

# VAMDC: atomic data production and curation in data-intensive e-science

- Claudio Mendoza
- Juan González
- Franck Delahaye
- Claude J. Zeippen
- Luis A. Núñez



# Virtual Atomic and Molecular Data Center

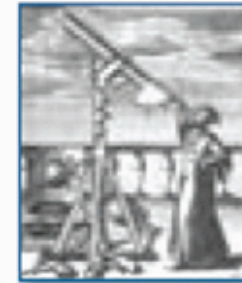
- VAMDC aims at building an interoperable e-infrastructure for the exchange of atomic and molecular data. VAMDC involves 15 administrative partners representing 24 teams from 6 European Union member states, Serbia, the Russian Federation and Venezuela.
- VAMDC is supported by EU in the framework of the FP7 "Research Infrastructures - INFRA-2008-1.2.2 - Scientific Data Infrastructures" initiative. It started on the 1st of July for a duration of 42 months.



# E-science is collaborative data-intensive science

## Science Paradigms

- Thousand years ago:  
science was **empirical**  
*describing natural phenomena*
- Last few hundred years:  
**theoretical** branch  
*using models, generalizations*
- Last few decades:  
a **computational** branch  
*simulating complex phenomena*
- Today: **data exploration** (eScience)  
*unify theory, experiment, and simulation*
  - Data captured by instruments or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files using data management and statistics



$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

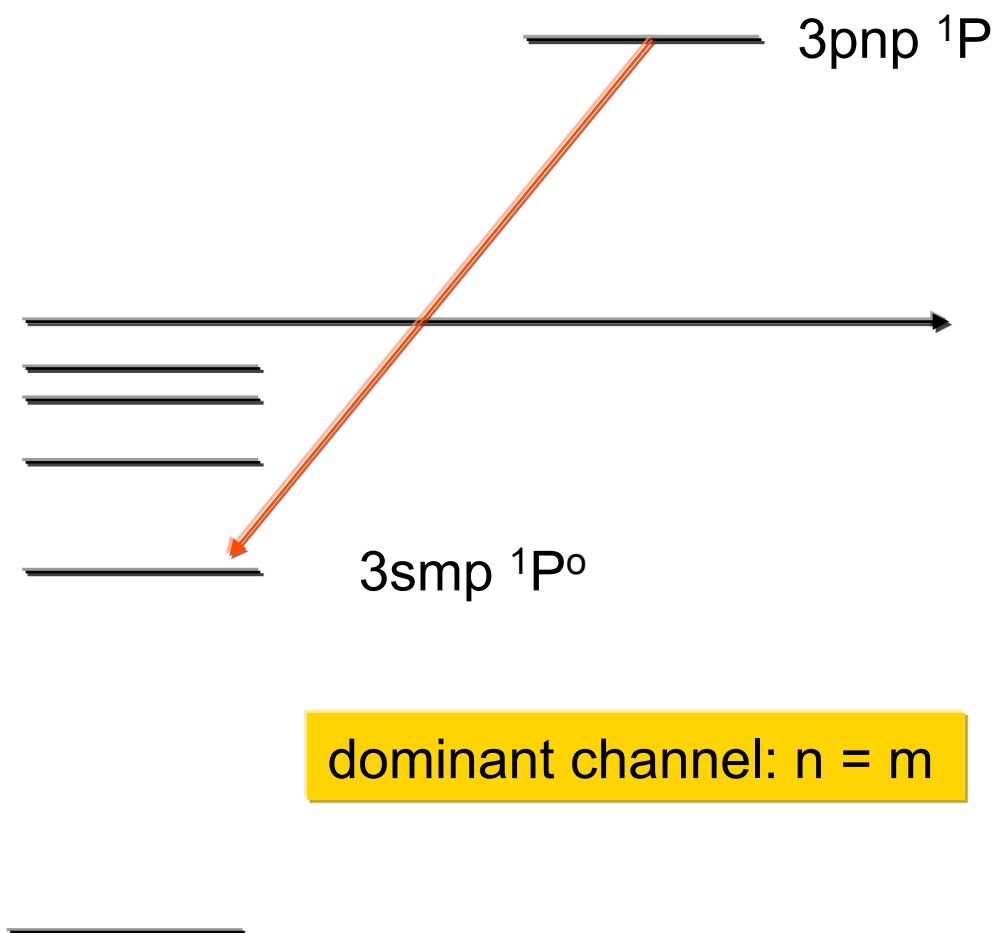


Original image from Hey, Tansley & Tolle (2009)

# Radiative decay of 3pnp $^1P$ states in Mg-like ions

Mg I

State	RLT (ns)
3p4p $^1P$	3.47E+00
3p5p	3.67E+00
3p5p	3.72E+00
3p6p	3.73E+00
3p7p	3.74E+00
3p8p	3.75E+00
3p9p	3.78E+00
3p10p	3.87E+00



Butler et al (1990)

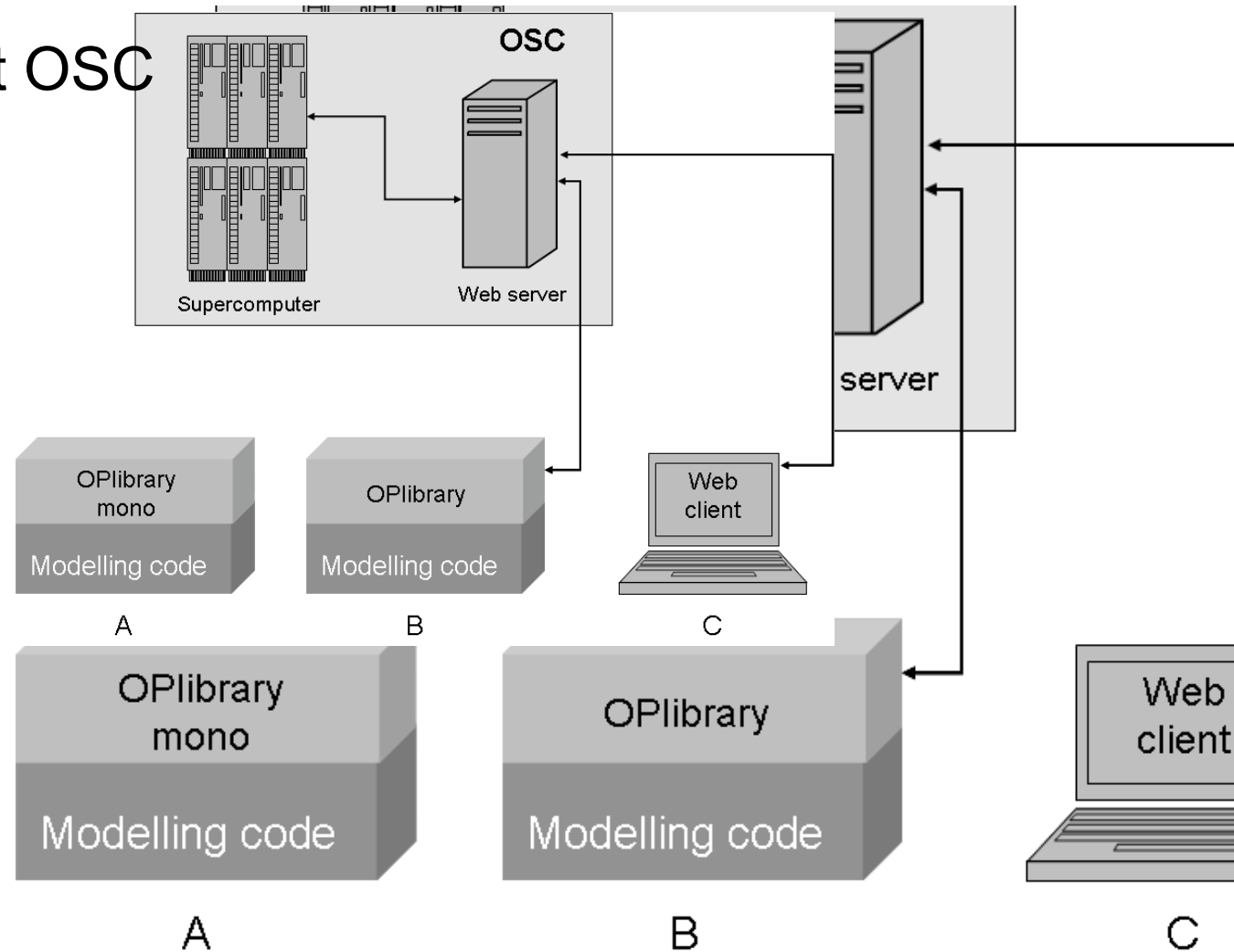
# Application and data access in e-science

Application and data access in e-science is evolving:

- ❖ Command-line based
  - Unix shell
- ❖ Web interface
  - Browser
- ❖ Distributed web services
  - Scripts
  - Workflow systems

# Application and data access in e-science

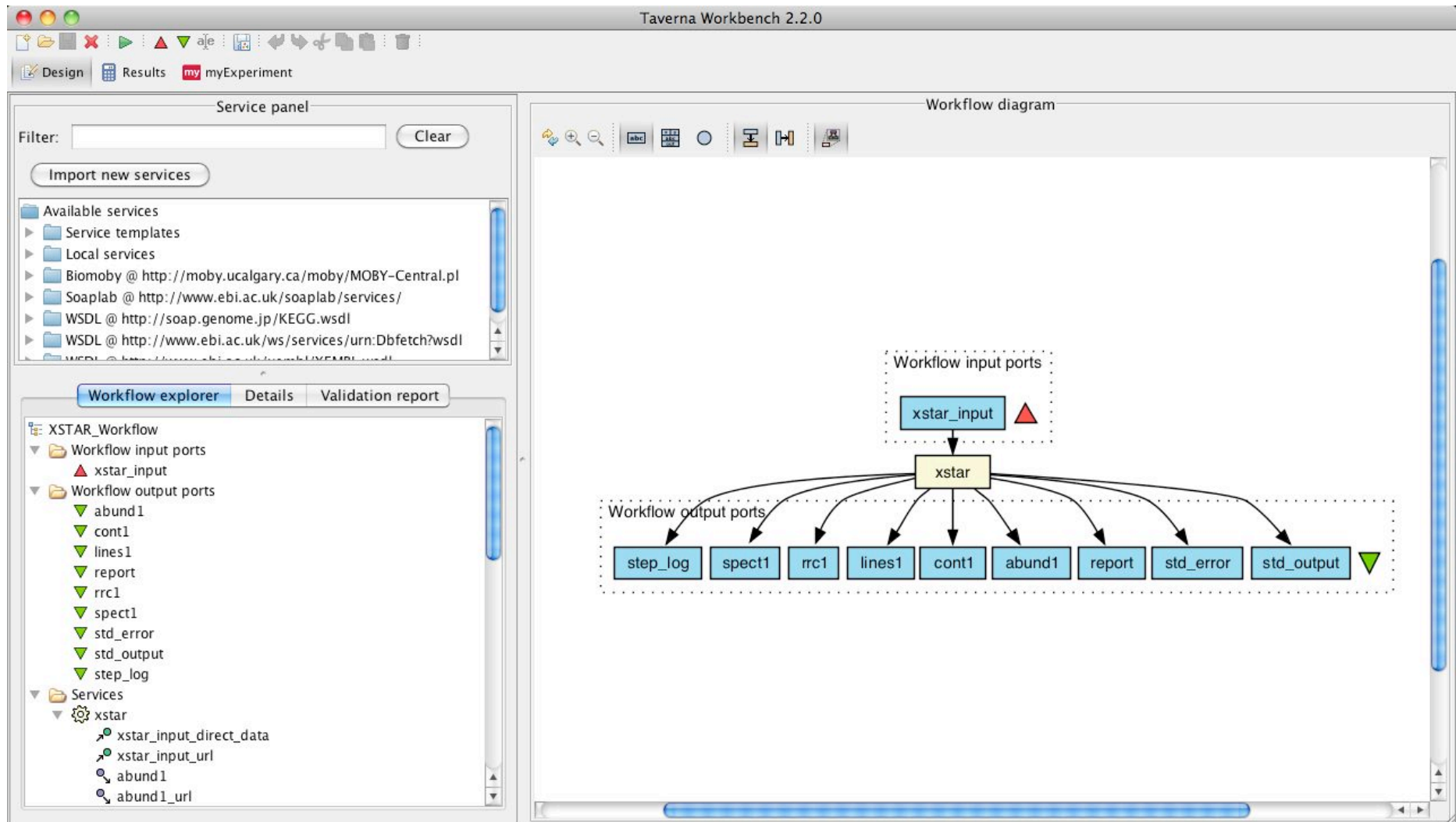
OPserver at OSC



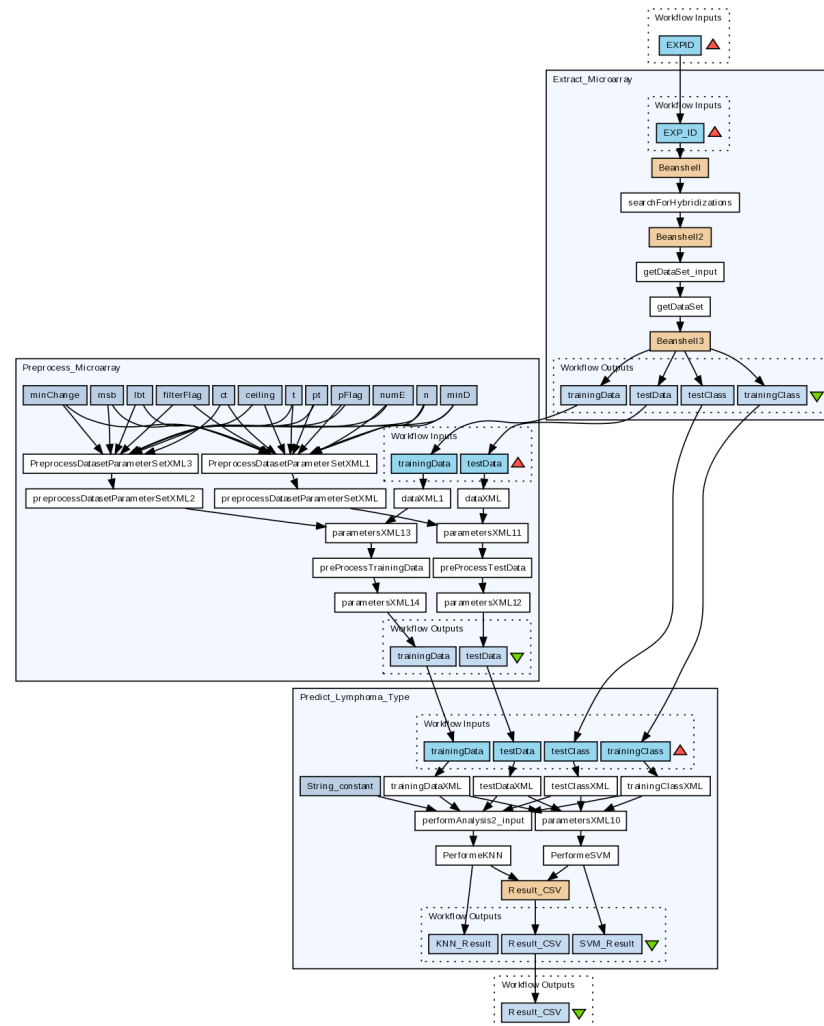
From Mendoza et al (2007)



# Workflow systems



# Workflow systems





# Workflow systems

Taverna Workbench v1.7.1.0

Design Results T2 Activity palette preview Discover Taverna 2 preview my myExperiment (beta)

Example Workflows Latest Workflows Search Workflows Tags Browser

10 example workflows found Refresh

Open in myExperiment

Preview Open

**Fetch Dragon images from BioMoby (version 2)**  
Uploader: Alan Williams

Fetch images and annotations of snapdragons

Open in myExperiment

Preview Open

**Demonstration of configurable iteration (version 1)**  
Uploader: Alan Williams

This workflow shows the use of the iteration strategy editor to ensure that only relevant combinations of inputs are used during an implicit iteration.

Open in myExperiment

Preview Open

**Retrieve sequence in EMBL format (version 2)**  
Uploader: Alan Williams

This workflow retrieves a sequence associated with its features in embi format given a sequence id.

Open in myExperiment

Open Import into current workflow

**Workflow Preview**

Workflow information found. Last fetched: Mon Sep 01 10:23:23 BST 2008 Sync Clear Refresh

**Workflow Entry: Fetch Dragon images from BioMoby (version 2)**

Uploader: Alan Williams  
Created at: Wed Mar 05 14:09:35 GMT 2008  
License: Creative Commons Attribution 3.0 License

```
graph TD
    id[id] --> Object[Object]
    namespace[namespace] --> Object
    Object --> getDragon[getDragonSimpleAnnotatedImages]
    getDragon --> getJpeg[getJpegFromAnnotatedImage]
    getJpeg --> ParseMobyDataJpeg[Parse_Moby_Data_JPEGImage]
    ParseMobyDataJpeg --> DecodeBase64[Decode_base64_to_byte]
    ParseMobyDataJpeg --> ParseMobyDataSimpleAnnotatedJpeg[Parse_Moby_Data_SimpleAnnotatedJpegImage]
    DecodeBase64 --> WorkflowOutputs
    ParseMobyDataSimpleAnnotatedJpeg --> WorkflowOutputs
    subgraph WorkflowOutputs
        images[images]
        annotations[annotations]
    end
```

Fetch images and annotations of snapdragons

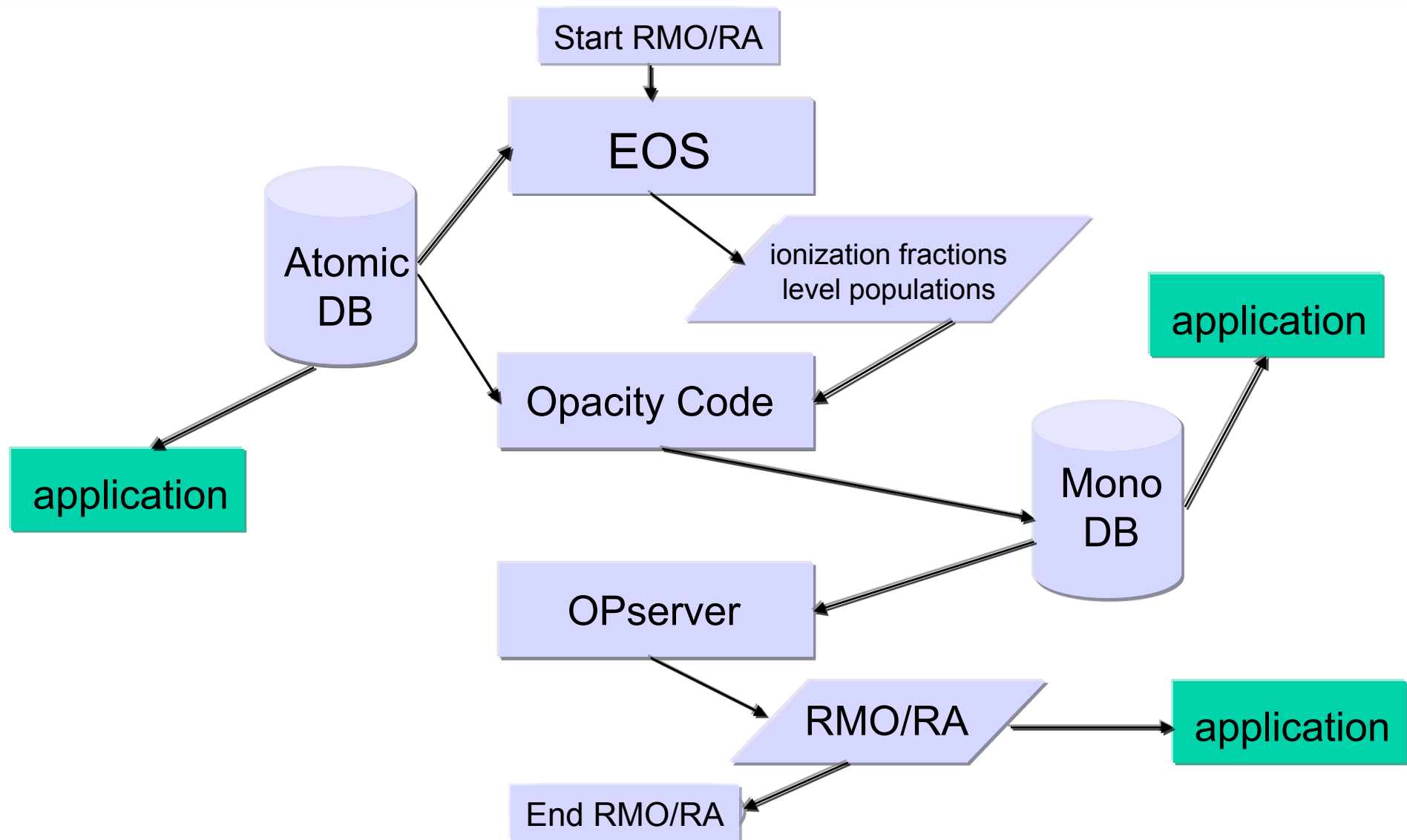
Tags

mygrid example taverna biomoby bioinformatics snapdragon image annotation

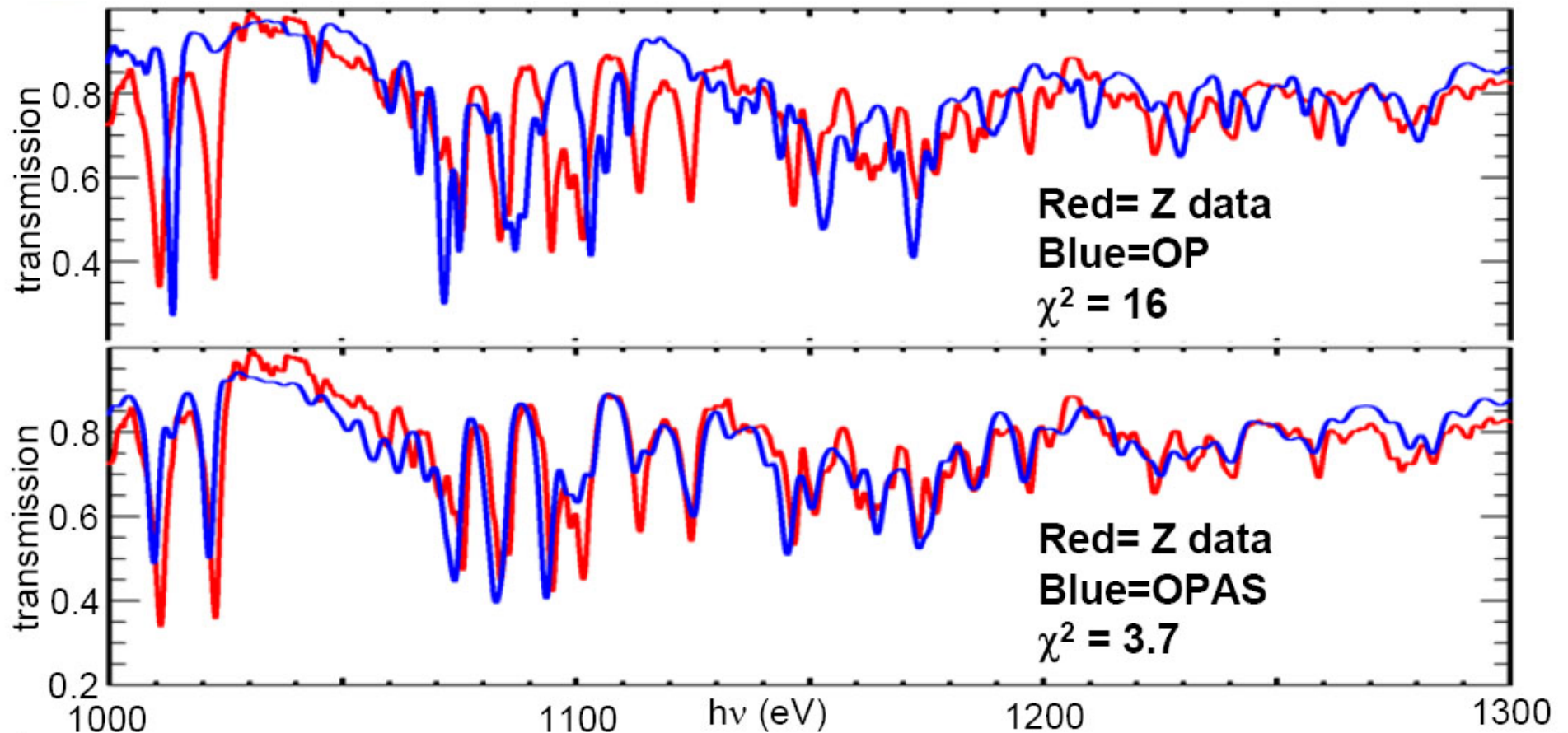
Credits

Tomoinn

# Current OP opacities flowchart



# OP Fe RMO at Z conditions is being questioned



**OP Rosseland mean is  $\sim 1.5\times$  lower than OPAS at Z conditions.  
If this difference persisted at solar conditions, it would solve the CZ problem**

Z conditions:  $T \sim 156$  eV,  $n_e \sim 10^{22}$  cm $^{-3}$

From Bailey (2008)

# Atomic data projects for e-science scale

Massive accurate and complete atomic data sets are in great demand:

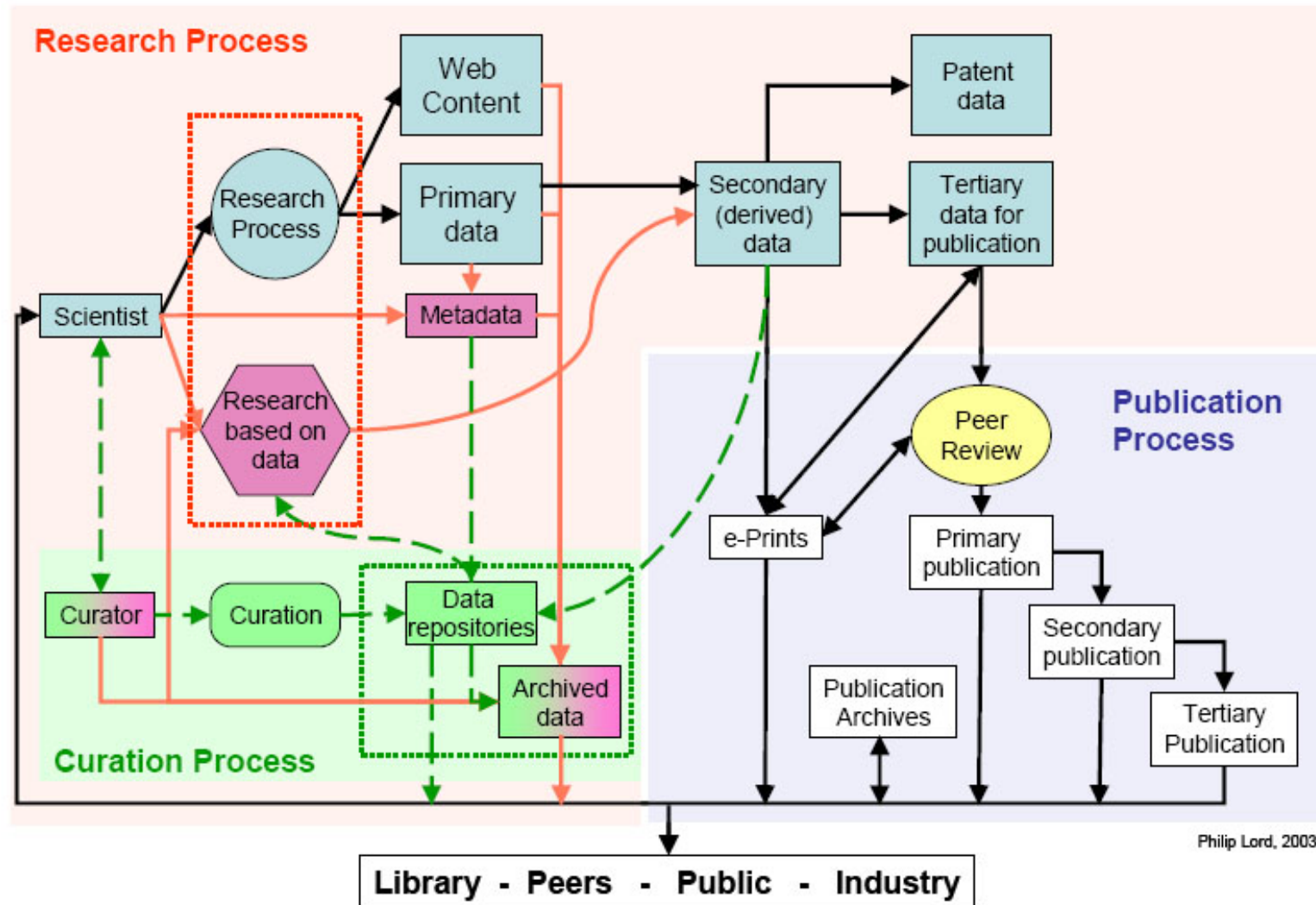
❖ Opacity Project III (IC)

❖ Ni Project

❖ NLTE Project

❖ W Project

# Model of the curation process



Original image from Lord et al (2004)

# Data management presents generic problems

## X-Info

- The evolution of X-Info and Comp-X for each discipline X
- How to codify and represent our knowledge



## The Generic Problems

- Data ingest
- Managing a petabyte
- Common schema
- How to organize it
- How to reorganize it
- How to share it with others
- Query and Vis tools
- Building and executing models
- Integrating data and literature
- Documenting experiments
- Curation and long-term preservation



- ✓ Scientific research is becoming increasingly collaborative and data-intensive (e-science)
- ✓ Atomic data production must be scaled up to the extreme requirements of virtual organizations
- ✓ Data repositories must be kept fit and integral for contemporary purpose, discovery and reuse (e-science curation)
- ✓ Data preservation is of vital importance