



Grant agreement no.: 27092

Mashing-up Science Collaborative Digital Experiments

Jose Enrique Ruiz - @bultako
IAA-CSIC



April 4th 2011

2011 Astronomy Conference - Oxford



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

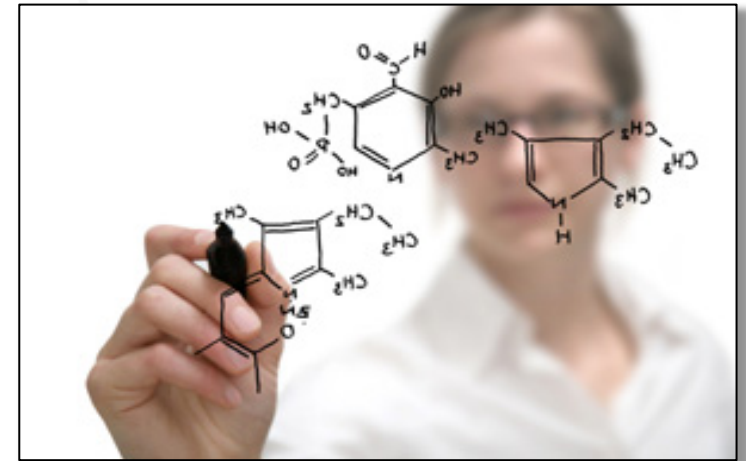
- Ancillary data
- Air masses
- Exposure time
- Meteorological conditions
- Instrumental signatures
- Observed raw datasets
- Science-ready datasets

- Observational programs
- Pipelines for automated data reduction



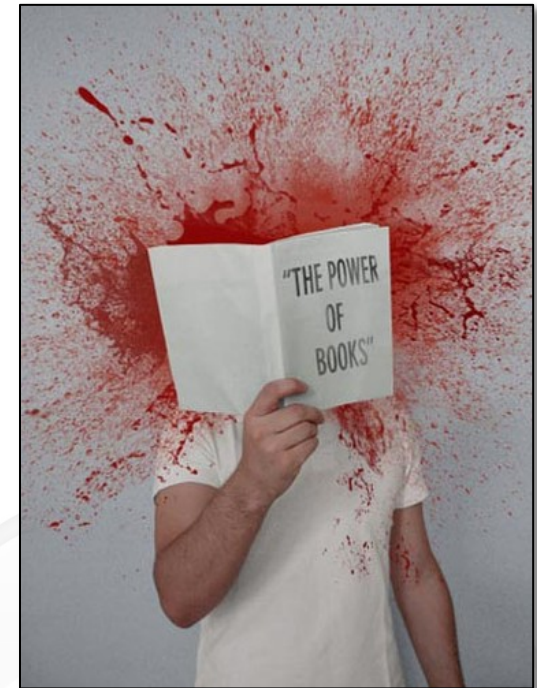
Analysis of the data

- 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

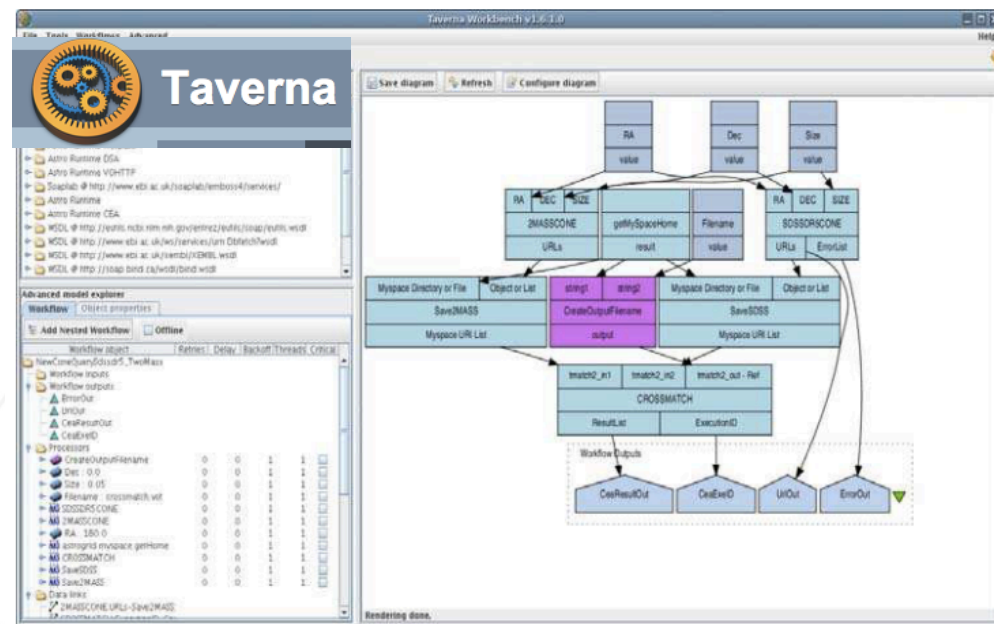
- Enable automation
- Make science reproducible
- Sometimes repeatable
- Encourage best practices
- Modular nature allows reuse
- Exposes the scientific method
- Formative
- Scientist friendly



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



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

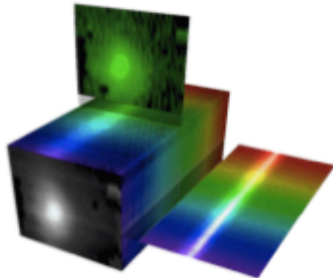
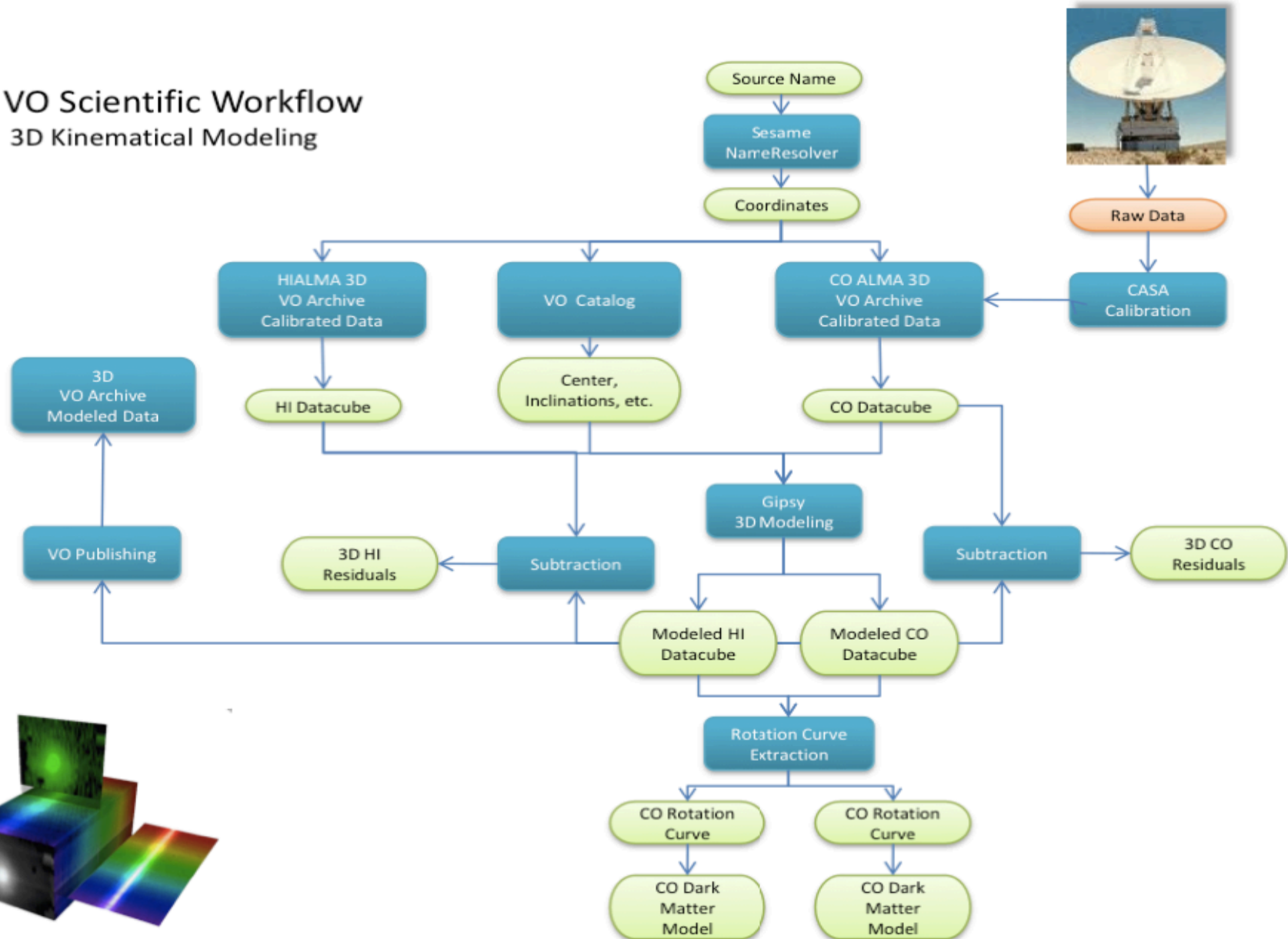


The screenshot shows the 'myexperiment' website interface. At the top, there are navigation links for 'Home', 'Users', 'Groups', 'Workflows', 'Files', and 'Packs'. A search bar is present with a dropdown menu set to 'All'. Below the navigation, the 'Workflows' section is active, displaying a list of search filter terms and a 'Filter by type' section with various workflow categories and their counts. The main content area shows two workflow entries:

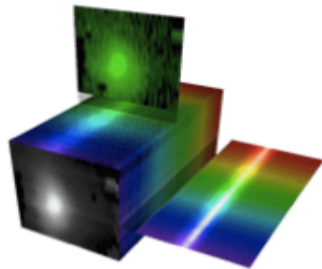
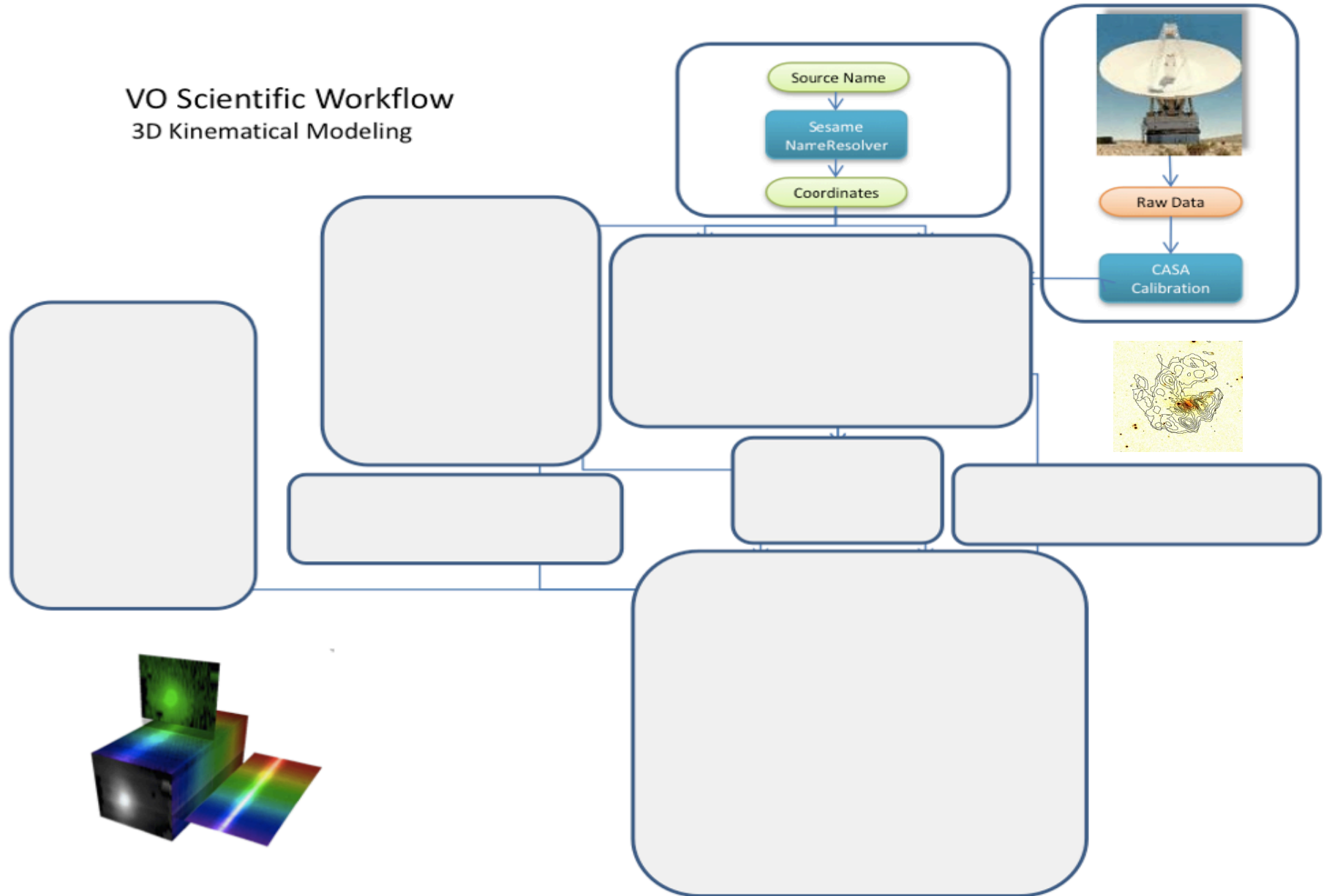
- Workflow 1:** 'Mapping OligoNucleotides to an assembly (v7)' by Wassinki. It includes a description, a version history table, and a detailed version info box explaining a change in transcript handling. The workflow has a rating of 0.0/5, 7 versions, and 0 reviews.
- Workflow 2:** 'Add Mesh String to Biological Process (v2)' by Paul Fisher. It includes a description and a diagram showing the integration of a MeSH term into a KEGG pathway. The workflow has 0 ratings, 1 version, and 0 reviews.

On the right side of the page, there is a 'New/Upload' section, a 'Log in / Register' form, and a 'Popular Tags' section listing various tags like 'benchmarks', 'bioinformatics', and 'BLAST'.

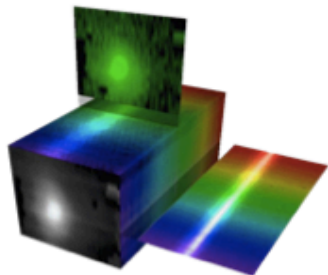
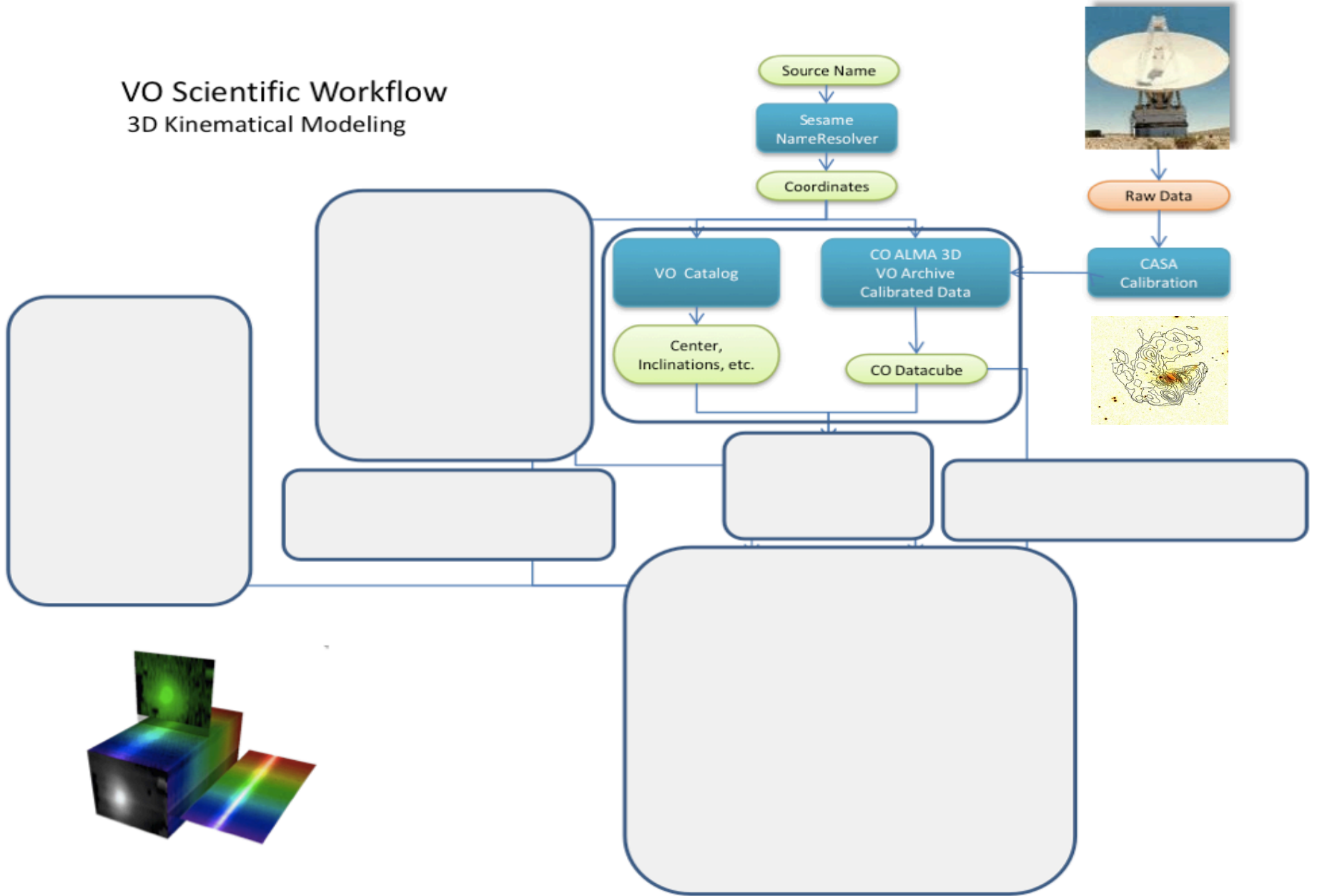
VO Scientific Workflow 3D Kinematical Modeling



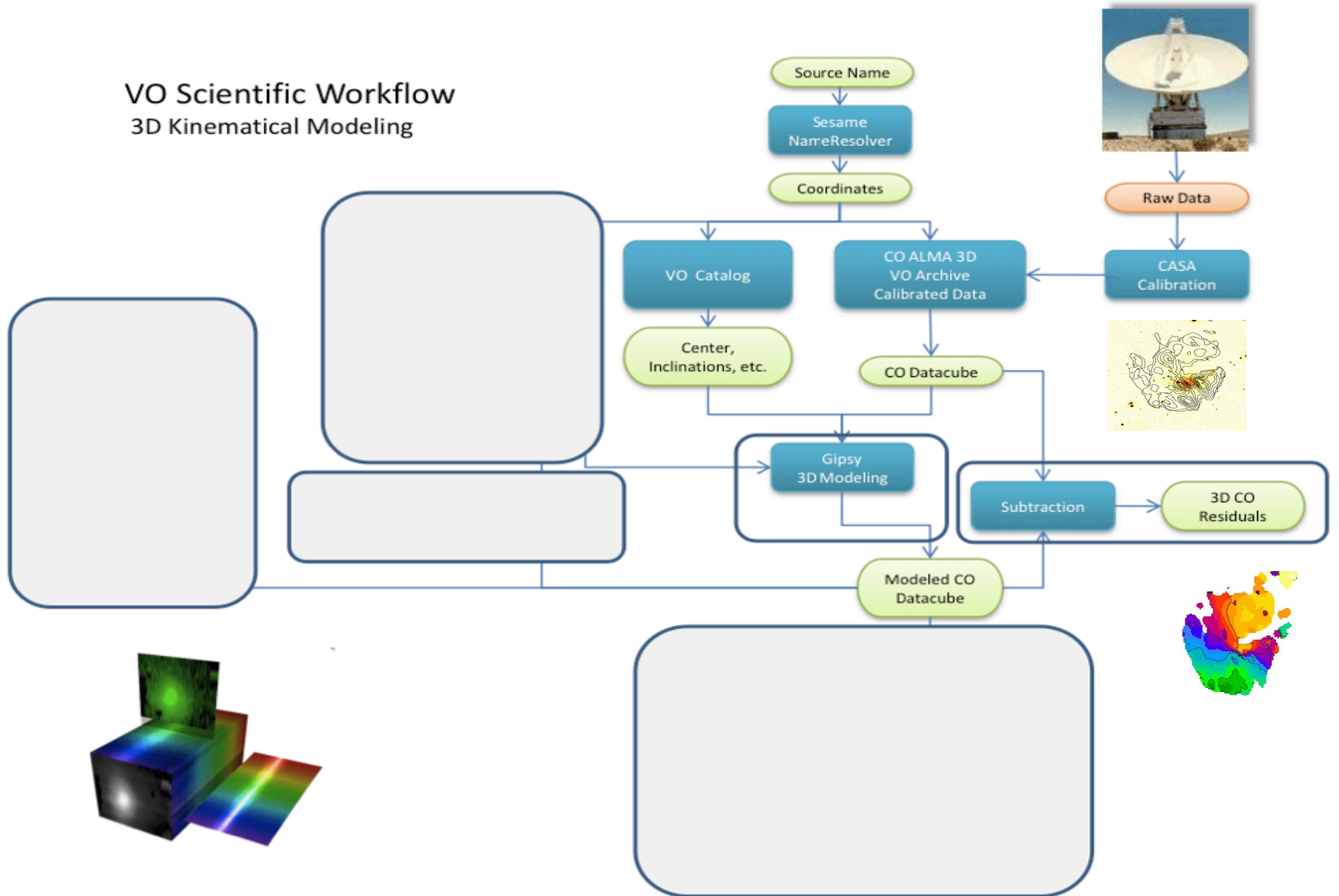
VO Scientific Workflow
3D Kinematical Modeling



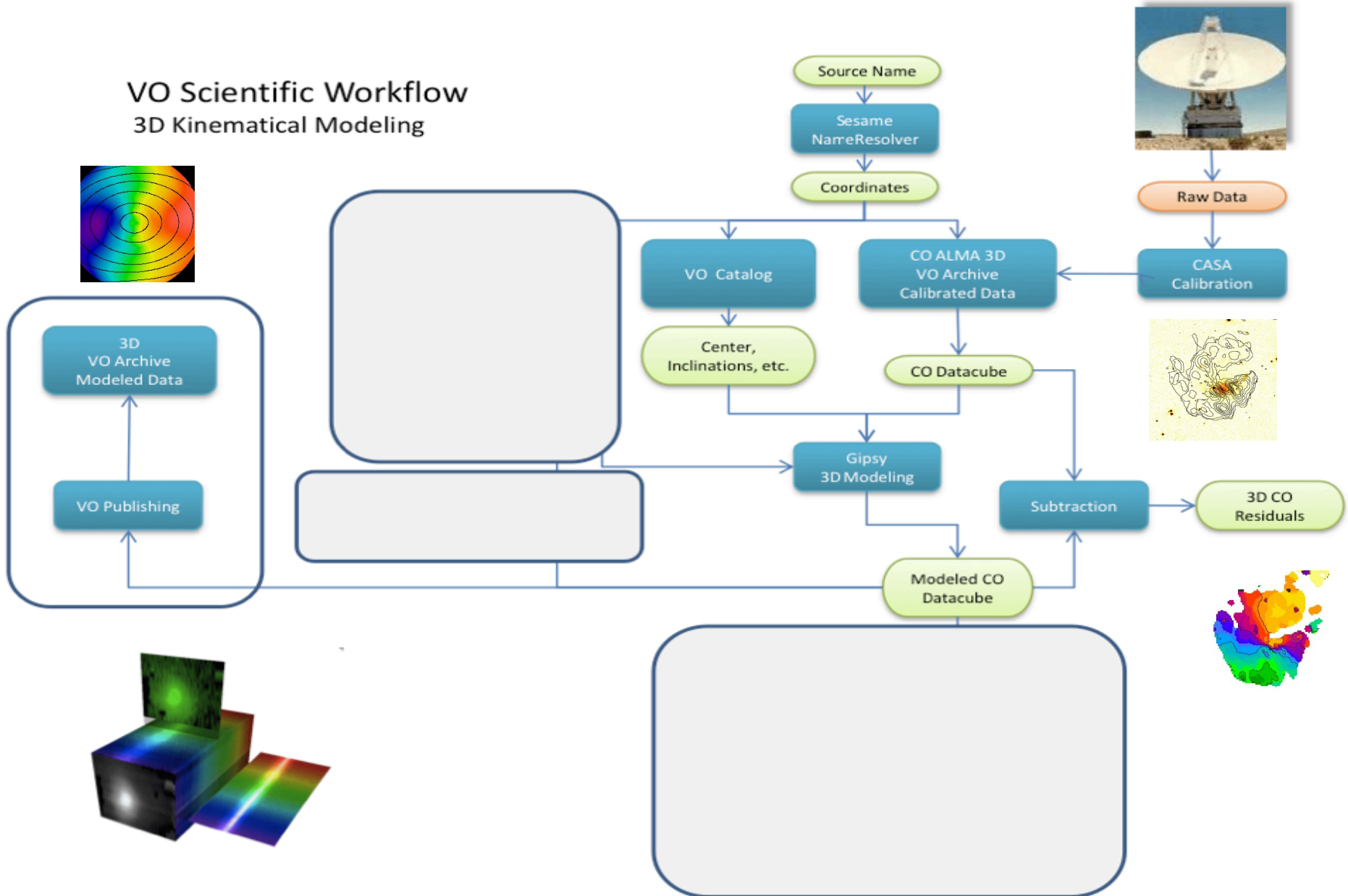
VO Scientific Workflow 3D Kinematical Modeling



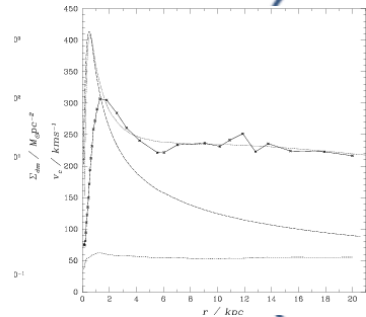
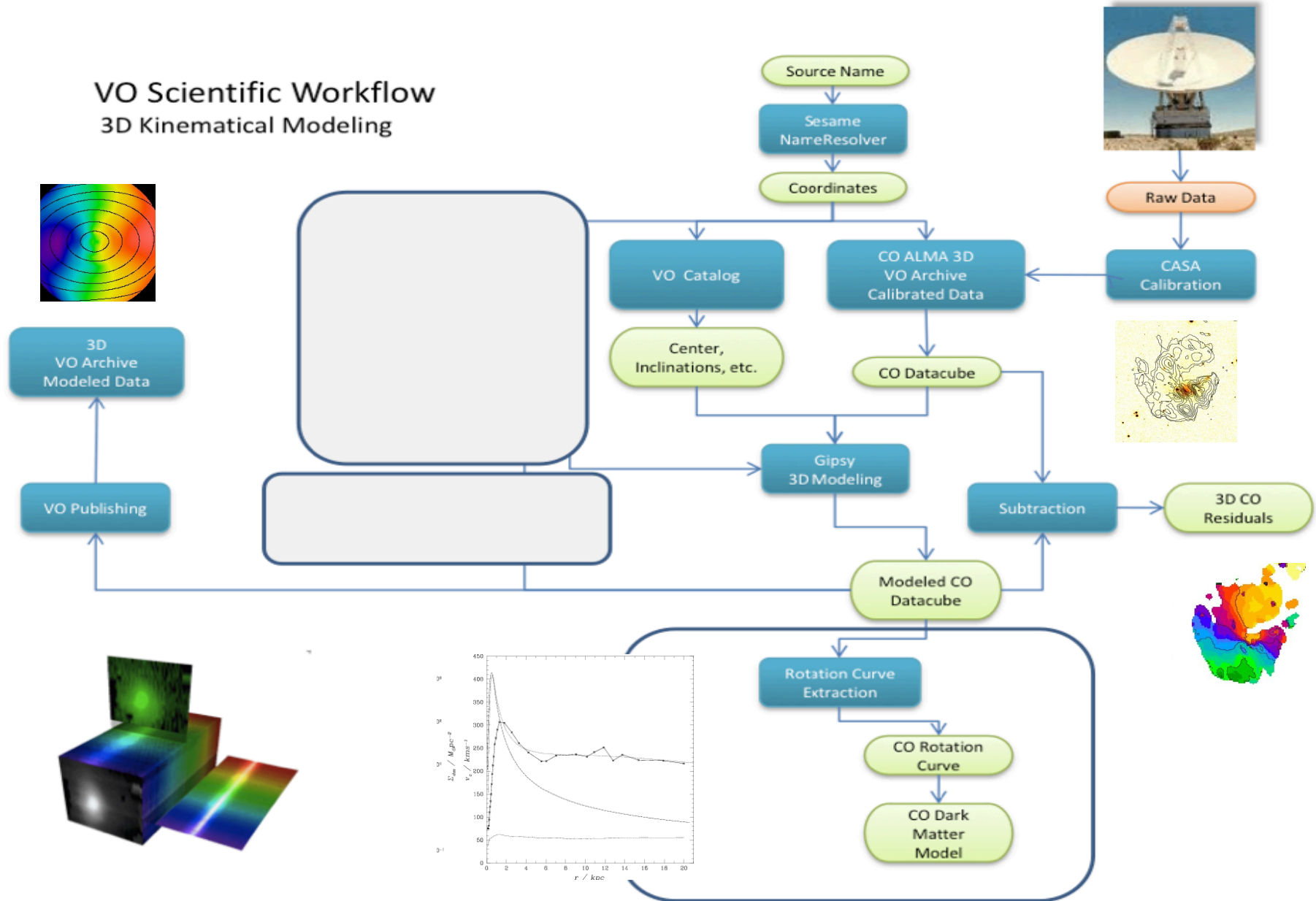
VO Scientific Workflow 3D Kinematical Modeling



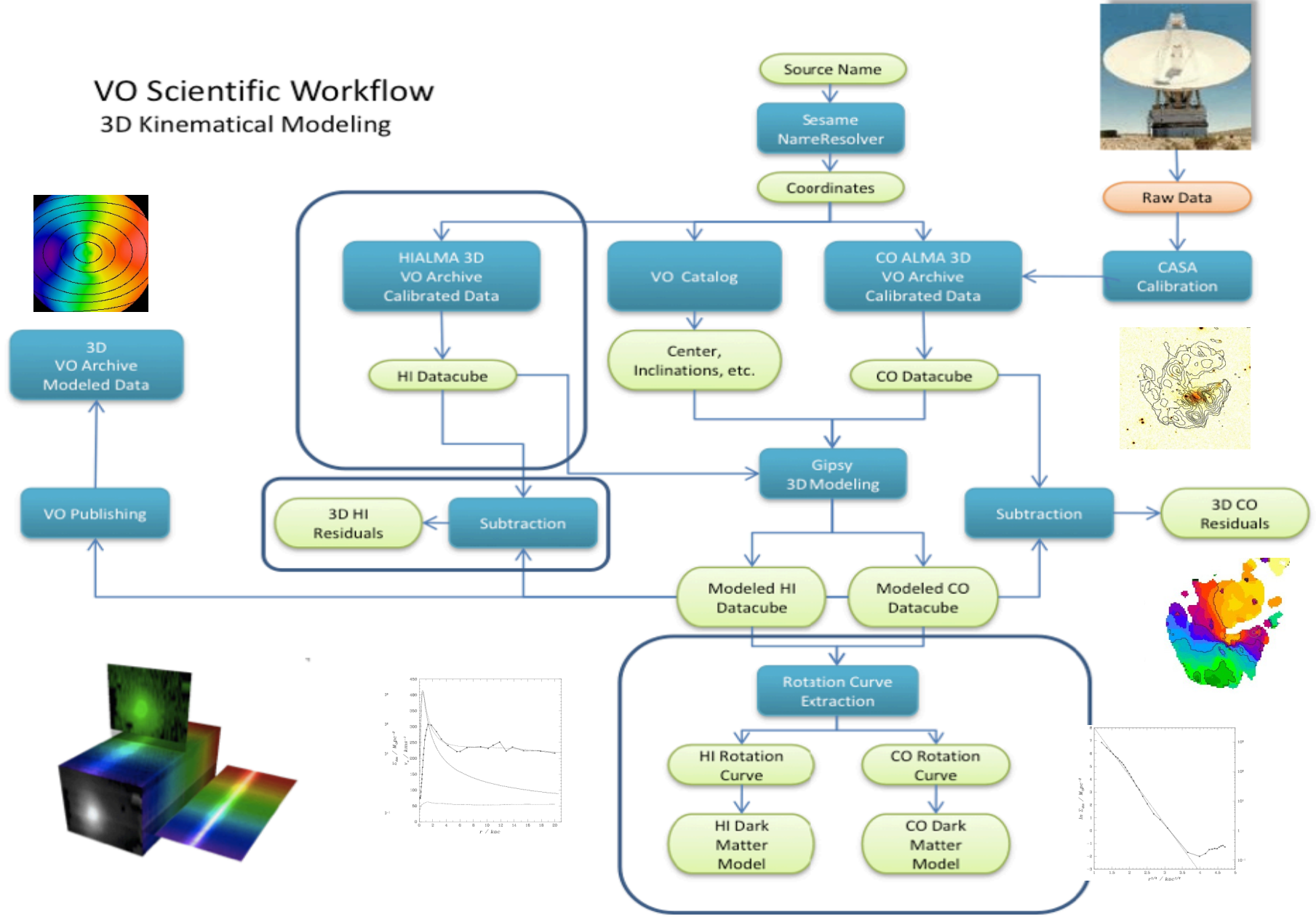
VO Scientific Workflow 3D Kinematical Modeling



VO Scientific Workflow 3D Kinematical Modeling



VO Scientific Workflow 3D Kinematical Modeling



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



Wf4Ever ROBox
Connecting Research Objects in your [Dropbox](#) folder

Sign out
User: Wf4Ever Test User

Manage your Research Objects (ROs)

The Wf4Ever ROBox app connects up your Dropbox and synchronises your Research Objects (ROs) with [dLibra](#)

Getting Started

- ✓ 1. Sign in or connect with your Dropbox account
- ✓ 2. Specify the ROs folder in your Dropbox
- ✗ 3. Wait for the app to sync [Status](#)
- ✗ Ready! View your Dashboard

© 2011 Wf4Ever
[Icons](#) by [Axialis Team](#) and [famfamfam](#)

Works with
Dropbox

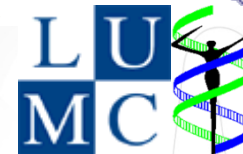
The team



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. University of Oxford (OXF, UK)
6. Instituto de Astrofísica de Andalucía (IAA, Spain)
7. Leiden University Medical Centre (LUMC, NL)



The University
of Manchester



Thanks for your attention !

<http://www.wf4ever-project.org>



WF