Grant agreement no.: 27092

Mashing-up Science
Collaborative Digital Experiments

Jose Enrique Ruiz - @bultako
IAA-CSIC

April 4th 2011
2011 .Astronomy Conference - Oxford

.astronomy

SEVENTH FRAMEWORK PROGRAMME

European Commission
Information Society and Media

Information Society Technologies

Summary

- Astronomy research lifecycle
- Research objects: the ingredients
- Scientific workflows: the cooking recipes
- Wf4Ever: preserving knowledge

Proposals Submission

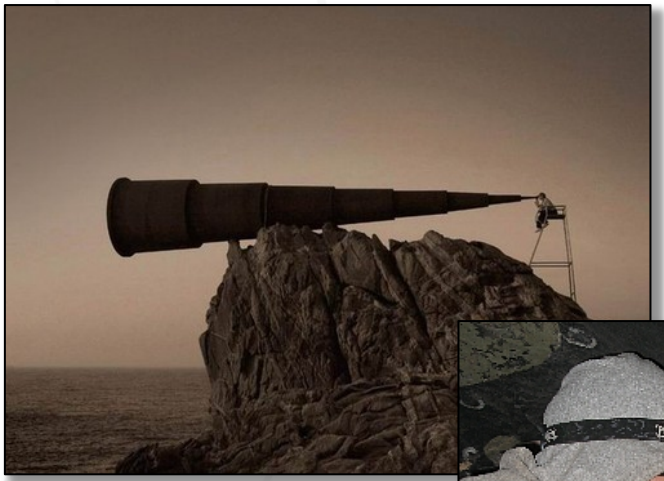
- Funding
- Observing time

- Description of the experiment
- Precedent related studies
- Problems to solve and strategy to follow
- Requirements and expected results

- The digital sky vs. Observing time proposals

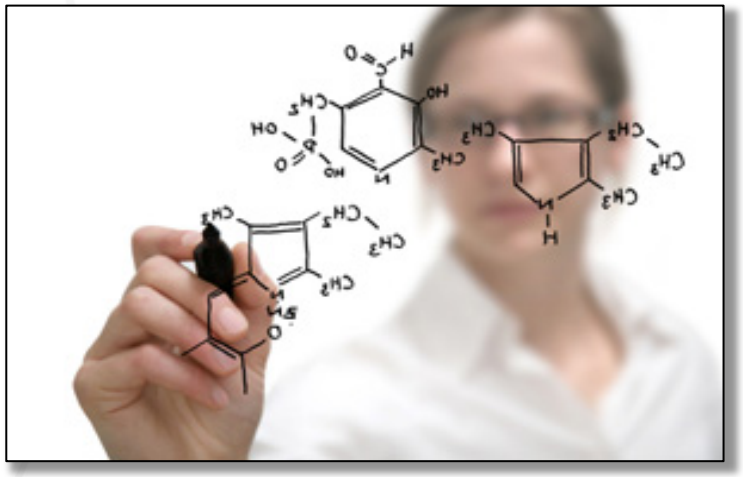
Observation and data reduction

- Ancillary data
- Air masses
- Exposure time
- Meteorological conditions
- Instrumental signatures
- Observed raw datasets
- Science-ready datasets



- Observational programs
- Pipelines for automated data reduction

Analysis of the data



- Specific computing environments
- Specific local interactive software
- Commercial packages (IDL)
- Grid and clusters for simulations
- High level programming languages
  - Fortran, C++
  - Python recipes community
- Use of web archives for well known properties of objects
  - Vizier, Simbad, NED
- SaaS approach slowly coming with the Virtual Observatory

Publishing

- Electronic PDF files
- Provenance concerning the Analysis
- Data Results are hidden behind the plots
- Most of them are public
- Interlinking provided
  - Astronomical objects
  - Catalogs and missions
  - Related publications
  - Citations, datasets, proposals..

Astronomy research is entirely digital
Time has come to go "Beyond the PDF"

- Methodology "in action"
- All data exposed
- Indexed experiments
- Reproducible
- Repeatable
- Participatory
- Formative
- Collaborative
- Cross-boundary

Research Object

A digital entity capturing all the components needed for the execution of a digital experiment and also the results produced, describing and characterizing the overall experiment, every single one of its components and the links existing among them.

The R's Dimensions

- Repeatable
- Reproducible
- Replayable
- Refreshable
- Reusable
- Reliable
- Retrievable
- Roll-backable
- Referenceable
- Research cross-boundary

Towards exchange and reuse of digital knowledge
Bechhofer, S., De Roure, D., Gamble, M., Goble, C. and Buchan, I. (2010)

Research Objects in Astronomy

- Metadata characterization
- Description of the experiment
- Related bibliography
- Ancillary and raw data
- Reduced science-ready data
- Digital environment needed
- Scripting and software used
- Links to web archives
- Final data products
- Scientific discussion

Scientific Workflows

The combination of data and processes into a configurable, structured set of steps that implement semi-automated computational solutions in scientific problem-solving.

A digital recipe which can itself be cooked in order to produce repeatable results.

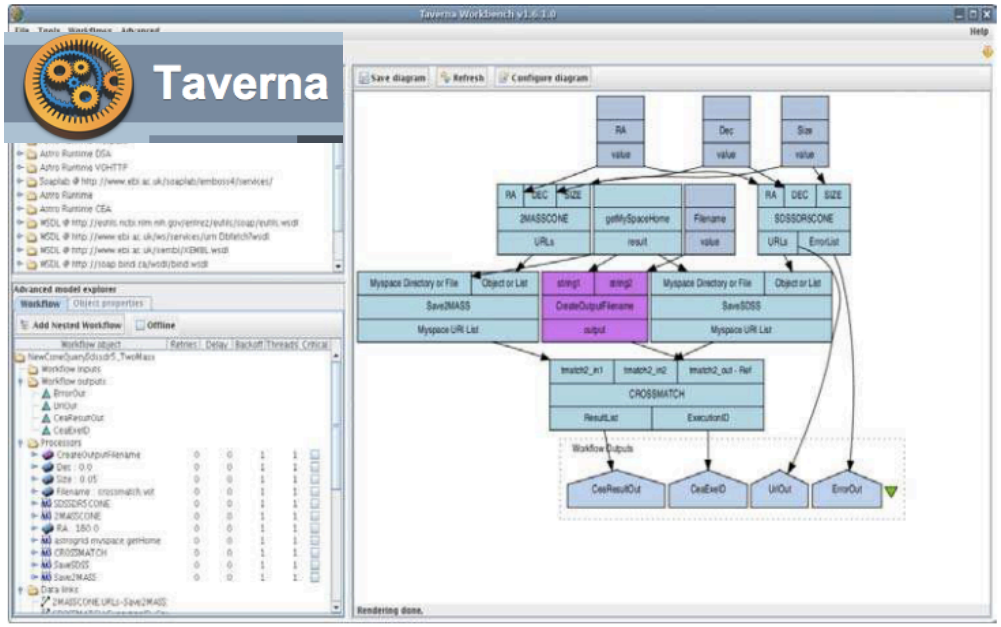
Scientific Workflows

- Enable automation
- Make science reproducible
- Sometimes repeatable
- Encourage best practices
- Modular nature allows reuse
- Exposes the scientific method
- Formative
- Scientist friendly

The oven

A workflow Management System

- Scalable suite of tools
- Workflow design
- Workflow execution
- Workbench
- Server implemented
- Domain independent
- Access to remote resources
- Open source

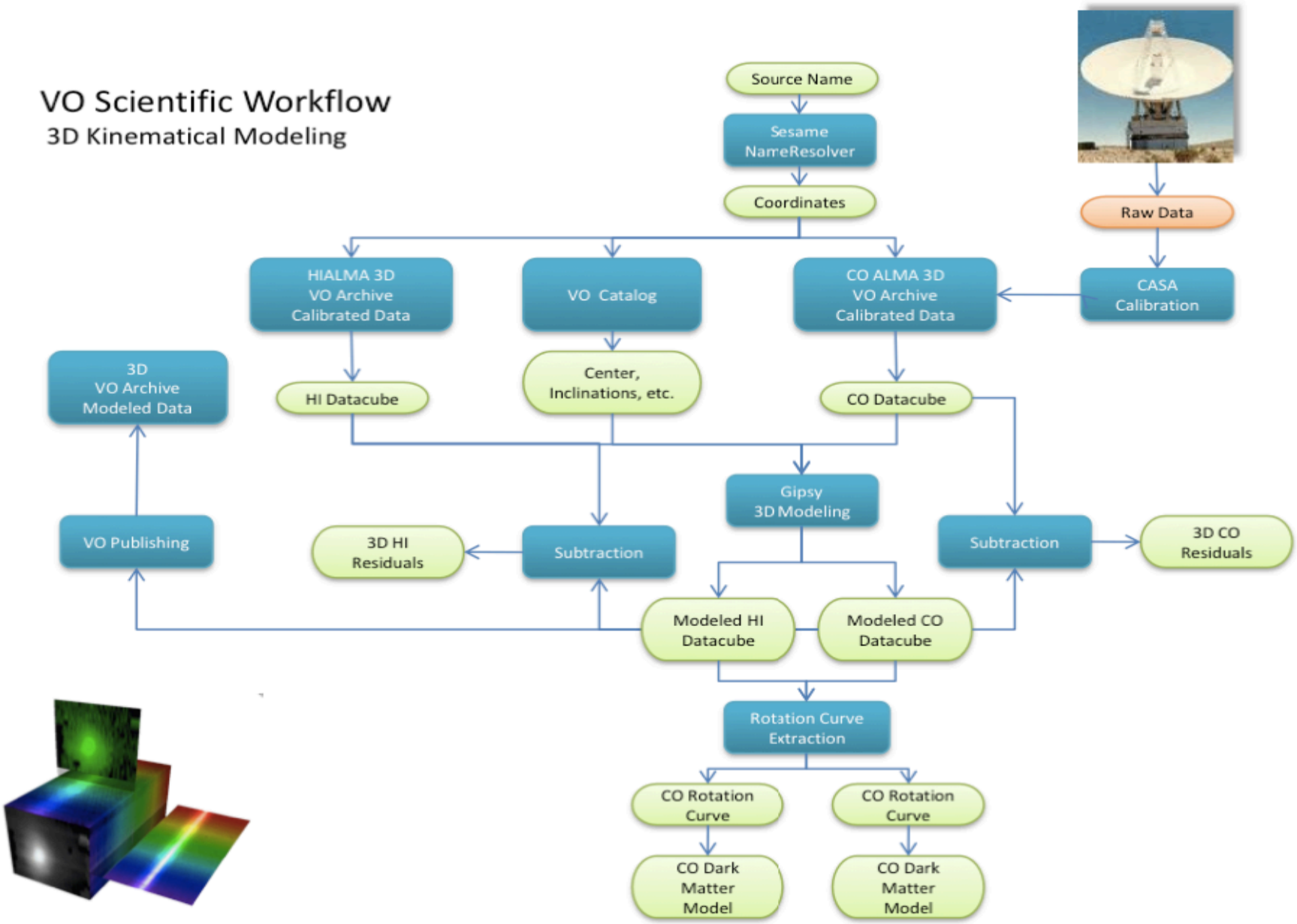


Scientific workflows: the cooking recipes

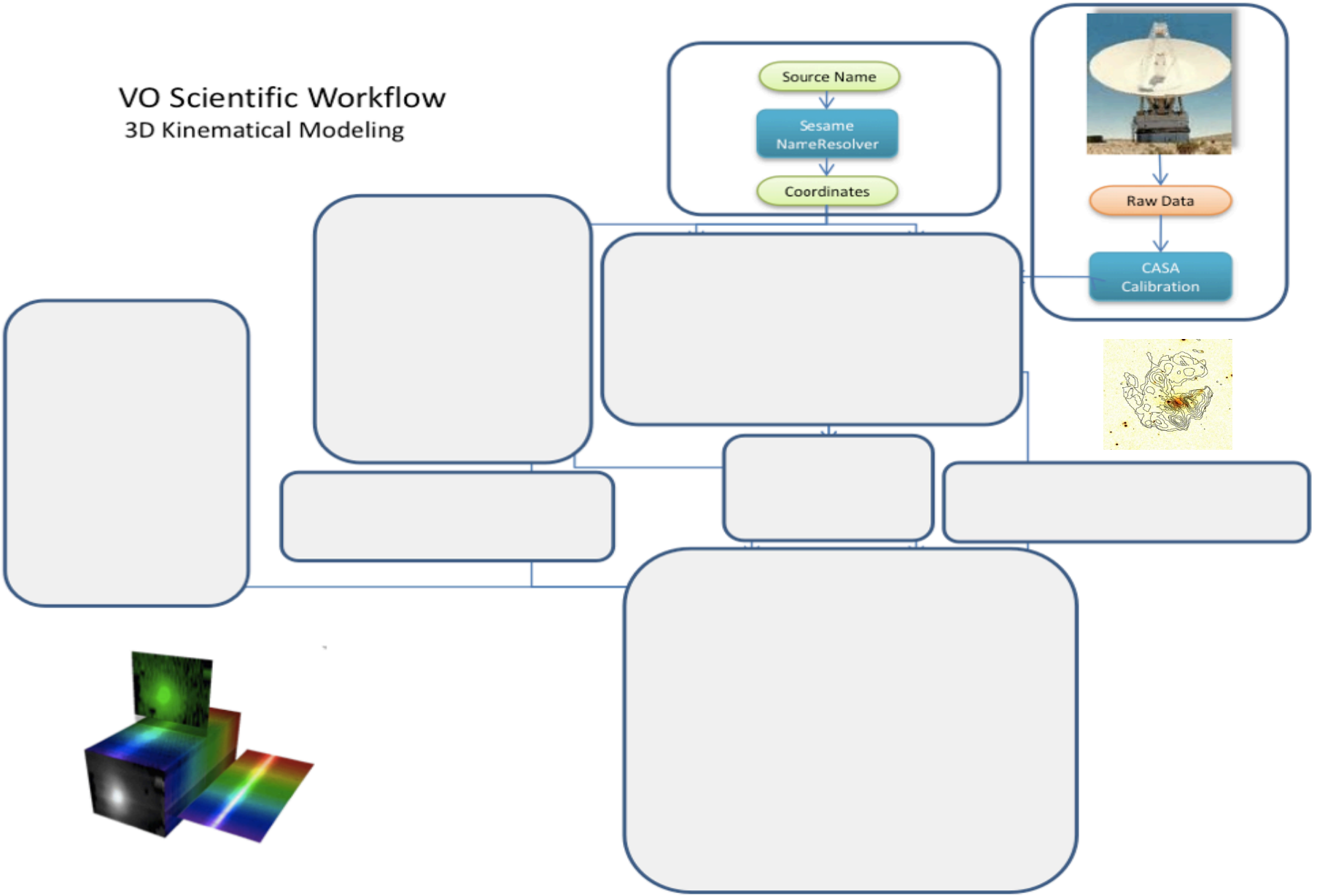
The recipes store

- Find workflows
- Share workflows and files
- Find people
- Build communities
- Publish packages
- Tag workflows
- Score and rate workflows
- Comment on workflows
- Write reviews

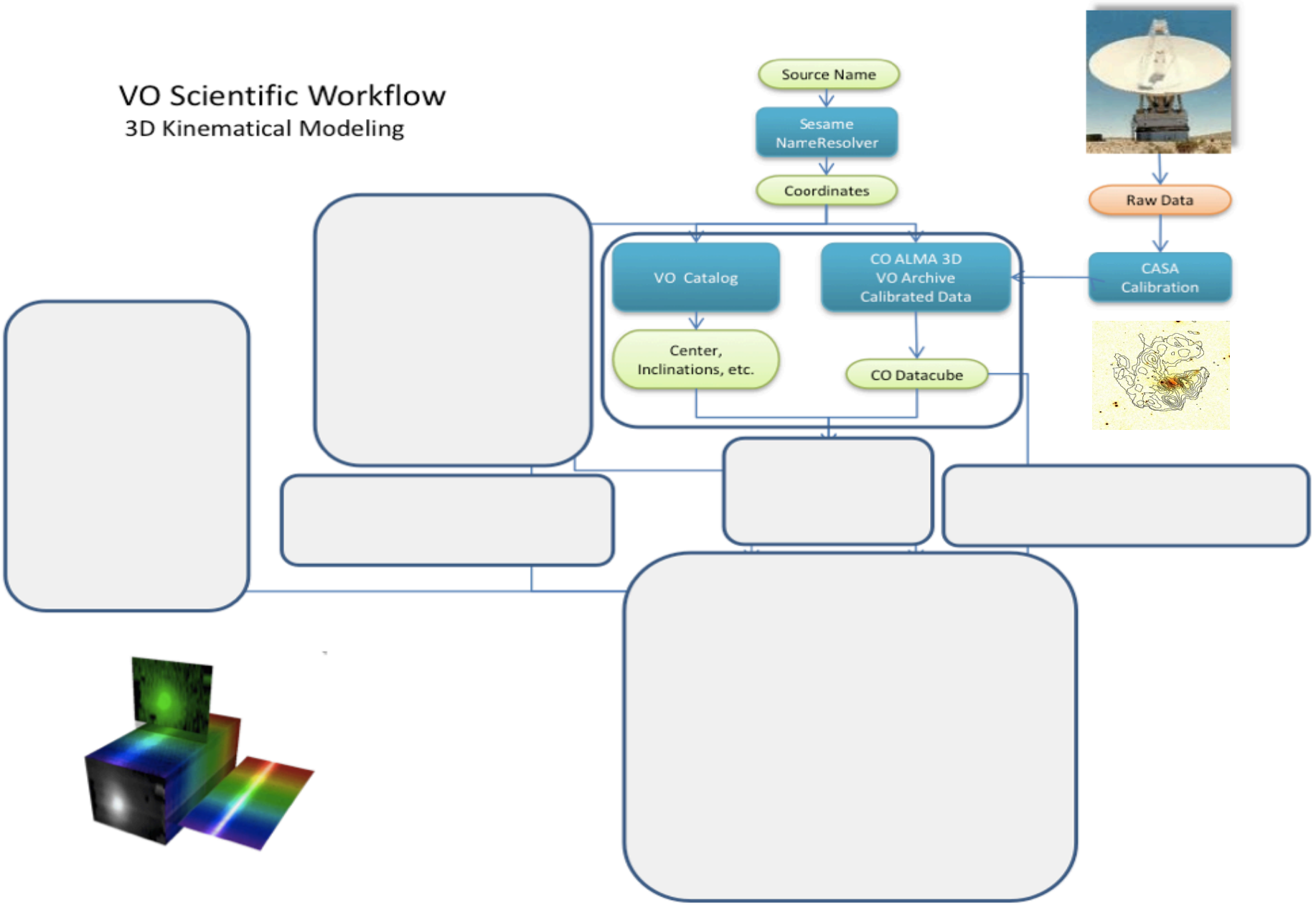
Scientific workflows: the cooking recipes



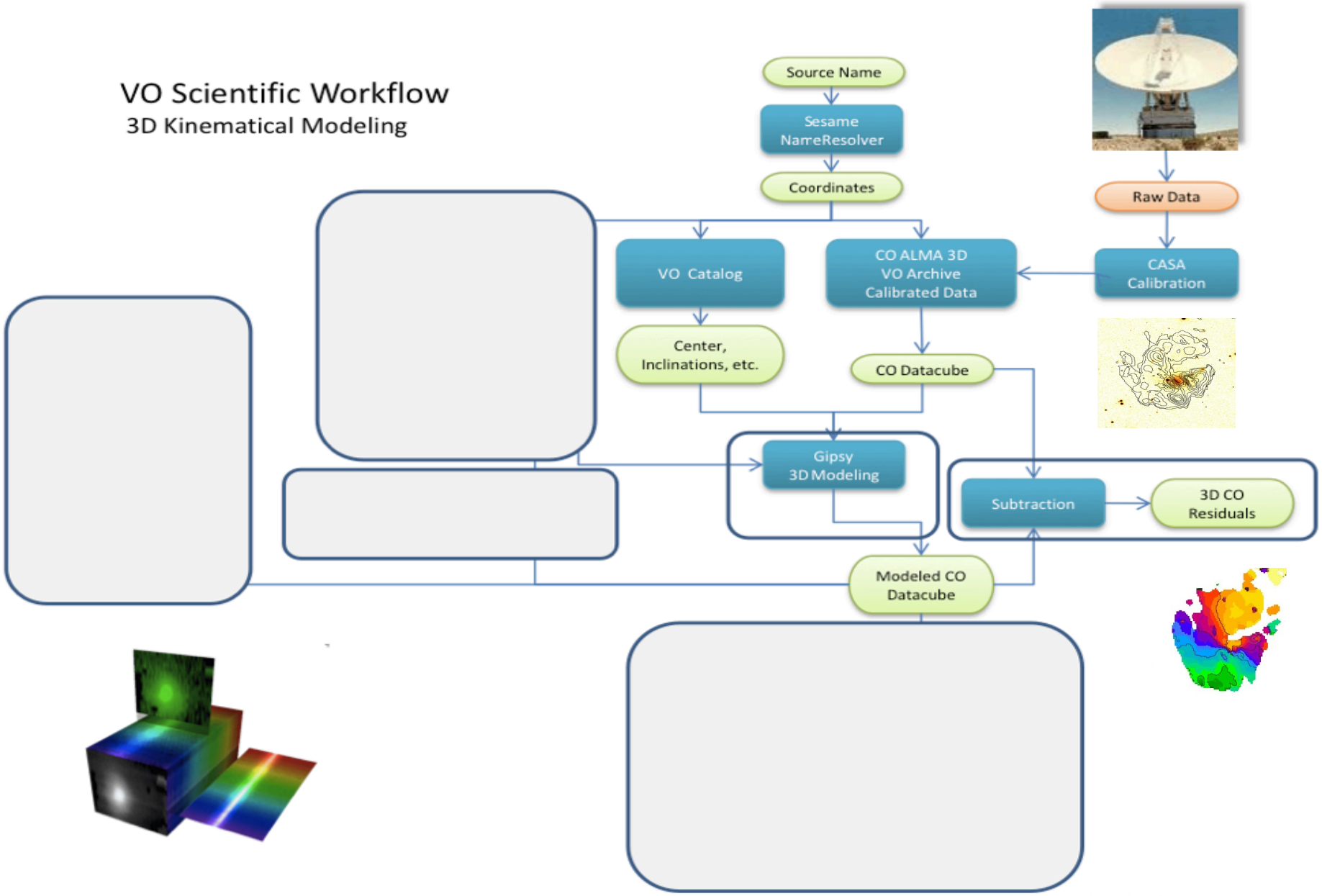
Scientific workflows: the cooking recipes



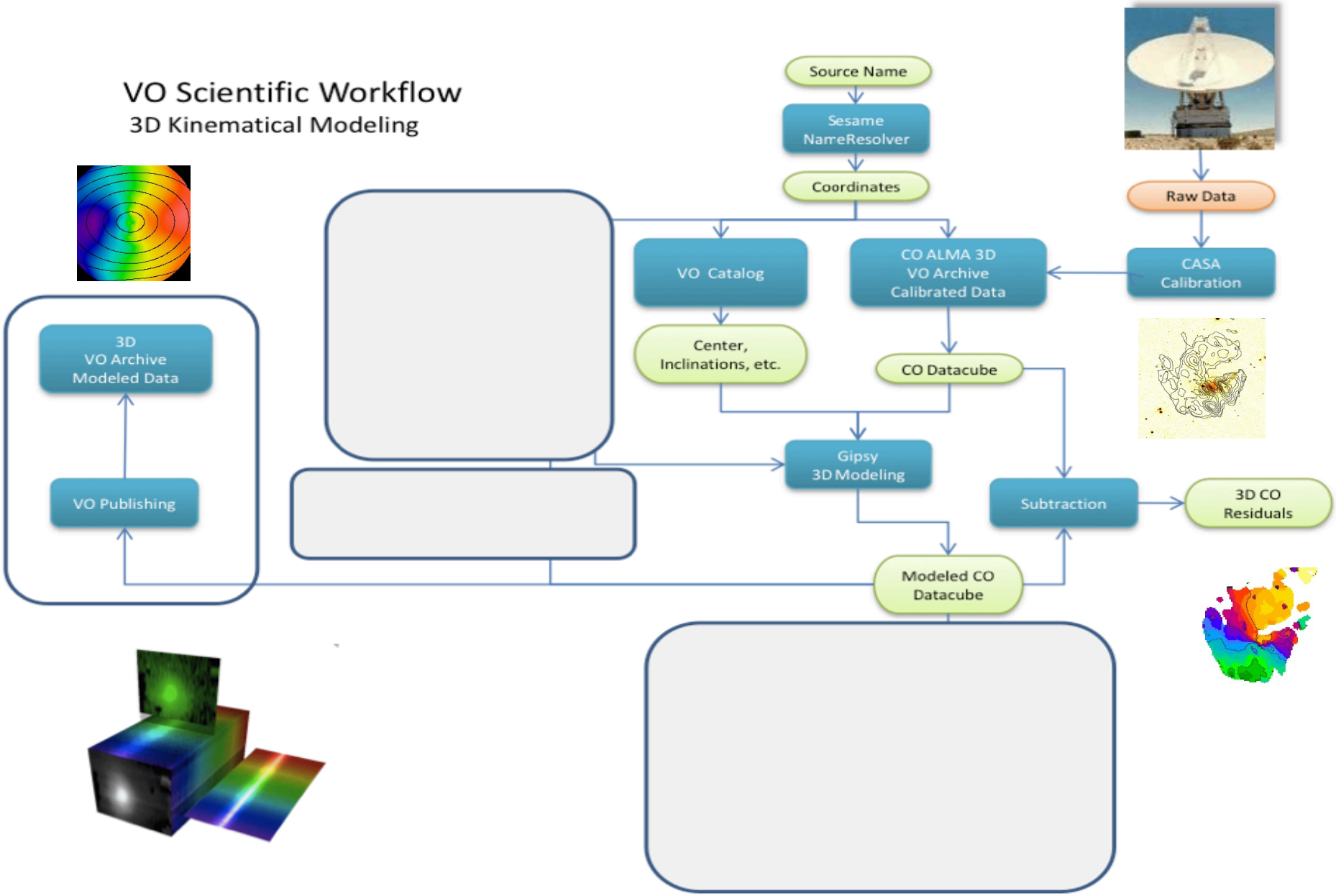
Scientific workflows: the cooking recipes



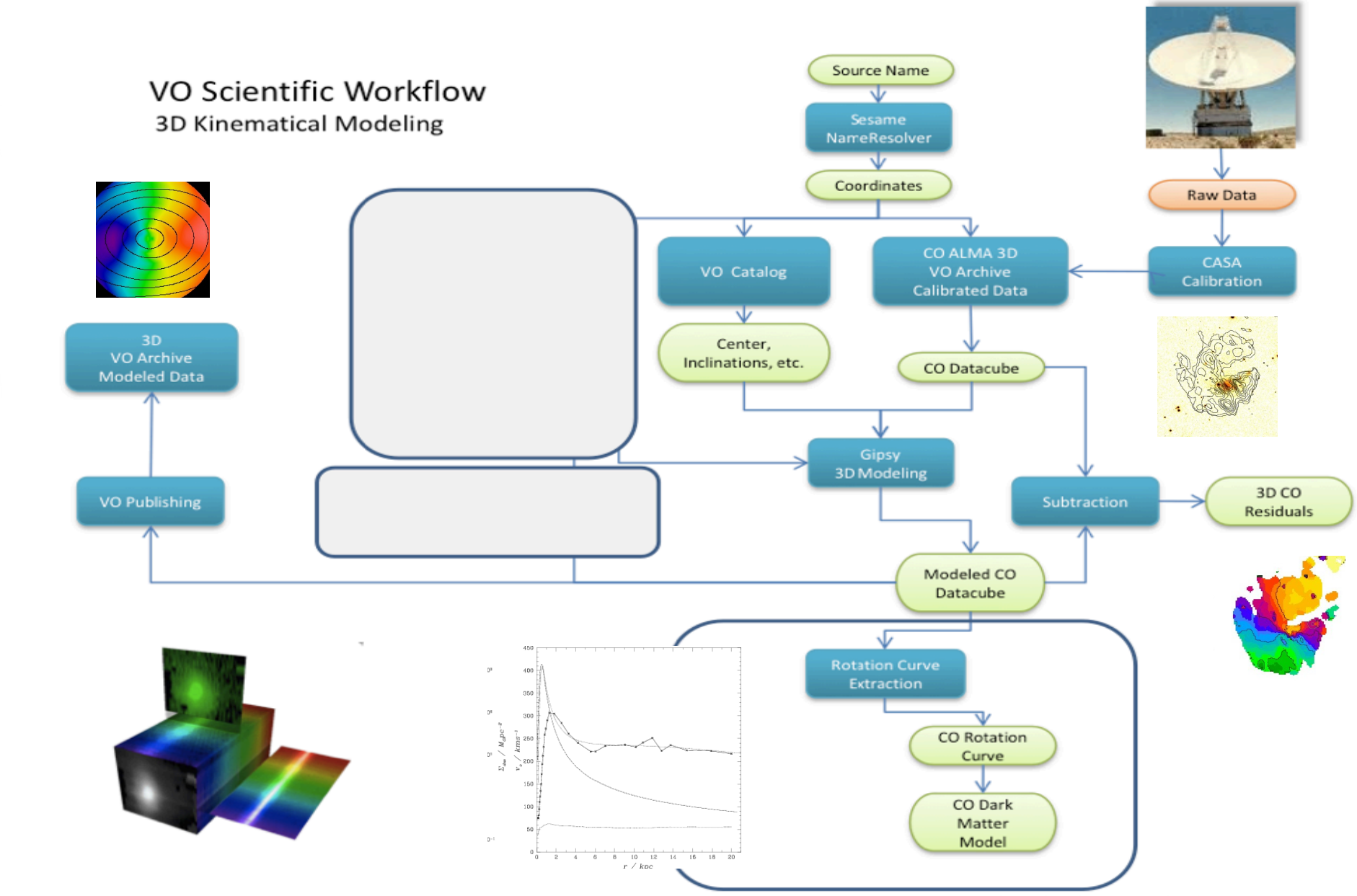
Scientific workflows: the cooking recipes



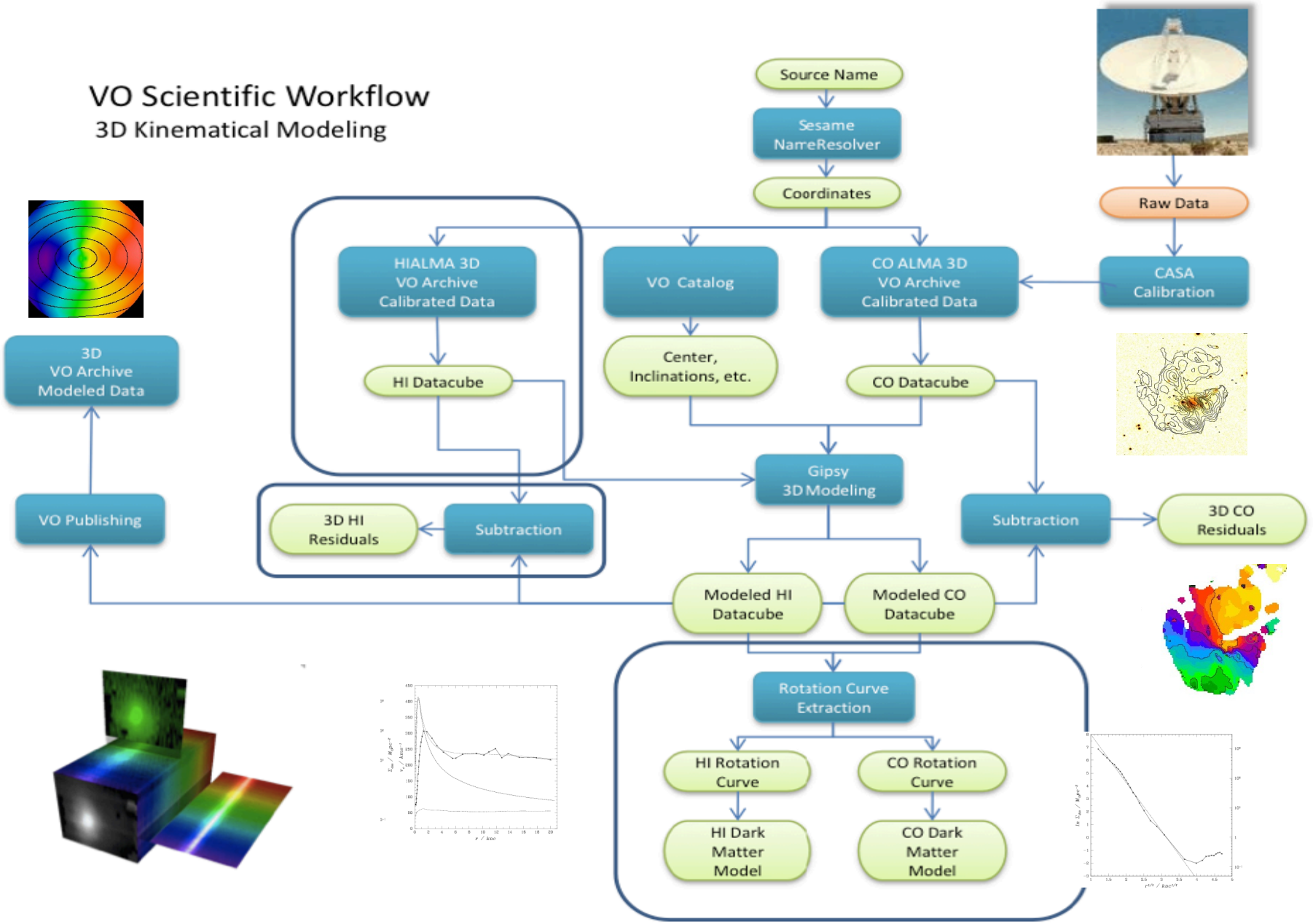
Scientific workflows: the cooking recipes



Scientific workflows: the cooking recipes



Scientific workflows: the cooking recipes



Wf4Ever goals

Preservation of Research Objects

- Data
- Resources
- Workflows

Workflow preservation is complex

- Interpreted through their execution
- Complex models are required to describe them
- Provenance is a complex issue in a cloud of services
- Need of Web Semantics, Ontologies, Linked Data, RDF, etc..
- Resources are often beyond control of scientists

Wf4Ever goals

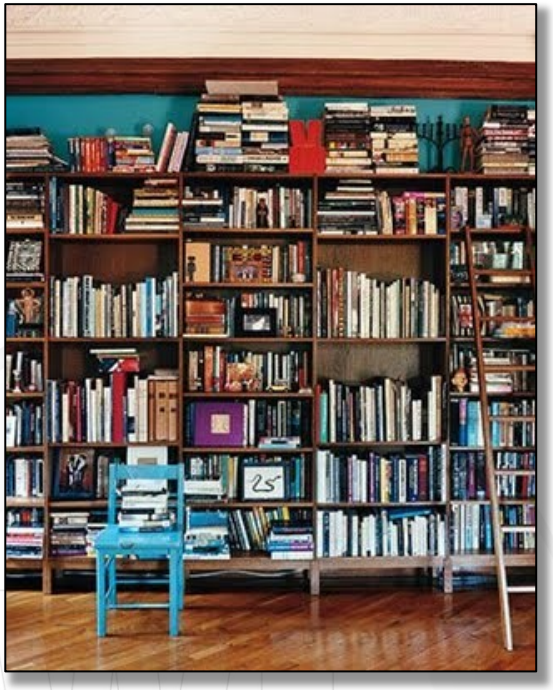
Creation of workflows and research objects for astronomical digital experiments

- Data archives
- Web services
- Virtual Observatory standards
- Ubiquitous storing and computing
- Python based community
- Interlinked digital libraries

Wf4Ever goals

Build an infrastructure for Research Objects management which stimulates the development of new scientific knowledge via collaborative work

- Creation
- Archival
- Classification
- Indexing
- Retrieval
- Community reuse
- Rating, scoring and annotations
- Scalable in semantic repositories



Users roles

Reader

Skims titles and abstracts of published research objects

Comparator

Looking for similar research objects to those he's working with at present

Re-user

Extract and replace modules from the workflow and use it for his own purpose

Publisher

Wants to the community to check his digital experiment

Evaluator

He is allowed to evaluate, comment and rate a specific research object

ROBox: the basket

The team

1. Intelligent Software Components (ISOCO, Spain)
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Centre (PSNC, Poland)
5. Universisty of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)

Thanks for your attention !

http://www.wf4ever-project.org