



**Wf4Ever: Advanced Workflow Preservation Technologies for Enhanced Science**

**STREP FP7-ICT-2007-6 270192**

**Objective ICT-2009.4.1 b) – "Advanced preservation scenarios"**

---

# D5.1: Astronomy Workflow Preservation Requirements

---

**Deliverable Co-ordinator:** Dr. Lourdes Verdes-Montenegro (IAA)

**Deliverable Co-ordinating Institution:** IAA

**Other Authors:** José Enrique Ruiz (IAA); Dr. António Portas (IAA); Dr. Juan de Dios Santander Vela (ESO); Alexander de Leon (UPM)

| | | | |
|---|---|---|---|
| Document Identifier: | Wf4ever/2010/D5.1/v1.0 | Date due: | 31/05/2011 |
| Class Deliverable: | Wf4ever 270192 | Submission date: | **31/05/2011** |
| Project start date: | December 1, 2010 | Version: | V1.0 |
| Project duration: | 3 years | State: | Final |
| | | Distribution: | Public |

## Wf4Ever Consortium

This document is a part of the Wf4Ever research project funded by the IST Programme of the Commission of the European Communities by the grant number FP7-ICT-2007-6 270192. The following partners are involved in the project:


| | |
|---|---|
| **Intelligent Software Components S.A.**<br>Edificio Testa<br>Avda. del Partenón 16-18, 1º, 7ª<br>Campo de las Naciones, 28042 Madrid<br>Spain<br>Contact person: Dr. Jose Manuel Gómez-Pérez<br>E-mail address: jmgomez@isoco.com | **University of Manchester**<br>Department of Computer Science,<br>University of Manchester, Oxford Road<br>Manchester, M13 9PL<br>United Kingdom<br>Contact person: Professor Carole Goble<br>E-mail address: carole.goble@manchester.ac.uk |
| **Universidad Politécnica de Madrid**<br>Departamento de Inteligencia Artificial<br>Facultad de Informática, UPM<br>28660 Boadilla del Monte, Madrid<br>Spain<br>Contact person: Dr. Oscar Corcho<br>E-mail address: ocorcho@fi.upm.es | **University of Oxford**<br>Department of Zoology<br>University of Oxford<br>South Parks Road, Oxford OX1 3PS<br>United Kingdom<br>Contact person: Dr. Jun Zhao / Professor David De Roure<br>E-mail address: jun.zhao@zoo.ox.ac.uk, david.deroure@oerc.ox.ac.uk |
| **Poznań Supercomputing and Networking Center**<br>Network Services Department<br>Poznań Supercomputing and Networking Center<br>Z. Noskowskiego 12/14, 61-704 Poznan<br>Poland<br>Contact person: Dr. Raúl Palma de León<br>E-mail address: rpalma@man.poznan.pl | **Instituto de Astrófísica de Andalucía**<br>Dpto. Astronomía Extragaláctica<br>Instituto Astrofísica Andalucía<br>Glorieta de la Astronomía s/n 18008 Granada, Spain<br>Contact person: Dr. Lourdes Verdes-Montenegro<br>E-mail address: lourdes@iaa.es |
| **Leiden University Medical Centre**<br>Department of Human Genetics<br>Leiden University Medical Centre<br>Albinusdreef 2, 2333 ZA Leiden<br>The Netherlands<br>Contact person: Dr. Marco Roos<br>E-mail address: M.Roos1@uva.nl | |

## Change Log

| Version | Date | Amended by | Changes |
|---|---|---|---|
| 0.1 | 09-05-2011 | José Enrique Ruiz | Initial Draft |
| 0.2 | 13-05-2011 | António Portas<br>José Enrique Ruiz<br>Alexander de Leon | First three sections added |
| 0.3 | 15-05-2011 | José Enrique Ruiz | Requirements added |
| 0.4 | 18-05-2011 | António Portas<br>Juan de Dios Santander Vela<br>José Enrique Ruiz<br>Lourdes Verdes-Montenegro | General review of document |
| 0.5 | 27-05-2011 | António Portas | Modifications based on QA feedback. |
| 0.6 | 28-05-2011 | Lourdes Verdes-Montenegro | Comments to version 0.5 and QA feedback |
| 0.7 | 28-05-2011 | José Enrique Ruiz | Modifications based on QA feedback |
| 0.8 | 29-05-2011 | Lourdes Verdes-Montenegro<br>José Enrique Ruiz<br>António Portas | Modifications based on QA feedback |
| 0.9 | 29-05-2011 | Lourdes Verdes-Montenegro<br>José Enrique Ruiz | Modifications based on QA feedback |
| 1.0 | 30-05-2011 | Lourdes Verdes-Montenegro<br>José Enrique Ruiz | Modifications based on QA feedback |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

## Executive Summary

This document provides a characterization of the domain of scientific workflows in Astronomy, as well as the lifecycle of astronomical experiments. It compiles as well the initiatives for the development of standards, models and platforms, which enable the preservation of collaborative scientific workflows. The requirements stemming from the astronomical field for the preservation of digital objects which encapsulate data, processes, methodology, and results are derived in 3 ways: 1) a selection of Golden Exemplars, 2) interviews with astronomers from different fields, and 3) identification of user roles for the Wf4Ever collaborative platform.

The Golden Exemplars were selected to cover a wide range of the current astronomical digital experiments, going from 1D catalogues of physical quantities to 3D galaxy data. They make use of data, which is either locally stored and processed by users at their workstations, or distributed, over a variety of external data repositories. Data may be accessed and analysed through Virtual Observatory compliant web services, or with the help of local software and scripts. Interviewees were selected from different fields within Astronomy, and asked to make their particular methodologies explicit. The requirements have been articulated around a set of user roles, where users will be gaining status from collaborators to publishers, that define the way astronomy users interact with their data, and would interact with the Wf4Ever digital objects. We also foresee a transversal role of the entire structure in the role of an Evaluator.

## Table of contents

## List of Figures

## 1. Motivations and aims

One of the current challenges in Astronomy is the *efficient exploitation of the huge volume of data* currently available, whether generated by experiments/observations, or computed by means of numerical simulations. This efficiency is needed in order to ensure the prompt return of the big investments made in terms of facilities to obtain those data, something that clearly the traditional methods of analysis are not currently achieving. This is the reason why scientific workflows are becoming a need in Astronomy. The systematic capture of the scientific process enables researchers to create, re-use, and share the full methodology and agents of an experiment, whilst reducing effort duplication and ensuring repeatability of the discovery process, which finally leads to a complete change of paradigm in the way research is performed.

Ideally, astronomers should work only with data ready for their scientific interpretation, by means of a set of analysis tools that would cover all possible use cases. As we will explain in Sect. 2, this is not the current situation, due to a vast variety of specific needs for the analysis of the processed data, which are not covered by the most commonly used software packages. As a result, the most widespread working methodology combines general-purpose software with specific tools developed within a single research group, and based on the extremely valuable knowledge and experience of a limited set of people. This procedure does not scale with the complexity level and the size of the data being generated by the new facilities, which double every year. Hence reducing reinvention and effort duplication in Astronomy is a must.

Astronomy is a collaborative science, but it has also become highly specialized, as many other disciplines. Workflows will enable astronomers to greatly benefit from each other's highly specialized know-how. Scientific workflows constitute a way to push Astronomy to share and publish not only results and data, but also specially processes and methodologies. Astronomers must hence be encouraged to document their methodology in a clear, concise and modular "universal language". This will:

- Allow the methodology to be preserved and become re-usable by others.
- Increase the efficiency of the collaborative process in Astronomy.
- Contribute to the repeatability of the experiments at any given time by any researcher, independently of their own knowledge level.
- Ease the learning process for young researchers, presenting them with well-organized and documented methodologies.

As we will explain below, at the start of this project most of the required individual infrastructures are either in place, or being developed with the goal of the interoperability of both data and methods, although essentially no scientific workflows exist in Astronomy. Hence the next step is to build a representative set of those, named here Golden Exemplars, and subsequently assemble all of the involved pieces through preserved scientific workflows.

# 2. State of the art

## 2.1 Characterization of the domain of the scientific workflows in Astronomy

Although Astronomy is a strongly digitized science, when compared with the application domain in WP6 the starting point for scientific workflows is in a significantly less mature stage. This requires a larger effort in this project to build up a set of astronomical workflows, but provides the advantage of those instantly benefiting from the Wf4Ever preservation architecture.

### 2.1.1 The Virtual Observatory

It is of relevance here the existence of the Virtual Observatory (VO) infrastructure, an astronomical initiative to provide the community electronic access to numerous sources of information. The VO provides standards to describe all astronomical information resources, enabling standardised discovery and access to interoperable data and services [8, 10]. This information technology provides hence the means, in form of interoperable data services, from which the tasks related with automating the processing of large volumes of astronomical data can be achieved, and will easily be integrated in the development of workflows for Astronomy. The *ASTRONET Infrastructure Roadmap* [5], a group of European funding agencies which came together to reflect and plan the future of European Astronomy, clearly states that “In the long term (ten years) the development of the VO is expected to merge into the standard practices for delivery of astronomical data. The scientific development is expected to be rich in innovations as VO leverages on data mining and semantic technologies”.

The VO community, federated through the International Virtual Observatory Alliance[1], is composed of several smaller initiatives, both national and supranational.

### 2.1.2 Census of astronomical workflows

We have investigated the state of the art of astronomical workflows. Most of the approaches we have found deal with very specialized data reduction pipelines or on the contrary written recipes for highly interactive software where decisions are often made on the visual inspection of data, which hinders a digital serialization of the methodology. We have identified that several efforts have come together in order to explore workflow-based working methodology in Astronomy, although none really in such an extended way as in the Genomics community. The most relevant are:

- Astro-Taverna [6, 11] is a version of the Taverna[2] software for enactment and management of workflows. It was developed in the frame of the UK's AstroGrid[3] project and consists of a set of VO plugins, which use AstroRuntime[4] API to integrate queries to VO services.

[1] http://www.ivoa.net

[2] http://www.taverna.org.uk

[3] http://www.astrogrid.org

- EuroVO[5] is a European organisation to foster the use of VO, and hosts several well-documented recipes ready for re-use. Although fairly unambiguous, these recipes are extremely difficult to digitize since they rely, to a certain extension, on human interactivity with specialised graphical interfaces and VO software. Recipes document the methodology in a non-machine readable way, hence not allowing automation and reproducibility of the experiments.
- The VO France Workflow Working Group[6] [9] is an early user and enabler of astronomical workflows. Although in this case the human interactivity was not so strong, these are too much complex and specialized to enable re-use and collaboration, and thus will not appeal to a broad astronomical community.
- Space missions and large surveys usually provide their data in an advanced stage of processing through the use of the so-called *pipelines*, which calibrate the observations and remove instrumental signatures as well as non-desired external effects. Even if being a significant improvement, this solution is not yet ideal: while for novice researchers or those just non acquainted with the field, these pipelines have a high degree of complexity, expert users typically demand information about the automated processes involved, usually not being fully documented.
- ESO Reflex [7] is a graphical workflow system for running ESO[7] (European Southern Observatory) reduction recipes and related tools in a flexible manner. It allows the user to define and execute a sequence of processes using an easy and flexible GUI. It was focused on ESO pipelines (see above) for astronomical data reduction.
- The HELIO[8] project is a domain-specific virtual observatory for solar physics that is being built, not only with data access and sharing in mind, but with the actual description of the knowledge in the field (via ontologies), and their processes (via workflows). One of its main achievements is having enabled Taverna to run on Grid or Cloud based resources, thus greatly expanding its potential in Astronomy.

None of them has reached a large enough community to promote a collaborative workflow-based environment. Compared to the other application domain selected in Wf4Ever, there is no single entry point for the collection of workflows in this domain e.g. the number of public workflows available in the Euro-VO site -is currently thirteen, and they are not formalized in a workflow description language.

This perspective, in our view, confirms that Astronomy is ready to dive into a scientific workflows-based collaborative working methodology, which needs to be preserved in order to provide re-use and reproducibility of the experiments.

---

[4] http://www.astrogrid.org/wiki/Help/AstroRuntime

[5] http://www.euro-vo.org/pub/fc/workflows.html

[6] http://www.france-ov.org/twiki/bin/view/GROUPEStravail/Workflow

[7] http://www.eso.org

[8] http://www.helio-vo.eu

### 2.1.3. The lifecycle of astronomical experiments

The digitized-based foundation of Astronomy is shown in many scenarios/stages accounting for the astronomical experiments lifecycle, which is described in the following.

*Proposal Submission:* One of the main activities in Astronomy is the access to observational data. Even when more and more current astronomical surveys provide open access to the "digital sky", and the VO gives access to all of them, follow-up observations are many times required, so that applying for telescope time is still needed to answer specific questions. Proposal elaboration requires researching on the available data related to the subject, a comprehensive literature review, as well as access to the detailed capabilities of the instrumentation. These actions rely heavily on online facilities, such as data archives as well as digital libraries. The submission process itself is performed via an online environment where the user is guided through a number of steps before submission. Many of these infrastructures also enable users to closely follow the after-submission process (review, score, result).

*Observations and data reduction:* Due the exponential increase of the automation in astronomical observations, those may be performed by the observatory staff without the presence of the astronomer, or sometimes remotely by the astronomer from a terminal. Observation execution and proper data calibration both generate and require auxiliary data. This kind of information is often located in built-in catalogues associated with the instrument, while in other occasions needs to be retrieved from external online digital catalogues. Astronomical observations are either retrieved under the form of raw datasets requiring further correction factors to be applied or, when automated data reduction pipelines exist, data are provided in a science-ready stage.

*Nature of science-ready data:* There are several types of data collected in Astronomy: images and spectra, recorded in arrays of pixels (mostly at optical, infrared and ultraviolet wavelengths), the intensity of the signal being collected by an antenna or an array of them, later converted into spectra and/or images using Fourier transformations (radio wavelengths), or positions, arrival times and energies of detected photons, from which spectra or imaging data can also be created (high-energy regimes)

*Astronomical Databases:* Most astronomical databases are available online, although in some cases data access may be restricted, usually for a limited proprietary time. Most of the data repositories are compliant with VO standards and provide access to data for an unrestricted type. Online astronomical databases are essentially of two types:

- The *telescope-specific data repositories* (oriented to pointed observations) usually provide raw observation data from the telescope, together with data required for the data calibration, as well as data reduction cookbooks.
- The *large astronomical data repositories* are typically associated with surveys covering a large fraction or even all sky regions. These data have been previously processed by an automated data reduction pipeline, and may provide additional data analysis tools e.g. ability to mosaic fields of the sky. Since most of them take into account VO standards, data re-use provided by these repositories is becoming the norm. For instance, it has been found that the current use of archival data from the Hubble Space Telescope exceeds the use of new data observed by the telescope [1].

*Analysis of science-ready data:* Another aspect to consider is the provenance of different software used by astronomers to handle data. Grid and cluster platforms will be selected when a large volume of data needs to be processed e.g. N-body simulations, which mock a galaxy structure and allow its time evolution. Astronomers also make use of on-line public codes and libraries, however the most common software sources for data handling are locally installed at the researchers' workstation, and might have either a commercial origin —Interactive Data Language (IDL), Matlab— or be freely accessible — such as the Groningen Image Processing System (GIPSY). It is also worth mentioning that astronomers also make extensive use of programming and scripting languages such as Fortran, Python, C/C++, or Java, in order to create custom recipes to analyse the data.

*Publishing:* Astronomy research is published in a paper format via PDF files, which are submitted to bibliographic repositories. These "traditional PDF papers" mainly contain the provenance of the analysis of data in a non-machine readable format. Data results are usually hidden behind plots and there is no current solid platform to allow readers' access to the data supporting the paper. Reused data are also difficult to track, rather than unambiguously identifying these data using dataset identifiers astronomers describe how the data can be obtained.

Most of the published papers are on occasion only accessible through the payment of a subscription of the journal where these were published. In parallel, articles can also be made available, with no costs associated, on an on-line platform, *arXiv*, which supports different disciplines, with the *astro-ph*[9] being the repository for astronomical papers. If in one hand *arXiv* suppresses publication costs, in the other it is a blind repository in the sense that no *peer review* process is associated[10]. This fact can obviously compromise the quality of published material. The common usage of *arXiv* is to make papers available here after they have been accepted for publication in traditional peer reviewed journals.

Bibliographic archives: Digital libraries are usually accessed via an online search engine such the NASA Astrophysics Data System ADS[11] or *arXiv, and* together with specialised literature databases. These bibliographic archives are the main source of literature review. ADS provides interlinking to Astronomical Objects Databases such as Vizier[12], SIMBAD[13], and NED[14] and vice versa. From the presented state of the art in the research lifecycle of astronomical experiments, it can be stated that although most of the components and steps involved are widely digitized, the reproducibility and re-usability of the experiments can only be achieved if scientific workflows exposing the scientific methodology, the data and the provenance in the processes are developed and preserved. For it to be possible all digital artefacts

[9] http://arxiv.org/archive/astro-ph

[10] Different initiatives to provide arXiv with peer-review are being discussed, through mechanisms such as endorsement, or pseudonymous review.

[11] http://adswww.harvard.edu

[12] http://vizier.u-strasbg.fr

[13] http://simbad.u-strasbg.fr/simbad

[14] http://ned.ipac.caltech.edu

related to the research lifecycle have to be available, preserved and easily retrievable, and scientific workflows should benefit of the same privileges acquired by data so far concerning preservation aspects.

### 2.2 Preservation initiatives

Astronomy was one of the first disciplines to benefit from the early developments of Internet and web-based technologies to enable cross-linking of resources across archives [2]. Seventeen years ago, thanks to a collaboration between the NASA Astrophysics Data System (ADS) and the Centre de Données Astronomiques de Strasbourg[15] (CDS), it became possible to follow the path from a list of articles to the abstract of an article, and from there to the list of astronomical objects described in that article, and to a set of measurements on one of those objects. Fourteen years ago, again thanks to the collaboration between the ADS and several major data centres, it became possible to actually go from an article abstract to the actual observational data used to write the article, and then back to all publications describing each observation. These connections have enabled astronomers to use the search capabilities of any of the main archives to locate datasets or publications of interest, and then follow the appropriate links to find related information provided by another archive. This procedure still requires users to click through all the data links provided in the list of returned papers. Automating this activity is currently not practical because of the current lack of semantic and contextual information among resources. The Linked Data effort[16]aims to build a global graph of resources built on the architecture of the web. The fact of exposing metadata related to these astronomical resources and the links among them following Linked Data principles may allow people and applications to transverse, analyse and compute over this global graph.

In the last year, the US Virtual Astronomical Observatory[17] (VAO) Data Curation and Preservation Group has launched an initiative [3] to create an infrastructure supporting curation, discovery and access to VAO resources. The two main objectives of the project are to capture and describe the linkage between astronomical objects, archival datasets from surveys and catalogues, observing proposals and publications and to capture and describe as much as possible the lifecycle of the research process, thus enabling to track the provenance of both data and publications produced by researchers. Thanks to this backend server infrastructure a new ADS search prototype is being built in order to expose these links, making it easier for astronomers to explore the space of astronomical concepts and phenomena using an iterative process through an interface which exposes key relationships among them [1, 4].

[15] http://cdsweb.u-strasbg.fr

[16] http://www.linkeddata.org

[17] http://www.usvao.org

# 3. Golden Exemplars

In the following we describe the three Golden Exemplars of use cases proposed in this project, which, together with the interviews described in Section 4 and the User Roles foreseen for the Wf4Ever platform described in Section 5, have been used to derive the astronomy preservation requirements. The Exemplars have been selected to cover a wide range of the present astronomical digital experiments — from 1D catalogues of physical values to 3D modelled galaxy data, both locally stored and processed in the user workstation, or distributed over a variety of external data repositories, accessed and analysed either through VO compliant web-services and tools, or with the help of local software and scripts.

## 3.1 Propagation of physical quantities

This use case deals with large sets of tabular data (1-D catalogues of numbers) curated by the user, that rely on basic experimental values coming from external data repositories and combined by means of mathematical equations. Updating of the external data repositories has an impact on the preservation of the digital experiment, since it is essential to know how and when these databases are updated, so that the propagation of these changes through the existing internal relation among the data can be triggered and registered. This question is particularly relevant due to the increase in the number of databases yet to be exploited.

A practical example is described in Figure 1. Here the aim is to compare values of apparent magnitudes and distances of an ensemble of objects (galaxies), against values stored in VO databases. From these values intrinsic luminosities are derived for each object based on a mathematical dependence, so if any of the two first quantities are updated, the third one also should be. The first step of the process is to query the VO databases using object's names and positions provided by user local ASCII file as shown by module 1.0. The retrieved values (magnitudes and distances) will be then compared (module 1.1) with the user locally stored values and if found greater than a threshold, the users' local values of magnitudes and distances are updated (module 1.2). The next action to take place is to derive new luminosities (module 1.5) using the updated values of magnitudes and distances and making use of local scripts embodying the mathematical relationship (module 1.3 and 1.4). The final action of this workflow will compare the calculated values of luminosities against those owned by the user (module 1.6). Again using a difference threshold, luminosities may or may not be updated locally in the user tabular data (module 1.7).

The data collection to be updated is usually stored in local ASCII files or in a database engine e.g. MySQL. The external data needed to calculate the new values for this data collection may come from VO services, but the parsing of HTML pages or ASCII files should also be considered. The execution of the workflow is done in the local desktop of the user with Internet access to the referenced data sources, a scripting environment may be needed for the parsing and updating scripts included in the workflow. Original before update and final data are stored locally, while the intermediate values needed for the calculations are gathered from external sources.

Preservation of the experiment is impacted in several parts of the process:

- Updates in the external repository need to be versioned, which is not always the case currently. This would permit to keep a timeline of the evolution of these quantities, allowing the extraction of potentially relevant scientific information.
- The user needs to be advertised of each of these updates, in order to be able to decide on the update of the tabular data. The reason for this update needs to be preserved, e.g. it could be based on the differences found in values obtained from the new data in a dry run.
- Modification of the mathematical equations involved, e.g. due to new values of physical parameters involved in the formula, have to be documented, often providing the bibliographic reference on which the update has been based.
- Al the above items combined would give place to a new version of the full catalogue owned by the user.
- In the current practice, the inability to link the several actions in an automatic fashion does not allow an easy reproducibility and re-use of the experiment. Moving this experiment into a workflow environment will automate the entire process and provide processing modules that may be re-used in similar experiments dealing with triggered propagation of any interlinked astronomical quantities, exposing the provenance of highly valuable data.

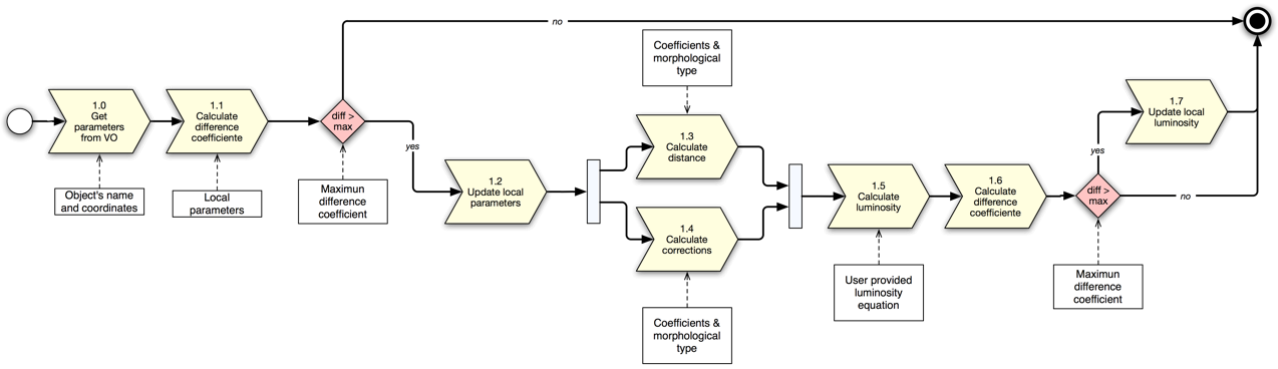



Figure 1: Propagation of quantities. The update of galaxies properties like magnitudes and distances triggers the re-computation of mathematically dependent quantities such as luminosities.

### 3.2 Extraction of galaxy samples from 2D data

In this case we focus on extracting a list of potential objects (galaxies) from a digital 2D (spatial coordinates) image coming from an external digital repository, fulfilling very specific criteria on their spatial distribution in the sky. The object extraction can be performed providing the digital image to a specific software (SExtractor[18]) running locally, or through external web services. The selection criteria are based on physical properties of the objects, such as magnitudes, or apparent sizes. The quality of the results of the experiment is tested (and eventually improved) by cross-matching the list of objects obtained via this source extraction with those coming from a query to VO tabular data archives,

[18] http://www.astromatic.net/software/sextractor

Figure 2 describes an embodiment of this use case. The main goal is to obtain a list of galaxies within a given radius from the target based on specific criteria. Querying the VO databases using the galaxies names or positions, as well as the selection criteria are performed in module 1.0, while extraction from a set of 2D digital images using local software or web services is performed in module 1.1, taking into account the same selection criteria and a software execution parameter list. Cross matching the results originated by two different methods allows assessment and eventually improvement of the results (module 1.3).

Preservation of the experiment is impacted in the following parts of the process:

- Updates in the source of the 2D data need to be versioned, as for the previous use case. This is more often found in Astronomy, e.g. one of the main sources of 2D data for galaxies, the Sloan Digital Sky Survey (SDSS) keeps track of all releases.
- The exact criteria used for the selection of the objects needs to be specified in a way that can be directly reused by another astronomer to apply to the same or a different dataset, either in the same exact way, or with modifications with the purpose of comparison. Those criteria are usually given only in the PDF publication as a set of sentences, and not always in sufficient detail to allow repetition.
- The methodology for extraction of sources from an image still requires some user intervention, and visual inspection. Improvements to this example by moving it to a workflow environment focus not only on the possibility of automation, reproducibility, provenance tracking and re-use of the experiment but also on the capability of comparing results extracted both from a list of positions or a digitized 2D image.

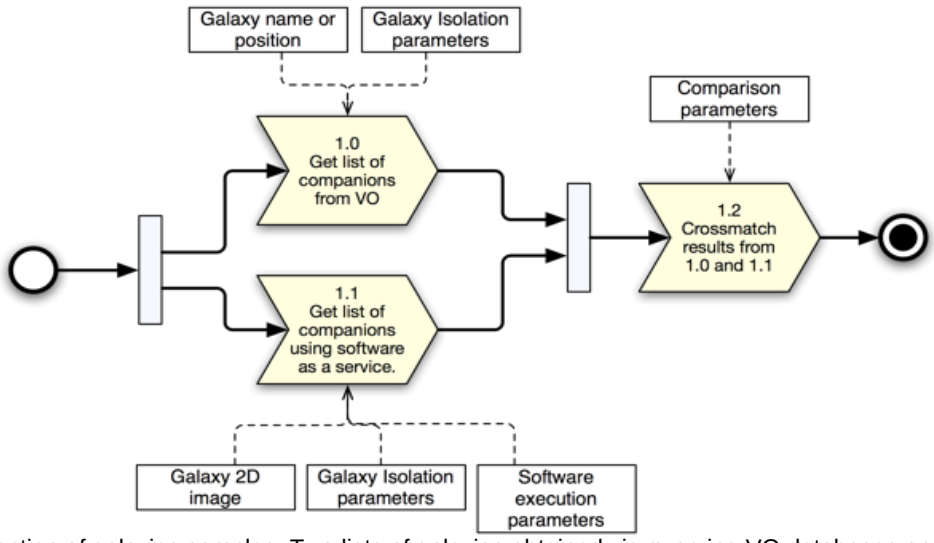



Figure 2: Extraction of galaxies samples. Two lists of galaxies obtained via querying VO databases and via source extraction of a 2D image from the sky are cross-matched.

### 3.3 Modelling of 3D data of galaxies

This use case addresses the issue of new and upcoming large astronomical facilities providing high volume of complex data, with the issues inherent to the need of large data transfers. The data here exploited

are 3D binary cubes (typically big sized) with two spatial dimensions and a third one associated to the velocity of the object. Those datasets will be analysed by means of modelling tools, and the produced models would be stored for comparison with the original data or similar data at different wavelengths in order to extract physical information on the targeted object.

In Figure 3 we present an example of this use case. First, a user file containing information on a large ensemble of galaxies names and/or their positions is used to query a VO archive. The VO archive will provide a collection of 3D data cubes at a specific wavelength (module 1.0). For each observed data cube a modelled data cube will be generated via a specific web service modelling task, using an extra user provided configuration file to set the modelling parameters e.g. orientation angles in the sky, 2D images of galaxies, etc. The process will loop through the difference between the model cube and the observed one until reaching a difference or residual threshold (modules 2.0 and 2.1). After this stage a rotation curve (1D data product that provides kinematical information on the analysed galaxy) will be also extracted from the final versions of the modelled data cubes (module 3.0), again making use of web services tasks, and will then be stored on ASCII files. The results of the experiment (modelled cubes and rotation curves) will be stored in the same VO infrastructure (module 3.1), closing the loop which connects observed and modelled data cubes through modelling parameters used in the experiment and enriching the whole with new entities like the rotation curves.

The execution of the workflow should rely on external web services provided by online archives, since the involved cubes are difficult to move from these archives to the user desktop due to their large volume. Intermediate cubes are generated on-the-fly through modelling tasks, which require as input a configuration file setting modelling parameters. The final products might be either 1D or 2D, and may be retrieved in the local environment of the astronomer.

Preservation of the experiment is impacted in the following parts of the process:

- Updates in the source of the 3D data need to be versioned, as for the previous use cases. For these complex datasets this is usually the rule, especially since raw data are in most cases provided to the astronomer. In the future it is however expected that processed data will be more and more provided, so that a special care will have to be taken for those to be properly registered, allowing for repeatability.
- The results of the modelling of the cubes depend on the set of input parameters defined by the user. Part of them relies on external (in general VO) resources, as positions or geometrical parameters, while others come from inspection by the user of the dataset (extent of the galaxy). Those set of parameters need hence not only to be registered for each modelling run, but also if relying on external resources those should ideally be versioned for proper tracking.
- In the current practise all the processes run locally, or in a cluster environment, and haven not been migrated to the VO web services platform. In addition to this, the current lack of interoperability between different tasks in the local software still requires some degree of user intervention. Bringing this example to a workflow environment will enable reproducibility through automation of these processes. Moreover, it will also expose the scientific methodology used in the experiment with the parameters used for the

modelling, and since all the entities involved in the workflow (external web services and data) are provided by online archives, reproducibility of the results is expected.

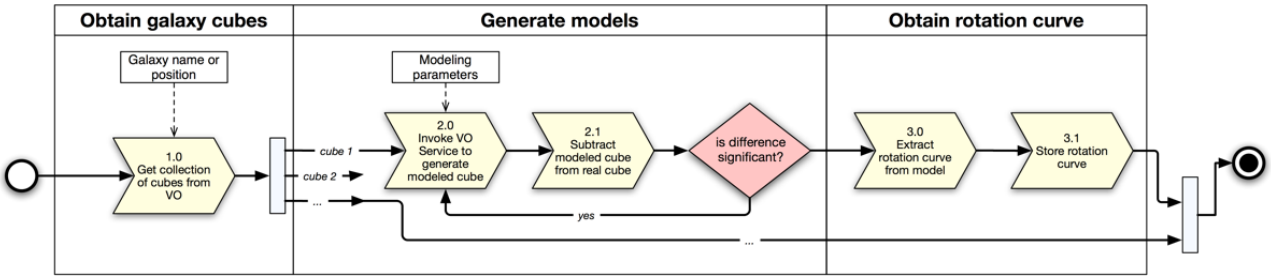



Figure 3: Modelling of 3D data of galaxies. Ensembles of 3D data cubes provided by VO databases are modelled and 1D/ 2D information from these models are retrieved locally.

# 4. Selective interviews sampling

The main user requirements have been compiled based on the Golden Exemplars, since those have been selected aiming to be representative of the most common methodologies used in Astronomy, Still we have made an effort to extend those requirements by means of selected personal interviews, where we tried to identify common patterns on the current working methodology among astronomers to impose additional requirements on the models for Research Objects (RO) and on the Wf4Ever platform. A comprehensive description of a RO has been presented in deliverable D2.1.

## 4.1 Method

We have interviewed five astronomers working at the Instituto de Astrofísica de Andalucía (IAA), who have been deliberately chosen to work in different areas of Astronomy— from the formation and physical processes occurring inside stars to those working in galaxies at different stages of their evolution.

The questions made cover general aspects of astronomy research. With this set of questions we aim to target fundamental aspects which we believe to be relevant towards the development of an infrastructure where preservation of digital experiments and their related scientific methodology will greatly improve collaborative work. We list below those questions and their connection with the aims of this deliverable.

- How the success of research is measured?

  This is relevant in order to infer how the community will assess the value of a RO. There will be an intrinsic need to rank preserved RO in order to allow ranked searches, in a similar way as the current practice of ranking publications with a high number of citations.

- How is the work organised locally? How is it accessed remotely?

  The goal of this question is to get insights on the organizational structure used in storing all artefacts involved in a specific experiment. Will people working on different areas share a common work structure? This will provide valuable information in potential functionalities of the Wf4Ever platform, how experiments should be preserved and infer information on topics such as Intellectual Property Rights, mixed ownership, access control and authentication in the RO.

- How to address the quality of the data in catalogues/articles?

  Here we want to address how the community will assess the quality of the different components of a RO. In this topic we distil information on integrity and authenticity aspects of RO which will feedback deliverable D4.1.

- Does the current “PDF paper” provide enough information?

  The objective of this question is to trace the gaps of the current publishing practice in Astronomy. This will help to build an RO model that captures, exposes and preserves the methodology and data interconnected in the workflows, and the provenance aspect in the experiment.

### 4.2 Results

- How does one measure the success of research?

  There is a personal perspective of success, where a scientist enjoys addressing, analysing problems and finding solutions. From a scientific community perspective, success is a merger of factors: published papers with number of citations, invitations to give talks in meetings, accessibility to funds. Recently, alternatives to a simple number of citations have been introduced. For instance, the h-index (Hirsch index) combines a measure of the total number of papers and the total number of publications (e.g. h=25 means 25 papers with at least 25 citations). Another alternative is to give more relevance to first author papers, or to ask researchers for their best 5 best publications. These answers clearly indicate that success of research is most of the times directly related with the publications.

- How is the work organised locally? How is it accessed remotely?

  The file system was the main organisational unit. Work is organized in directories with 3 distinct levels:

  - The first level divides the work into the different teams the researcher is involved with.
  - On a second level, each individual project is assigned to a single folder.
  - A further subdivision organise folders in 5 distinct areas: data, software, scripts, results and papers.

  Of particular interest in the organizational process is the search of information from this hierarchy. The interviewees were asked how they searched for a specific file. These are emerging conclusions:

  - The structure of the organisation of work is essentially the same from subject to subject, which allows users to know a priori, the location of a file.
  - The file naming convention is quite relevant, since it can encode a sequence of actions associated with the file, also providing a hierarchy amongst files under the same directory. "Filename_a_b" is different from "Filename_b_a", since actions a and b were applied in a different order. Hence, we can see that a part of processing provenance is encoded in the filename itself.
  - They make use of operating systems' facilities for file searching. This is especially useful for file name searches, and recently for file content. However, finding portions of scripts revealed to be more difficult, since this involves proper documentation inside the scripts.

  The remote access of work has also been identified as relevant. Although FTP, SSH, SFTP and Telnet are still used to access/download files, the use of Dropbox —or Dropbox-like alternatives— is becoming more generalised. Files that are thought to be important are dropped in Dropbox and can be accessed from multiple computing devices, while keeping the same

structure, and allowing data synchronisation. This feedback provided by the interviewees is of important relevance towards one of the initiatives of the architecture working package of this project. In fact the Dropbox RO Connector app (ROBox) has been developed to allow for a Dropbox folder to be integrated and preserved into the Wf4Ever platform.

- How to address the quality of the data in catalogues/articles?

  Our interviewees take full advantage of platforms such as ADS digital library, as well as Vizier and Simbad catalogues. Science-ready processed data from space missions e.g. Spitzer, HST, etc. are seen as trustworthy given the quality of the calibration via their dedicated pipelines. Given the large volume of unexploited data in catalogues and data repositories, there are many astronomers that no longer find necessary to apply for new observations, basing a large fraction of their research on the available databases. The latest release of each survey is preferred as data pertaining all-sky surveys are constantly being refined, with the pipelines to generate them improved, which leads to better quality data with each new release. When publishing articles there should be an unambiguous mention to the release version of the data used

  This raised the question on the assessment of the quality of published papers, summarized below:

  - A more dynamical review system on published material should be available for the community. E.g. it could comment on a paper and rate it. This would lead to quantification of the amount of information/ discussion generated by a paper, and could provide an additional measure of its interest.
  - The peer review system as it is, is very centralised. It only allows seeing the point of view of the authors and referee. Another relevant issue is that within Astronomy papers are only reviewed by one peer, while other fields can have from 3 to 5 reviewers. This can lead to lower quality publications.

- Does the current “PDF paper” provide enough information?

  All the interviewees agree that the traditional "PDF published paper" does not contain enough information to evaluate or reproduce an experiment, willing more information beyond a static document. All interviewees reveal that they have been contacted / or contacted other authors in order to get access to data from published material. There should be tools available to the community to argue beyond the peer review process, so that published papers could be questioned continuously and communication with the authors maintained. The lack of free access to data is mainly due to intellectual property right issues, or internal policies within a group. Access to private data by third parties will mostly rely on a mutual trust relationship, where there is a clear purpose for the use of data. The existence of a simple tool in ADS called

Dexter[19] which will allow you to retrieve approximate data from a simple plot, not its provenance, is a good example of a simple initiative that enables the reusability of data.

[19] http://adsdoc.harvard.edu/abs_doc/help_pages/dexter.html

## 5. User roles

We have identified and we propose the following user roles for the Wf4Ever platform. This collaborative platform should differentiate the specific basic stage of the RO lifecycle as presented in D2.1: Live Objects (mutable content), Publication Objects (immutable content and no preservation guaranteed) and Archived Objects (immutable content and preservation guaranteed), where users will be gaining status from collaborators to publishers. We also foresee a transversal role of the entire structure in the role of an Evaluator. In the following we describe the different user roles definitions.

*Collaborator*

The *collaborator* is working in a group that uses the Wf4Ever collaborative platform and tools, as the ROBox developed inWP1. By means of the ROBox and future evolutions, he/she takes advantage of the seamless integration of his/her own working environment into a sharing and ubiquitous platform. Hence newcomers do not even know that they are dealing with Live Objects.

*Reader*

The *reader* is looking for related works, state of the art, in his field of research. He/she skims the titles, keywords, themes and abstracts of the Publication Objects and Archived Objects, sometimes delving into its content, andmay be interested in re-use or comparison. Novice users of the platform will be readers and will gain new roles as they become more familiar with RO, scaling to comparators, (re)users and publishers.

*Comparator*

The *comparator* is looking for ROs similar to those he/she is using at present, and wants to know whether the work he/she is doing has been already published, and if there are RO that deal with similar tasks to those present in his/her own research, for re-use in a later stage. The comparator may come from a very different scientific domain e.g. workflows for statistical tasks in Biology may be also very useful for an astronomer. And may also take the role of re(user) if finding RO components suiting his/her specific needs.

*Re(User)*

The *re(user)* knows how to work with ROs and scientific workflows, and how to extract and replace modules from one workflow and insert them into the one is using. Most of the times has also taken the role of comparator, other times it is another colleague who took the role of comparator and the re(user) just goes right through the selected RO someone else has identified.

*Publisher*

The *publisher* wants his/her work and his/her group to be known among the community. This person is the main author of the Publication Object, though the real person who undertakes the action of publishing may be one of the co-authors. He wants his/her digital experiment to be known, and consequently the work done in his/her group, among the community. In order to achieve this it should be easy for the Publisher to place his RO in the right "*drawer"* e.g. extragalactic astronomers would a priori not be interested in solar system related ROs.

*Evaluator*

The *evaluator* has enough experience in his/her field of research to evaluate and score a Publication Object. He/she can provide comments and suggestions to improve the methodology showed in the RO from a scientific and also technical point of view.

# 6. Requirements

Requirements for the Wf4Ever platform have been distilled from the previous sections in this document, where three Golden Exemplars have been proposed, selective interviews to astronomers have been performed and user roles have been identified. Considering this feedback and the general goals pursued in the Golden Exemplars, we present below a list of user requirements on the Wf4Ever platform and identified digital entities that should be taken into account when developing the model of the content of an astronomical RO.

## 6.1 Wf4Ever platform

| As a … | I want to ... | Comments |
|---|---|---|
| General User | Identify and choose the platform interface. | The collaborative platform should differentiate the specific basic stage of the RO lifecycle (Live Objects, Publication Objects, Archived Objects). Since the potential actions and decisions to be taken by the user depend on these different levels of the RO lifecycle, the user should have the possibility to choose whether working in a living RO platform or consulting the digital library of RO. |
| Collaborator Re(User) | Build, manage, re-use and enact scientific workflows. | Given the current state of the art of astronomical workflows as characterized in this document, astronomers will need a tool for these tasks, as it is already the case with other scientific communities using software as Taverna. This workflow management tool should integrate access to both VO data repositories and VO compliant web services, considering communication and interoperability with widely used VO software as Topcat[20], Aladin[21] and VOSpec[22]<br><br>The re(user) would also like to have the |

[20] http://www.star.bris.ac.uk/~mbt/topcat

[21] http://aladin.u-strasbg.fr

[22] http://www.sciops.esa.int/index.php?project=ESAVO&page=vospec

| | | |
|---|---|---|
| | | possibility to concatenate existing workflows and to extract useful components from RO (data, processes, web services, scripts) in order to use them in his own research. |
| Collaborator | Create a and share and RO in a research group | Seamless integration of newcomers in the Wf4Ever platform |
| Reader<br>Comparator<br>Re(User) Evaluator | Search and Retrieve RO | The way searching and retrieving of research objects is performed would not be very different from current procedures for searching bibliography, allowing search in fields like author, abstract, keywords, publication dates, etc. Since a primary goal is to bring astronomers into the platform, thre are more chances of succeeding if proposing well-known friendly procedures. This is particularly relevant, and the concept of “familiarity” is a shared concern amongst astronomers and biologists (as mention in D6.1). In addition, providing the user with a good tagging system is also a common necessity amongst biologists and astronomers. |
| Collaborator | Update and Delete an RO or part of its content. | |
| Publisher | Publish a RO | The way to define how the RO is "advertised" or indexed should be an efficient and easy task for the publisher. |
| Collaborator | Manage access to RO | In general Publication Objects may be freely accessible to all users for re-use and Live Objects should be subject to more restraint access privileges. Functionalities for access management should be provided, modular and flexible enough to deal with users and groups, and at different levels of granularity in the components of the RO. |

| | | |
|---|---|---|
| Collaborator<br>Re(User)<br>Evaluator | Store and identify different versions of the RO or its components | Versioning may depend on whether we deal with Live Objects, Publication Objects or Archived Objects. When working with Live Objects, the platform should allow the users to choose not to backup full versions of the RO when a modification is made in the components (data or processes involved). Because of the big volumes of data involved it may be reasonable not to store multiple versions of the same RO. On the contrary, it may be useful in this case to allow versioning of individual components of a RO, those not suffering of heavy sized issues. For example, automatic versioning of metadata (contributors, access privileges, modification date, etc.) may be implemented to trace the timeline provenance of these metadata. Publication Objects or Archived Objects may depend on external resources that may evolve or being suppressed with time. In this case versioning of these RO may be useful to check their integrity and authenticity. |
| Collaborator<br>Publisher | Seamless integration of metadata | As opposed to other users, astronomers are not familiar with Linked Data principles and RDF syntax; because of this all declarations of metadata information related to a RO should be transparent at the moment of creating a Live Object, publishing a Publication Object. The platform should provide a protocol or interface for an easy publication of a RO with assistance on metadata, with the aim of not discouraging contribution. The ROBox developed in WP1 is a very good example of how research objects and users are seamlessly integrated in the Wf4Ever platform in an unobtrusive way. |
| Evaluator | Rate, comment, recommend RO | These actions apply not only the whole RO but also to specific components of it, relating to |

| | | |
|---|---|---|
| | | quality criteria in reproducibility, repeatability of the results and re-usability and usefulness. |
| Re(User) | Take advantage of the modular and decomposable nature of RO. | It should be possible to import other RO or their components into a Live Object. In this respect, RO should not be closed entities but fully decomposable with seamlessly accessible components. |

### 6.2 Content of a Research Object

We have identified the following entities related to the astronomical research lifecycle and that may be considered in the study of the content of an astronomical RO.

*Related information contextualizing the experiment:*

- Small description/abstract: clear summary of the scientific purpose of the RO.
- Precedent related studies: references to both literature papers with respective ADS bibliographic bibcodes and reference to other existing relevant ROs.
- Problems to solve and strategy to follow: description of the specific goals and methodology being used to tackle the problems involved.
- Expected results: which are the expected results from pure scientific point of view. This component is essential if we want to perform a control on the auto consistency of the RO in order to avoid or detect contradictions.
- Acknowledgements: facilities, instruments, other research objects, bibliography, etc.

*Metadata for the characterization of the experiment*

These data may come from external sources and may not be reachable after a period of time. The following information needs not only to be preserved but also to be accessible at the moment of its creation with queries on external data catalogues and repositories:

- Observing proposal and/or research project.
- Instrument used.
- Observer, research group and institution.
- Meteorological conditions at the moment of observation (opacity, humidity, etc.)
- Calibration parameters/data

*Original, Intermediate and Final datasets*

These data may consist of high volume data collections reaching sometimes tens of GB. Links and references to data repositories are needed, and data policy access issues would need to be solved. Data would mainly fall within the following categories

- Science ready input data: used as input of an astronomical workflow, these data may be collections of tabular values, binary images, spectra or data cubes. They may come from observations realized by the user or from on-line archives.
- Intermediate data: data involved in the workflow as the output of a process and the input of a subsequent different one. It may be useful to keep them since we should provide the possibility to enact only specific parts of the workflow.
- Final data results: final results provided by the workflow, they may be of the same nature as the science ready input data. Some of them may be retrieved in the local environment of the user or published and preserved in online archives.

*Operations involved in the experiment*

The relations among every single element in this list and the datasets involved are stored as a digital reproducible workflow and its representation

- Web services, where sometimes access and authentication accounts issues may need to be solved.
- Standard VO web services, they mostly provide access to data, though a new family of VO web services providing complex analysis tasks is under development.
- Scripts in high level programming language (Python, IDL, Fortran, Perl, shell scripts) mostly dealing with parsing and small operations on data.

*Discussion*

This point covers the standard PDF publication issued from the experiment where results are discussed and conclusions are exposed. The content should be as accurate, clear and concise as in a traditional science article. Plots should be easy to inspect and data behind them should be a visible part of the RO that reproduce them.

## 7. Discussion, conclusions and future work

The characterization that we have performed of the lifecycle of astronomical experiments shows that Astronomy is a science where most of the resources used in current research are now in digital form and are often available on the web. The involved entities are however not properly interlinked yet, though some of them in the process of being done. We have identified several initiatives for the development of standards, models and platforms, which enable the preservation of collaborative scientific workflows.

The combination of three approaches (exemplars, interviews and user roles) has provided the user requirements for the Wf4Ever platform. It is particularly relevant to consider the different user roles since they yield the basic functionalities of collaborative research where astronomers should benefit from the high degree of specialization they have acquired inside their field instead of taking this fact as an obstacle for making better science. As an analogy with the economics that rules our world, in an infrastructure where astronomers would feel themselves safe to freely trade their scientific methodologies, they would concentrate on areas where they have a comparative advantage, and avoid areas where they have a comparative disadvantage. This efficient collaboration market based on fruitful transactions of ideas will greatly improve astronomical research.

Given the role of the VO in Astronomy, and in particular that it is achieving a whole set of standards for the interoperability of data and services in this infrastructure of astronomical information, the next steps in the project have clearly to integrate the current VO initiatives as well as provide continuous feedback from Wf4Ever. This is especially relevant since a new generation of facilities is giving birth to a new kind of astronomical archives where observed data are not transferred to the user, web services are offered for their analysis and the research is moving from local desktops to online tools living near the data.

In this context of distributed storage and computing, since the data and processes are completely external to the astronomer, the repeatability of the experiments and reproducibility of the results are guaranteed as far as these remain unchanged. The proliferation of automated surveys yielding vast amounts of observations at all wavelengths is conducting Astronomy into a panchromatic research where interoperability among all these archives is needed, and in particular among the analysis tasks they provide by means of web services. In this cloud of services and data, web services should benefit of the same privileges acquired by data so far including interoperability, curation, preservation and discovery. This can only be achieved with the development of standards and models through the involvement of the different initiatives identified aiming preservation of collaborative research. Some issues to be solved deal with versioning, authoring and metrics for quality e.g. based on popularity, statistics of use, uptime server logs, etc.; other more technical with user authentication, permissions and platform licences.

All the ingredients are there for the astronomical community to benefit of a disruptive working methodology based mainly on the re-use of digitized RO that pack all components needed in the research lifecycle. However our study of existing or in progress astronomical workflows reveals a lack for digitized astronomical workflows describing the scientific methodology and accounting for analysis provenance of the experiment. Hence, the first step and our main goal is to engage astronomers in a living research

collaborative environment (Wf4Ever platform) that allow them to create astronomical workflows from re-use or from scratch and manage RO in a seamless manner. For this to happen we plan to develop the three proposed Golden Exemplars and adapt easy digital workflows in order to provide astronomers not only with tools but also with use cases for re-use

## 8. References

[1] A. Accomazzi, *Astronomy 3.0 Style*, ASP Conference Series, 2010, pp. 1-9.

[2] A. Accomazzi, *Linking Literature and Data: Status Report and Future Efforts*, 2011, p. 9.

[3] A. Accomazzi and R. Dave, *Semantic Interlinking of Resources in the Virtual Observatory Era*, ASP Conference Series, vol. 442, 2011, pp. 1-10.

[4] A. Accomazzi, M.J. Kurtz, and S.S. Murray, *Towards a Resource-Centric Data Network for Astronomy*, Proceedings of Science, 2010, p. 6.

[5] *The ASTRONET Infrastructure Roadmap: A Strategic Plan for European Astronomy.* ISBN: 978-3-923524-63-1. Ed: Michael F. Bode, Maria J. Cruz & Frank J. Molster.

[6] K.M. Benson and N.A. Walton, *AstroGrid : Taverna in the Virtual Observatory*, Memorie della Societá Astronomica Italiana, vol. 80, 2009, pp. 574-577.

[7] R. Hook, M. Ullgr, M. Romaniello, S. Maisala, T. Oittinen, O. Solin, V. Savolainen, and P. J, *ESO Reflex : a graphical workflow engine for data reduction*, Memorie della Societá Astronomica Italiana, vol. 80, 2009, pp. 578-583.

[8] Szalay, A. S., and Gray, J. *The World-Wide Telescope*. Science 293 (Sept. 2001), 2037– 2040.

[9] A. Schaaff, F. Le Petit, P. Prugniel, E. Slezak, C. Surace, *Workflow in Astronomy: the VO France Workflow Working Group Experience*. Astronomical Data Analysis Software and Systems ASP Conference Series, 2008, pp. 77.

[10] A. Schaaff, F. Bonnarel, M. Louys, E. Slezak, B. Gassmann, C. Pestel, and O. Benjelloun, *Workflow systems and VO standards*, Memorie della Societá Astronomica Italiana, vol. 80, 2009, pp. 559-564.

[11] N.A. Walton, D.K. Witherick, T. Oinm, and K.M. Benson, *Taverna and Workflows in the Virtual Observatory*, Astronomical Data Analysis Software and Systems ASP Conference Series, 2008, pp. 309-312.

## Appendix A – Interview Questionnaire

We present here the list of questions made to a set of target astronomers. We also present an example of the practical exercise we have asked our interviewees to perform.

1 - QUESTIONS

1. Which area of astronomy do you work on?

2. Could you explain in a couple of sentences what do you do, assuming that the interviewers are general public and know little about astronomy.

3. How do you measure the success of your research? (Number of publications, number of citations, accessibility to funding, invitations to meetings, etc)

4. Do you consider teamwork essential?

5. Do you work as part of a team?

6. How do you interact with your team members? Regular meetings? Emails?

7. In general astronomers develop most of their work in their desktop/laptop. Could you tell us about how you organize your work? (Folder per item? Random notes? Others?)

8. How do search for a specific file inside your computer? / How do you access information remotely?

9. Current astronomy research makes use of new technologies like online tools and catalogues. Do you take advantage of these? If so can you give us an example?

10. If you were presented with two catalogues containing relevant data for your research, how would you assess the quality of the catalogues? Which one would you choose?

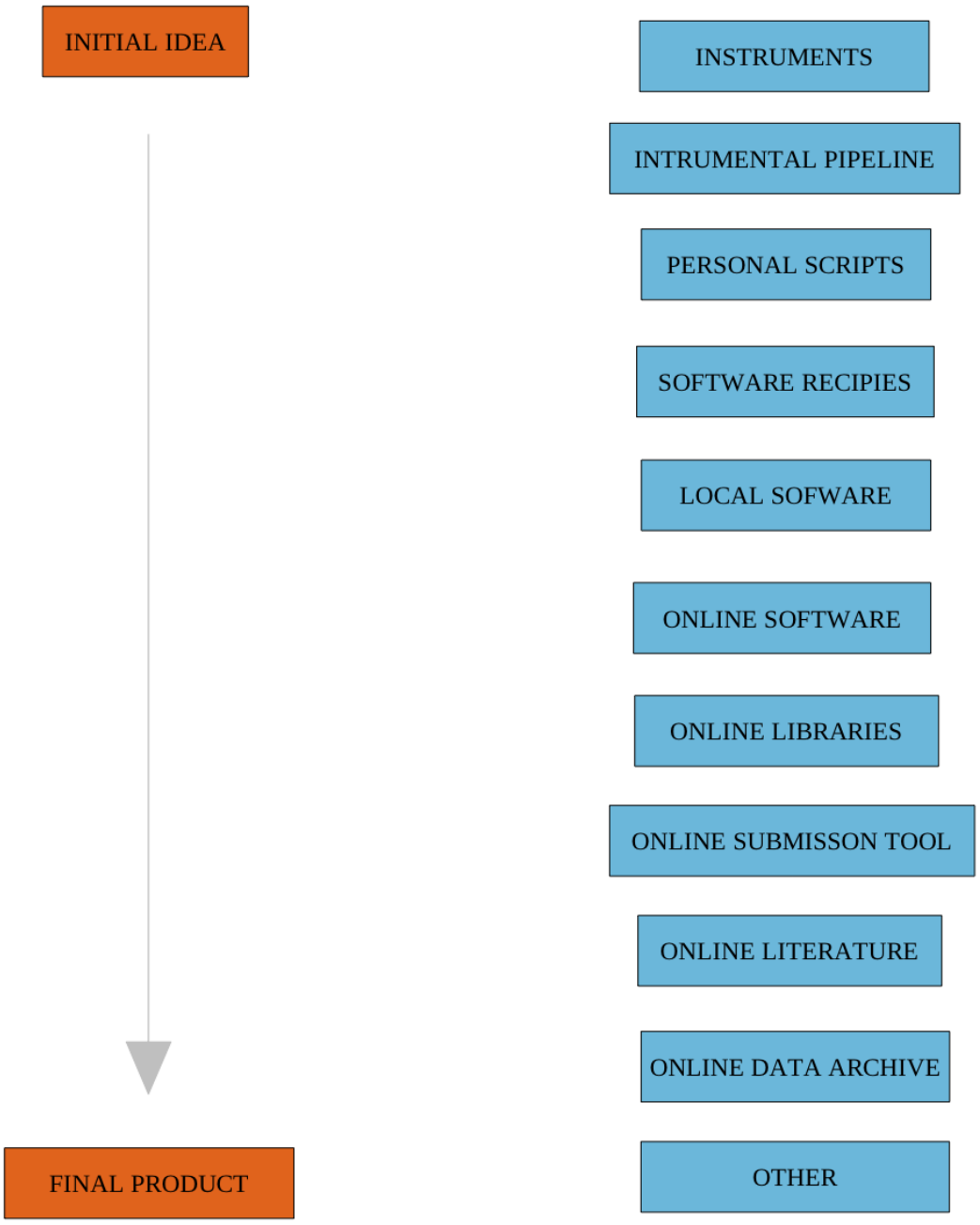
11. Same as above but for two scientific papers.

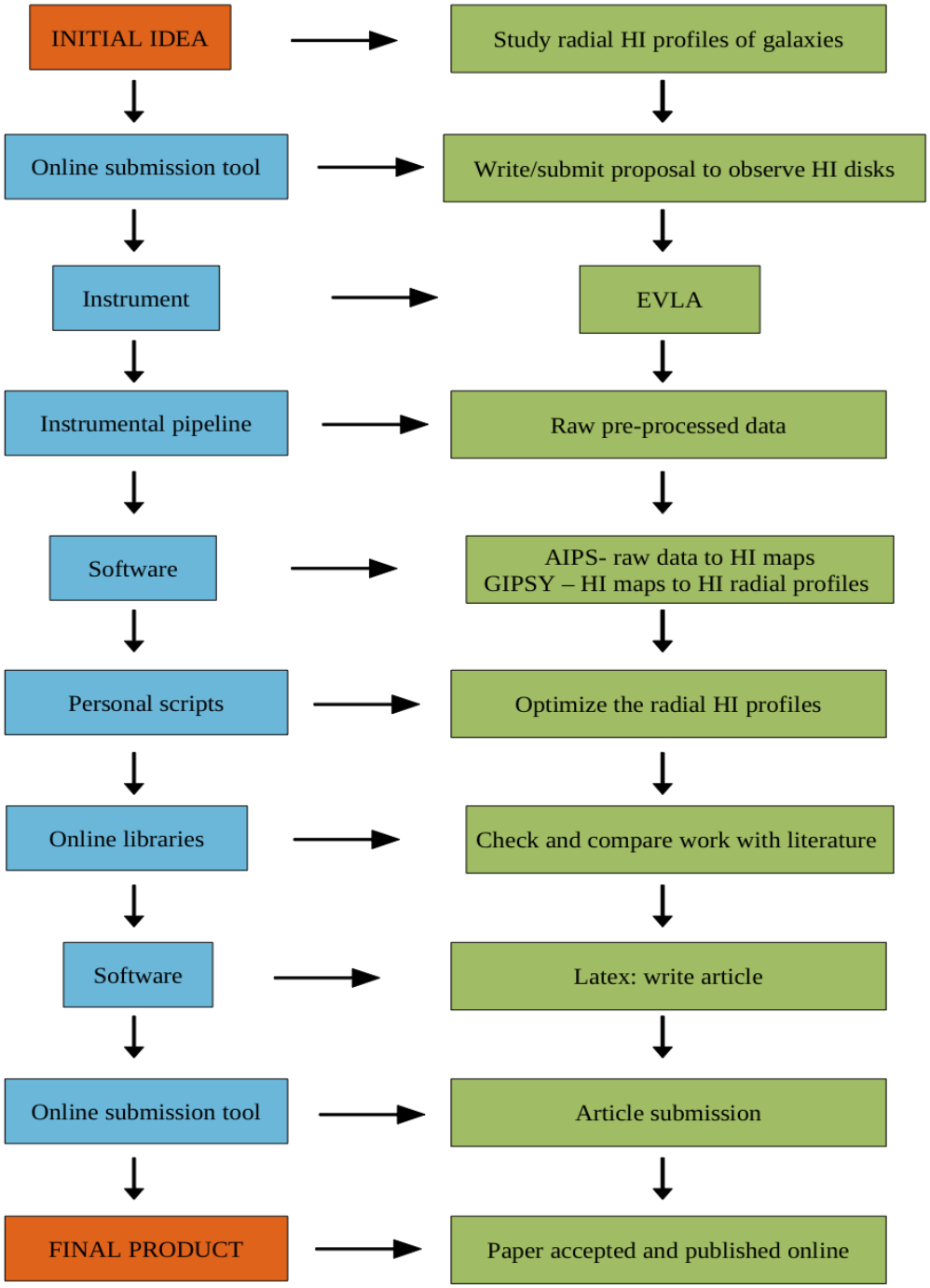
12. Did you ever need to contact a colleague because you felt that the available material (paper/talk) didn't contain the information you were looking for?

13. Which tags/ keywords would describe your work if 3$^{rd}$ parties were searching for it?

2- DIAGRAM

We ask you to sketch a diagram using the bullets points on the right column an applying it to a small portion of your latest work (a table or a plot from your own work). We provide you with a practical example for inspiration.




INITIAL IDEA
Study radial HI profiles of galaxies
Online submission tool
Write/submit proposal to observe HI disks
Instrument
EVLA
Instrumental pipeline
Raw pre-processed data
Software
AIPS- raw data to HI maps
GIPSY – HI maps to HI radial profiles
Personal scripts
Optimize the radial HI profiles
Online libraries
Check and compare work with literature
Software
Latex: write article
Online submission tool
Article submission
FINAL PRODUCT
Paper accepted and published online