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

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

VAMDC Virtual Atomic and Molecular Data Centre

CeCalCULA

IVIC

l'Observatoire de Paris

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.

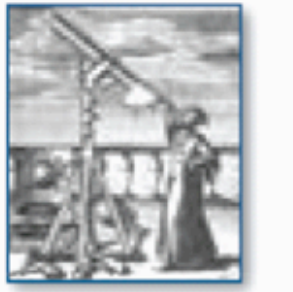
- VAMDC is supported by EU in the framework of the FP7 "Research Infrastructures - INFRA-2008-1.2.2 - Scientific Data Infrastructures" initiative. It started on the 1rst of July for a duration of 42 months.

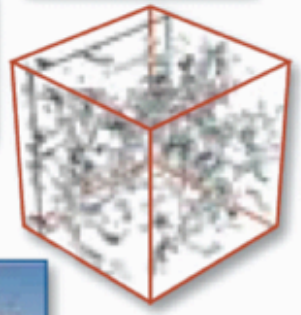
E-science is collaborative data-intensive science

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

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



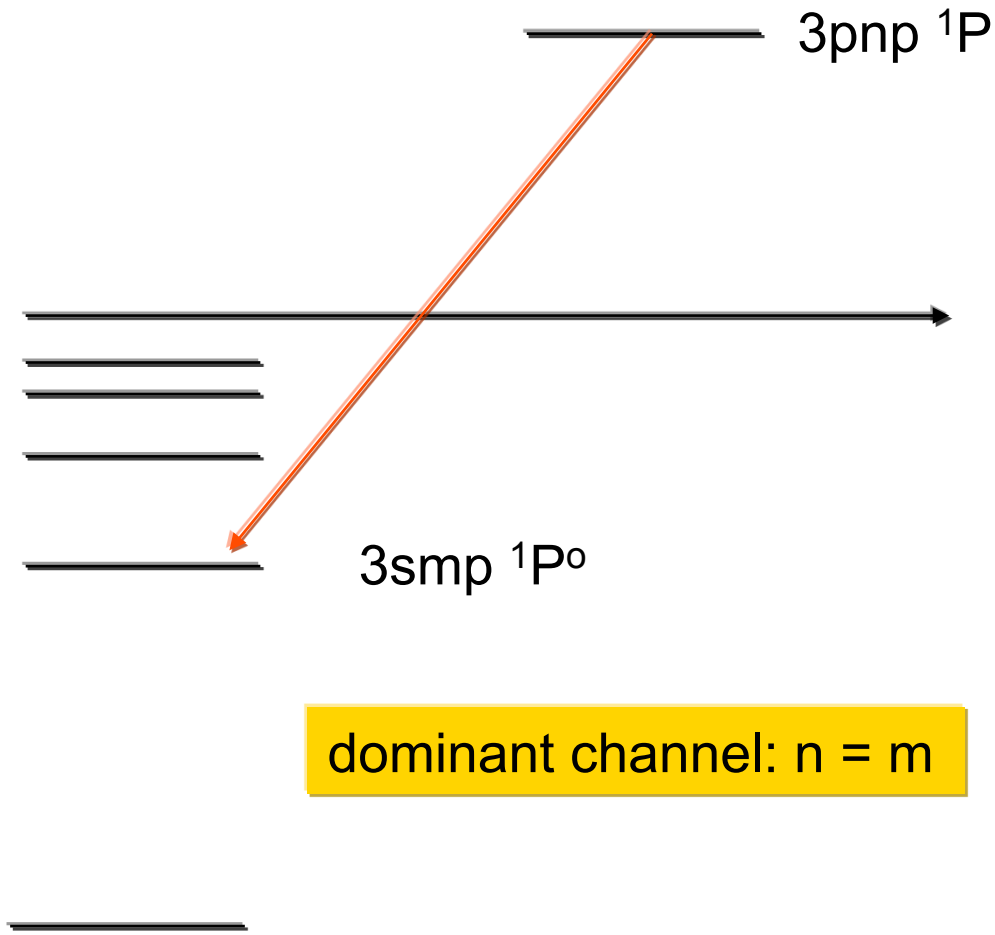


Original image from Hey, Tansley & Tolle (2009)

Radiative decay of 3pnp ^{1}P states in Mg-like ions

Mg I

State	RLT (ns)
3p4p ^{1}P	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



dominant channel: n = m

Butler et al (1990)

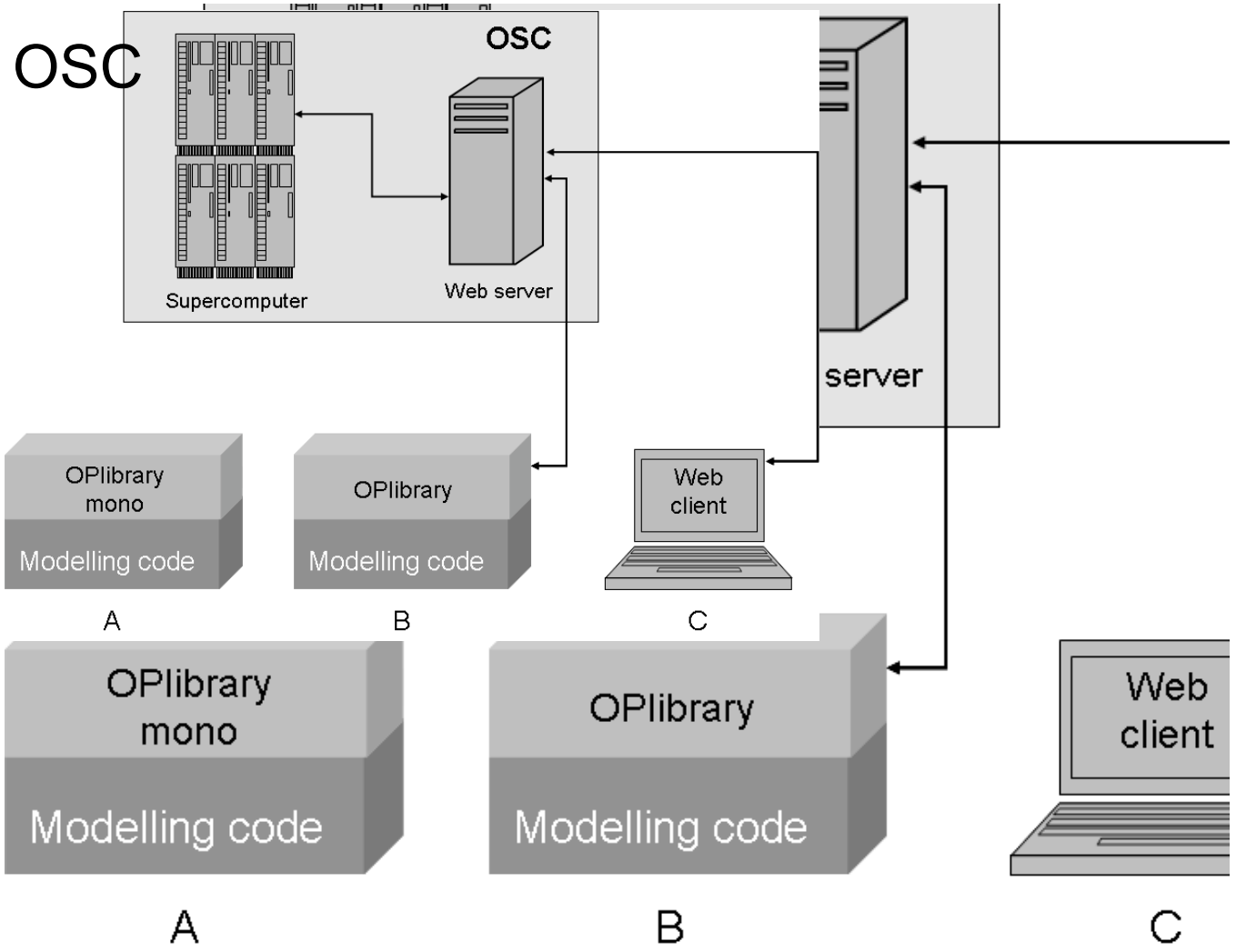
Application and data access in e-science

Application and data access in e-science is evolving:

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

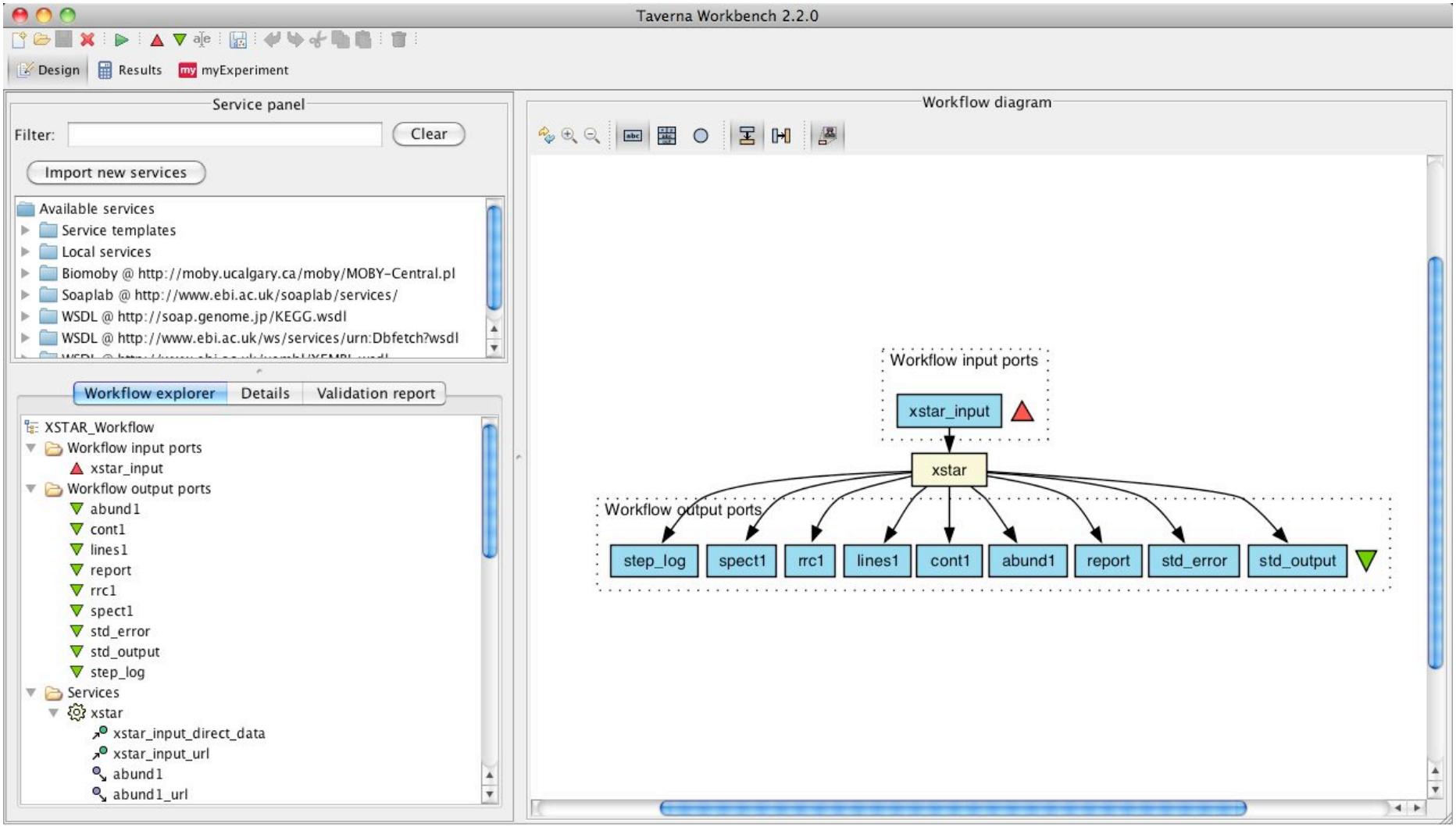
Application and data access in e-science

OPserver at OSC

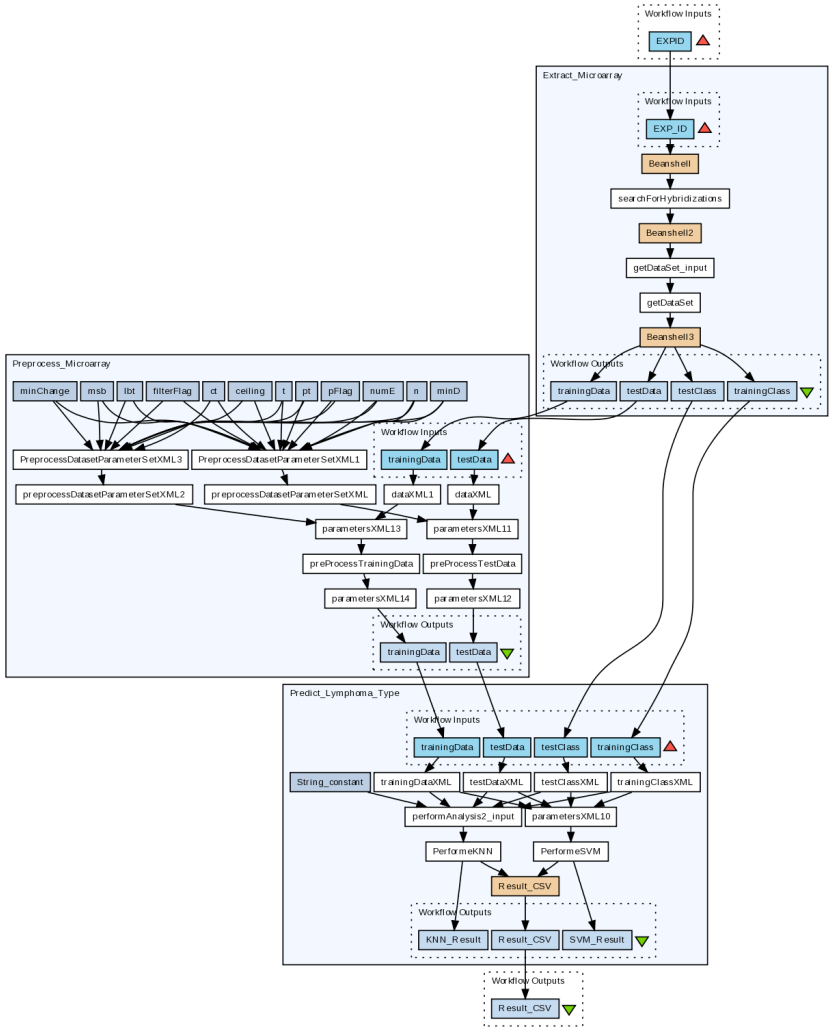


From Mendoza et al (2007)

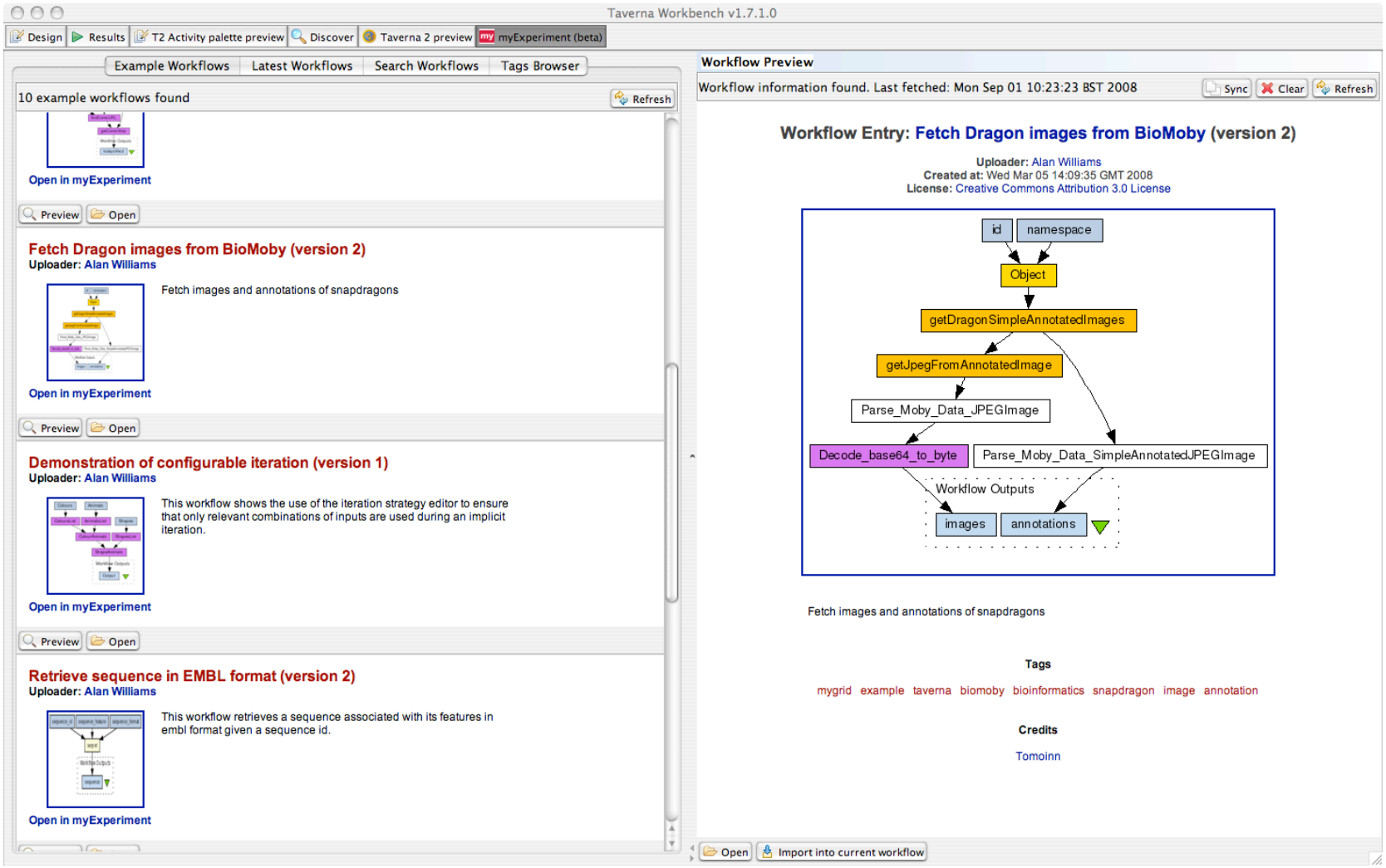
Workflow systems



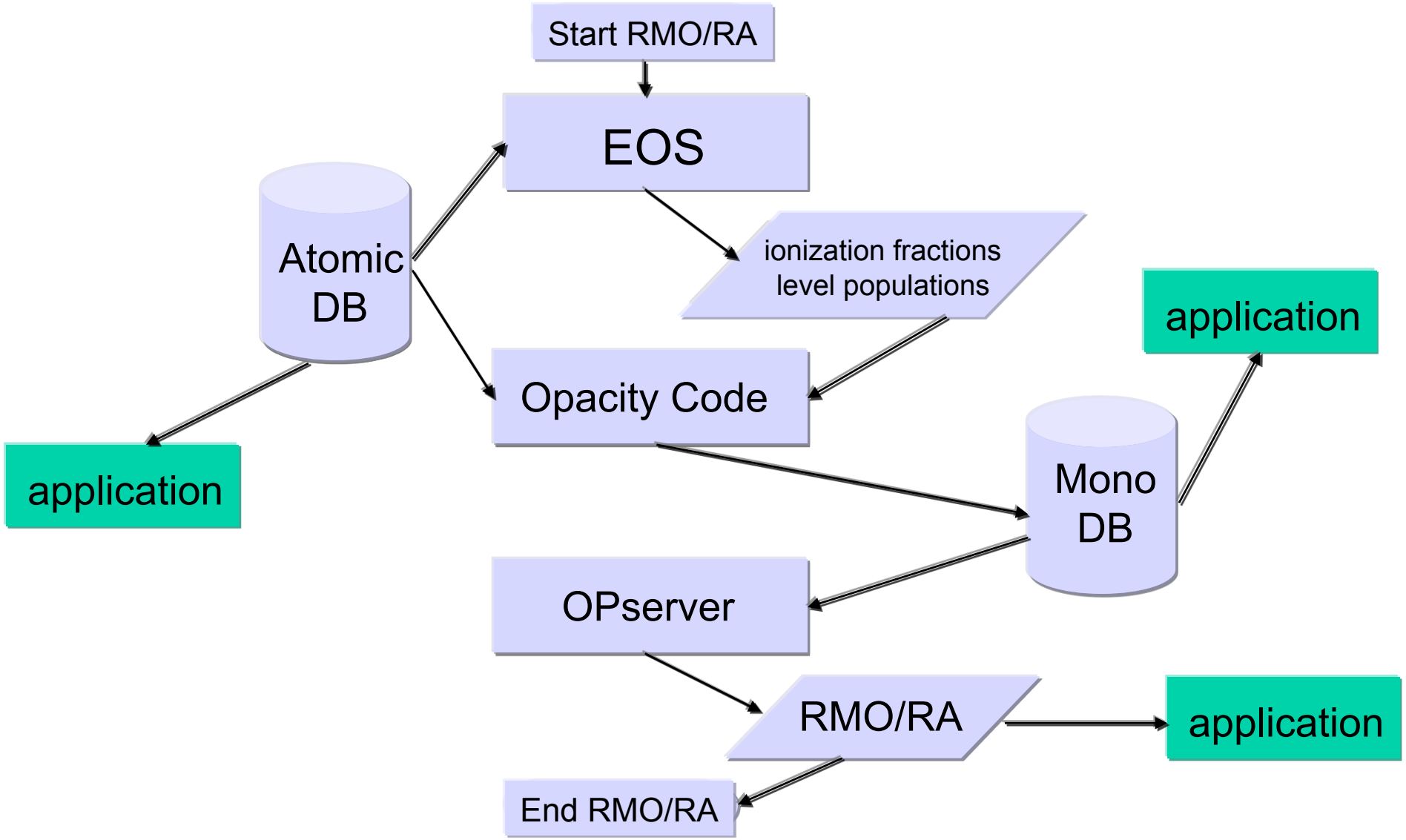
Workflow systems



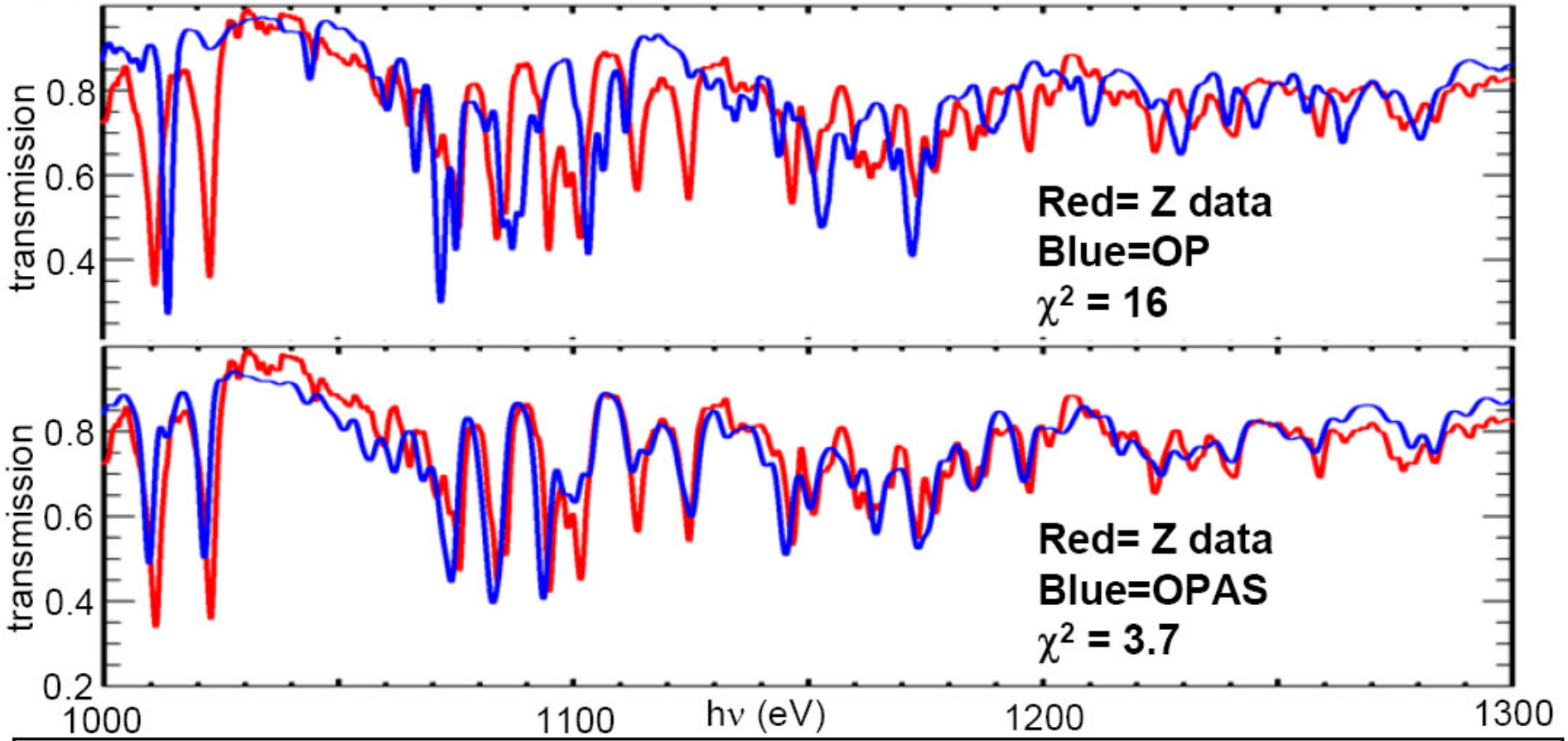
Workflow systems



Current OP opacities flowchart



OP Fe RMO at Z conditions is being questioned



OP Rosseland mean is ~ 1.5x lower than OPAS at Z conditions.
If this difference persisted at solar conditions, it would solve the CZ problem

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

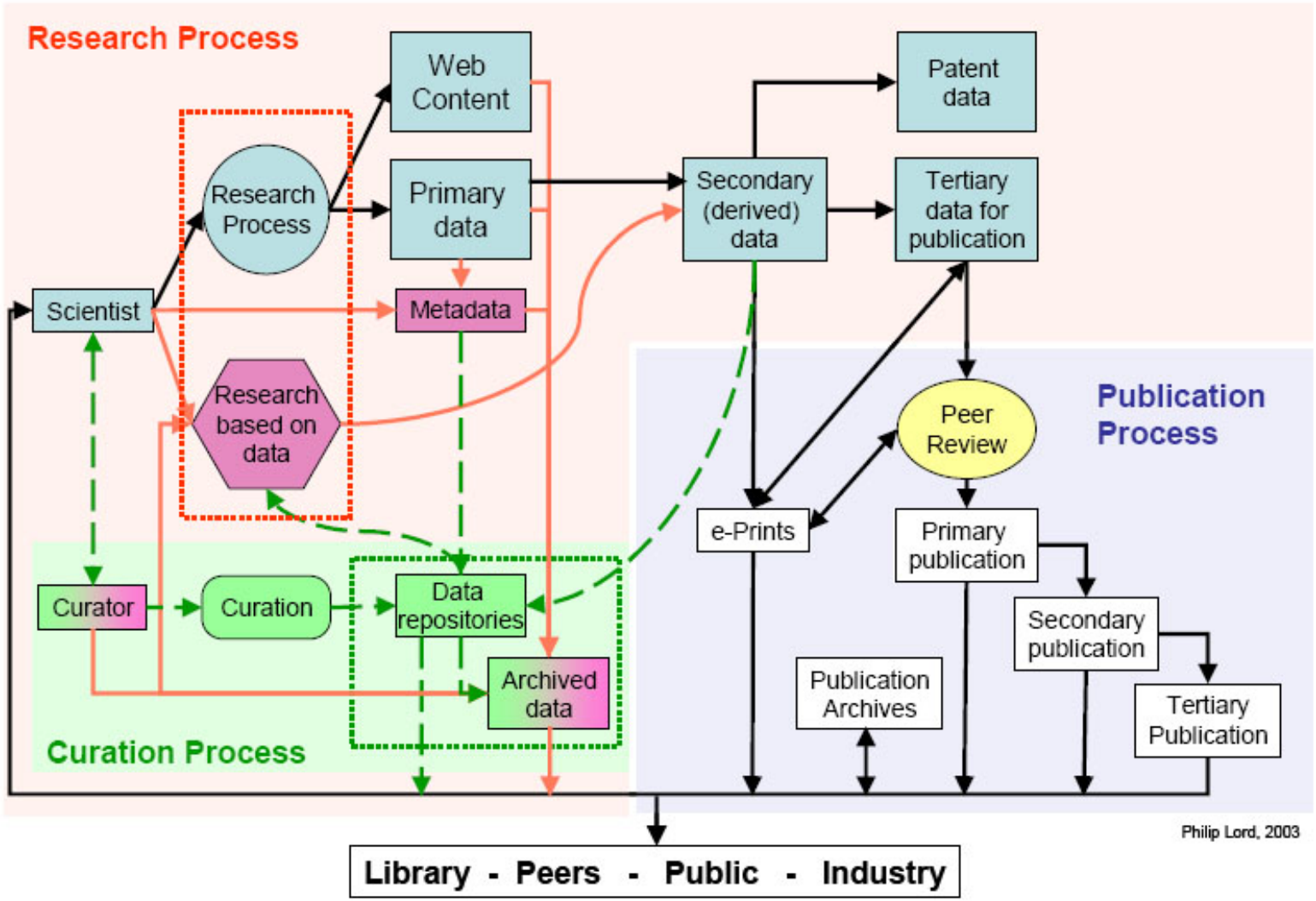
From Bailey (2008)

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

Conclusions

- ✓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