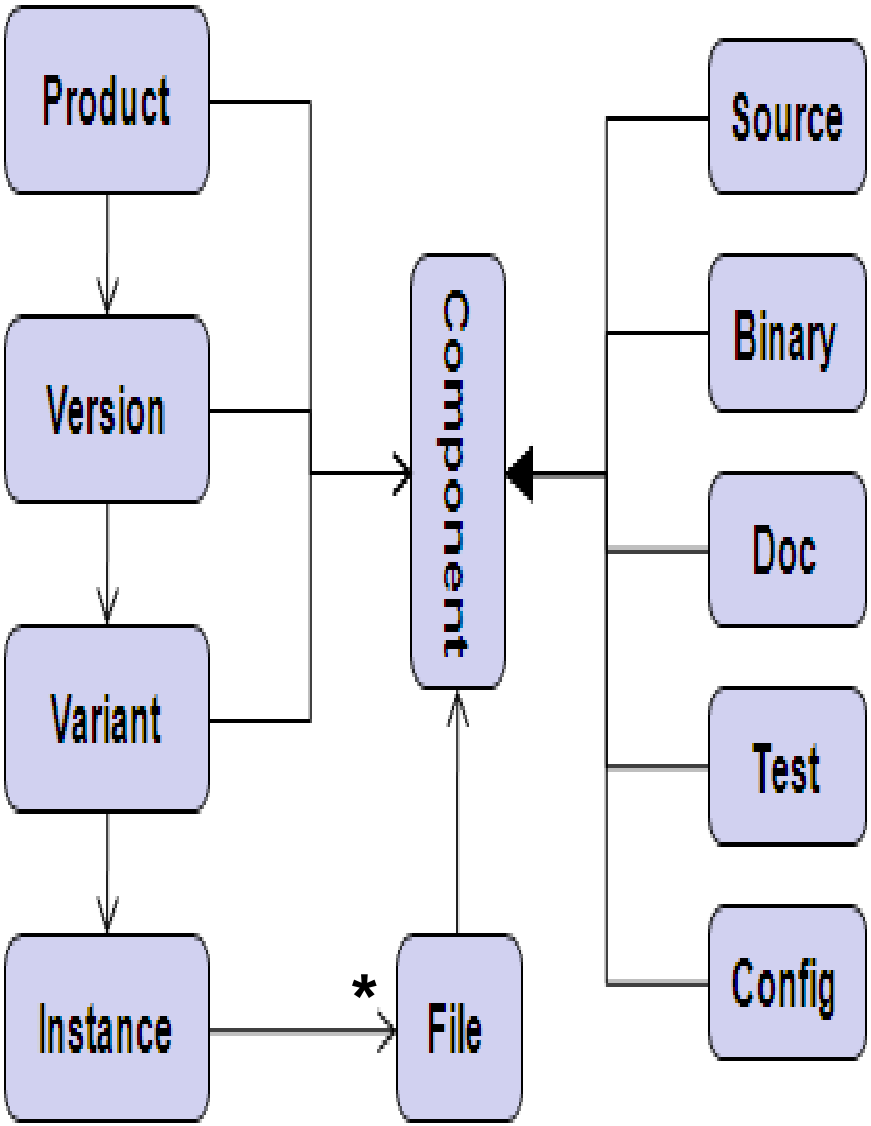
- Testing data performance to judge ***adequacy*** of the software performance.
- Important to maintain software test suite to assess preservation of significant properties of the software.

Adequacy of Software Preservation

A software package can be said to perform adequately relative to a particular set of "significant properties", if in a particular performance it preserves those significant properties to an acceptable tolerance.

- Significant properties are evaluable features of the performance
- After the recall and reconstruction phase
- Assesses the value of the replay
- ***Can be generalised to any digital object***

A Conceptual Model for Software

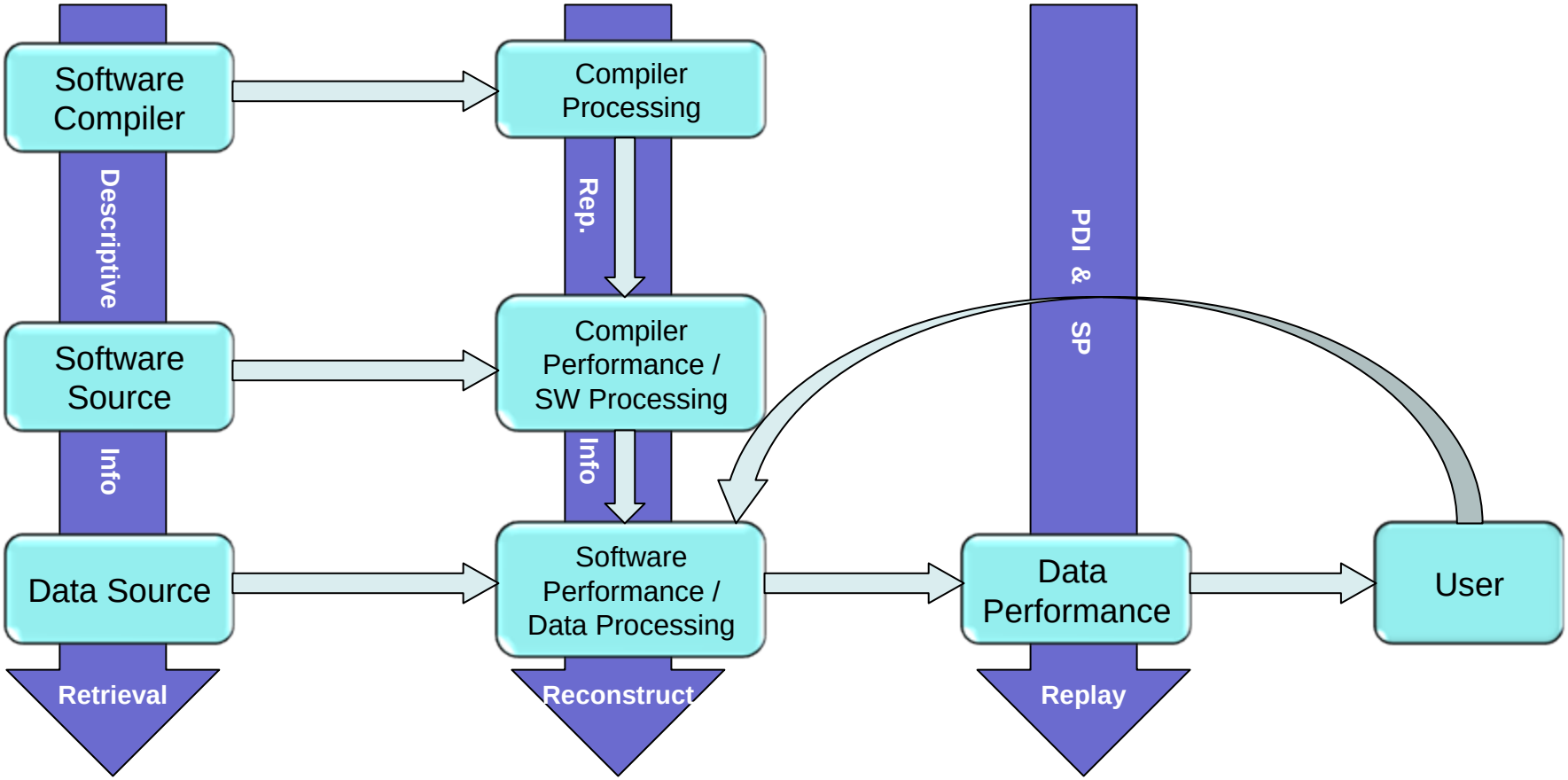


- Product
  - The whole software object under consideration
  - Could be single library module, or very large system (e.g. Linux)
  - Comes under one “authority” (legal control)
  - Defines “gross functionality”
- Version
  - Releases of the system
  - Characterised by changes in detailed functionality
- Variant
  - Versions for a particular platform
  - Characterised by operating system and environment
- Instance
  - A particular instance of a particular variant at a particular location
  - Ownership
  - An individual licence
  - Fixed to particular MAC or IP address, URLs etc.

Preservation Properties of Software

- What attributes of software do we need to take into account for long-term preservation?
  - Functionality
    - what it does and what data it depends on
  - Environment
    - platform, operating system, programming language
    - versions
  - Dependencies
    - Compilation dependency graph
    - Standard libraries
    - Other software products
    - Specialised hardware
  - Software is a Composite digital object
    - Collection of modules
    - Specifications, Configuration scripts, test suites, documentation
  - Architecture
    - Client/server, storage system, input / output
  - User interaction
    - Command line, User Interface
    - User model

Relationship to the OAIS model



- Open Archival Information System (OAIS) – ISO standard for the preservation of digital object.
- Software preservation properties are related to concepts in OAIS.

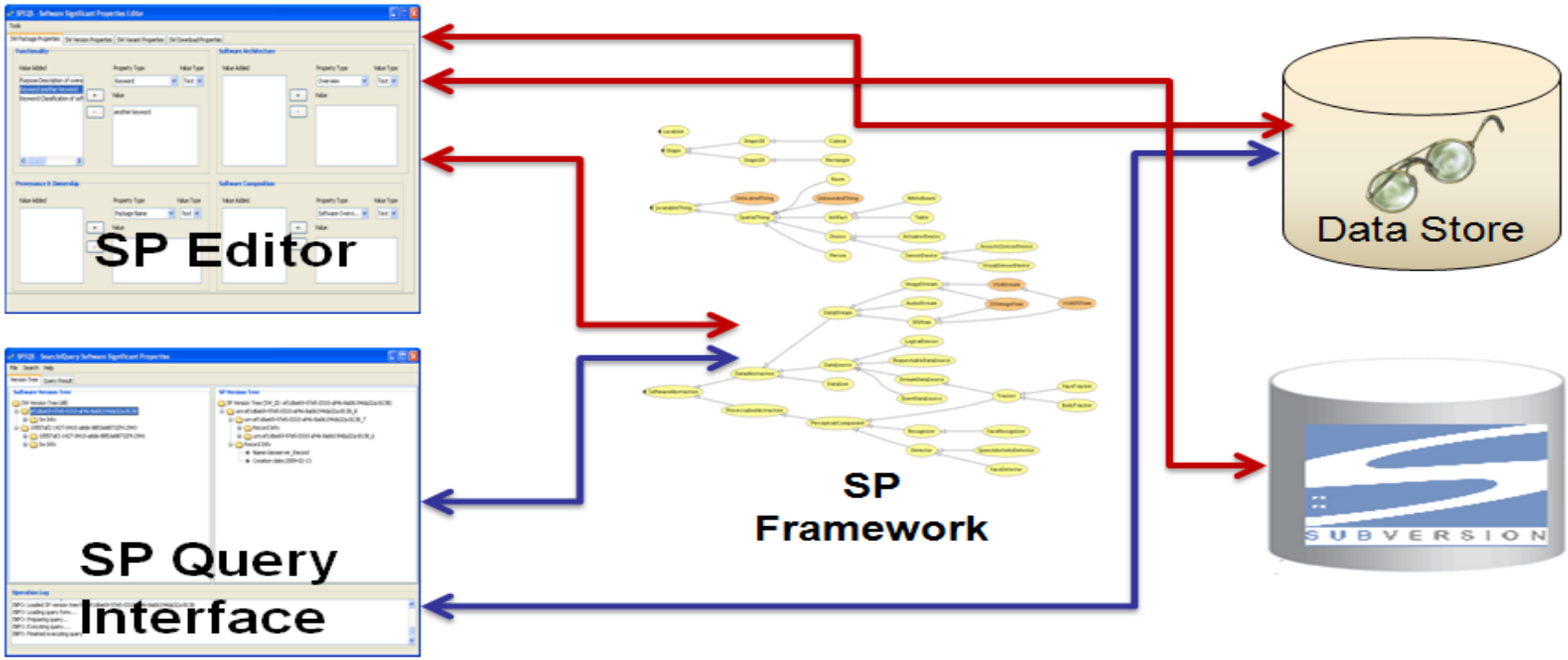
The BADC Case Study (1)

- The British Atmospheric Data Centre (BADC)
  - A NERC Designated Data Centre
  - Hosts over 250TB of atmospheric data for UK scientists and researchers
  - Also develops, supports, and provides access to a variety of software to facilitate accessibility and usability of data
  - Examples of BADC software: Trajectory Service, Weather Generator
- The BADC approach to Software Preservation
  - Long-term preservation is out of the current operational remit
  - Considers the high recurring costs of preservation as a prohibitive factor
  - Requires preservation solutions that could be integrated into the existing software management infrastructure

The BADC Case Study (2)

- Evaluating the preservation framework against some BADC software artefacts
  - Involved recording values for different preservation properties defined in the framework
  - The final result validates the relevance and adequacy of the framework
  - However, highlights that clear understanding of both the framework and different aspects of the software is also needed
  - Underlines the need for suitable tools with sufficient guidelines

Significant Properties Editing and Querying for Software (SPEQS)



- Java-based Eclipse plug-in; enables capturing software preservation properties during its development
- Demonstrates the concept of preservation tools that could be integrated within existing software development systems
- Used in CASPAR project

Summary

- Exploration of the s/w preservation space
- Defined reasons, audience, some basic concepts
- Defined a framework which enables s/w to be included in OAIS preservation framework
- Fits in a OAIS compatible preservation methodology
- Validated in some practical scenarios

Questions?

http://sigsoft.dcc.rl.ac.uk/twiki/bin/view

http://www.e-science.stfc.ac.uk/projects/software-preservation/softpres8985.html