Building a VO-compliant Radio Astronomical DAta Model for Single-dish radio telescopes (RADAMS)

**J.D. Santander-Vela · E. García ·
S. Leon · V. Espigares · J.E. Ruiz ·
L. Verdes-Montenegro · E. Solano**

Abstract The Virtual Observatory (VO) is becoming the de-facto standard for astronomical data publication. However, the number of radio astronomical archives is still low in general, and even lower is the number of radio astronomical data available through the VO. In order to facilitate the building of new radio astronomical archives, easing at the same time their interoperability with VO framework, we have developed a VO-compliant data model which provides interoperable data semantics for radio data. That model, which we call the Radio Astronomical DAta Model for Single-dish (RADAMS) has been built using standards of (and recommendations from) the International Virtual Observatory Alliance (IVOA). This article describes the RADAMS and its components, including archived entities and their relationships to VO metadata. We show that by using IVOA principles and concepts, the effort needed for both the development of the archives and their VO compatibility has been lowered, and the joint development of two radio astronomical archives have been possible. We plan to adapt RADAMS to be able to deal with interferometry data in the future.

J.D. Santander-Vela
European Southern Observatory, ESO, Karl-Schwarzschild-Strasse 2, 85748 Garching, Germany. E-mail: jdedsant@eso.org

J.D. Santander-Vela · E. García · V. Espigares · J.E. Ruiz · L. Verdes-Montenegro
Instituto de Astrofísica de Andalucía, IAA-CSIC, Camino Bajo de Huétor 50, 18080 Granada, Spain.

S. Leon
Joint ALMA Observatory/European Southern Observatory, JAO/ESO, Av. Alonso Córdova 3107, Vitacura, Santiago, Chile.

E. Solano
Centro de Astrobiología. Departamento de Astrofísica. P.O. 78. E-28691 Villanueva de la Cañada, Madrid, Spain.

1 Introduction

With the advent of digitisation, sharing astronomical data has become not only possible, but desirable in order to make better use of our community's valuable observing resources, and combine data from multiple instruments to enhance science, making data exploitation more efficient.

However, as more and more data are made available by observatories and other data providers in the form of digitally archived data, astronomers face new problems. They have to deal with different data access procedures, different data formats, and different data semantics (how to compare equivalent concepts originating from different instruments).

The VO (Szalay and Gray, 2001) tries to answer these concerns by specifying common data access protocols such as the Simple Cone Search (Williams et al., 2008) for spatially querying catalogues, or the Simple Image Access and Simple Spectra Access Protocols (Tody and Plante, 2009; Tody et al., 2008) for image and spectral data; data formats such as the XML-based VOTable (Ochsenbein and Williams, 2009) for tabular data interchange, linking to FITS files, and expressing observation metadata; and data semantics, by means of virtual observatory data models.

These common standards have been possible by the joint work of an international standardisation body, the IVOA (Hanisch and Quinn, 2003), a federation of national and supranational VO groups, which steers and sanctions the development of the different parts of the VO infrastructure, thanks to its different Working Groups (WGs). In particular, the Data Access Layer (DAL) and Data Modelling (DaM) WGs are the ones in charge of standardising access protocols and data models within the VO.

A data model can be defined (see, for instance, Hirschheim et al., 1995) as the description of the set of entities needed for information storage in a particular problem domain to be solved by a software system, and specifies both the data being stored, and the relationships among them.

Questions about the data must be answered by using the relationships encoded in the data model. Considered this way, a uniform, interoperable observation data model can be mapped into a uniform set of questions which can be answered within the VO.

VO data models affect how data from observatories' archives are exposed through VO services, as the way such data are stored by the archive might differ from the IVOA standards. However, there are particular attributes that need a precise description in order for the model to be useful, so VO data models need to fix the kind of metadata they support: a completely open model, where arbitrary metadata can be attached, can be very useful for annotations, but it is much more difficult to use for querying. And we must have present that, within the VO, data models apply not only to the scientific data, but to the metadata describing them.

Both within and outside the VO framework there is a long history and expertise on building optical, infrared and ultraviolet archives (e.g., MAST, see Christian et al., 1999; Kamp et al., 2005, for a VO perspective). However, that expertise is not as pervasive for radio astronomy, where only a few of the radio telescopes provide an online archive, and very few observations are accessible through the VO.

An interesting challenge, then, is building radio astronomical data archives both for single-dish and interferometry instruments, accessible through the VO, enhancing the multi-wavelength window available through the VO. But for that, we also need to provide an answer to the question on how to express the data access, data format, and data semantics of radio astronomical data within the framework of the VO, so that these datasets can be easily combined with data from other wavelengths. This article is focused on the data format and data semantics for such services, using as much as possible of the existing VO infrastructure and standards, and avoiding the reinvention of existing processes.

The work presented here was developed within the AMIGA project (Analysis of the interstellar Medium of Isolated GAlaxies[1]) at the Instituto de Astrofísica de Andalucía (IAA-CSIC) with the aim of providing an archival infrastructure for the Institut de RadioAstronometrie Millimétrique (IRAM[2]), in order to lower the barriers for radio astronomical data publishing in the VO, and to increase the number of available datasets in the radio wavelengths. This was possible because AMIGA scientific activities (e.g. Verdes-Montenegro et al., 2005) are being supported by a group of software engineers specialised in radio archives within the VO framework (see Ruíz et al., 2008, for a review of AMIGA VO activities).

2 Building blocks: available IVOA data models

Given the decision to bring VO compatibility to the radio astronomical archives to be developed from the beginning, we wanted to make as much reuse as possible of existing VO standards and best practices.

By 2005, the beginning of the development of the the TAPAS IRAM-30m (see Leon et al., 2012) and DSS-63 archives, and hence of the RADAMS, the IVOA had only one Recommendation (highest degree of endorsement by the IVOA; see Hanisch et al., 2009), the VOTable[3] (Ochsenbein and Williams, 2009). This provided a common format for tabular data interchange, but did not provide specific semantics for it.

The Data Model for Observation (ObsDM; McDowell et al., 2005) was drafted by the DaM WG to provide a framework that could be used to unify all kinds of astronomical observations from a discoverability point of view. It provides the concept of dataset characterisation, that was later developed in depth by the Characterisation data model (CharDM), which was finally promoted to the Recommendation status (IVOA Data Modelling WG, 2008).

The DaM WG has also distilled the minimum (core) metadata elements from the ObsDM (Louys et al., 2011). These ObsDM Core Components can be used to expose and discover through the VO lists of observations, by specifying most of the CharDM metadata for all kinds of astronomical observations in the spatial, temporal, spectral, and polarisation axes.

Other IVOA data models presently in the Recommendation status, and which are relevant to the description of radio astronomical datasets are the Spectral

---

[1] http://amiga.iaa.csic.es

[2] http://www.iram-institute.org/

[3] http://www.ivoa.net/Documents/VOTable/

data model (SpecDM; McDowell et al., 2007), and the Space-Time Coordinate metadata model[4] (STC; Rots, 2007).

In addition to data models, the IVOA also provides recommendations for standardised data semantics, so that equivalent information between data models can be tagged and identified as such. The Unified Content Descriptors (UCDs; Preite Martinez et al., 2007) are atomic units which can be used to identify the physical meaning of different data columns, or data attributes, with rules for combining them to provide more specific meaning, or to match them against search terms to find equivalent concepts (Derriere et al., 2004). On the other hand, UTypes where defined as an specific attribute which can be attached to a VOTable `<FIELD>` or `<PARAM>` tags, in order to make clear that the values for that field or parameter exactly correspond to a particular data model (see Ochsenbein and Williams, 2009, section 4.6). We will make use of these semantic constructs to be able to export archive data with the most rich metadata.

By directly using VO data models as the basis for astronomical archives, no intermediate layer for translation is needed between the storage and VO publication stage, and we ensure at the same time that all relevant VO metadata are present.

In case of building VO archives on top of existing, non-VO oriented archive, the translation layer needs to provide some metadata which are not just translated, but in many cases need to be extracted or even calculated from either the FITS headers, the actual FITS data tables/images, the observation logs, or possibly a combination of all of them.

When dealing with radio astronomical specific concepts not covered by any VO standard, we made use of the proposals from Lamb and Power (Lamb and Power, 2003) to the IVOA, the VO archive of the Australia Telescope Compact Array (ATCA; Murphy et al., 2006), and concepts from the Alma Science Data Model (ASDM; Viallefond, 2006).

3 Building the RADAMS

While using the VO, we can identify four different phases in the typical astronomer's workflow, with different data semantics needed during each one:

- *Discovery:* Datasets available in the VO have to be discoverable for them to appear in VO tools. In this phase we need a unified description of where in space-time, but also on frequency (possibly velocity) coordinates, spatial, temporal, and frequency resolution, the VO Registry holds data for existing datasets so that they can be easily discovered. The data model for STC (albeit in a simplified form most of the time), CharDM, Resource metadata (ResDM), and the UCD and IVOA thesaurus (IVOAT) are relevant in this phase.
- *Evaluation:* Datasets have to be evaluated in order to assess their applicability to the kind of analysis we might wish to perform; for instance, in order to do image mosaicing we need a certain coordinate overlap, and in order to do image stacking we need an almost complete overlap, and comparable resolutions. The

---

[4] As of this writing, the Simple Spectral Lines data model is in the Request For Comments (RFC) phase, and if no negative responses are presented, will be promoted the Recommendation status (Osuna et al., 2010) as soon as revised draft, taking into account the comments provided during the RFC, is submitted to the IVOA's Executive committee.

main data model involved in this phase is the CharDM, but also Curation (to
evaluate data based on authority), and Provenance (to evaluate data based on
observational and environmental restrictions).

- *Data Access:* There is an implicit data model in the IVOA data access protocols,
the Data Access Layer (DAL), which is focused on targets (coordinates with
tolerances/search radii), and uses several properties from the CharDM, such
as the Coverage in several axes, to allow for narrower searches. This DM is
always used by all systems.
- *Transformation:* When creating a new data set, or transforming an existing
one, a new CharDM instance needs to be created. If the transformed data set
is a spectrum, the Spectral data model (SpecDM) is needed both for obtaining
the complete description of the original data and describing the transformed
product[5]. In addition, in order to trace the origin of the transformed image
we would need to use interoperable Provenance information, suggested by the
ObsDM.

We can see, then, that in order to properly enable VO operations for dataset
Evaluation and Transformation we need to use Provenance metadata. Provenance-
related dataset metadata are part of some IVOA data models (i.e., the SpecDM
DataID), but RADAMS aims to provide a description of Provenance applicable to
all kinds of radio observations. Besides, Curation metadata, part of the Discovery
process, need to be rich enough to be linked with Policy information. Plus, if we
want to deliver data beyond FITS files, and allow characterisation information to
be available in native format offline, Packaging is needed. By adding extra, radio
astronomical specific metadata, we also enhance the specification of Sensitivity in
the CharDM.

This shows that the RADAMS is a radio astronomical observation data model
whose main emphasis is describing single dish radio observations (hence, Radio
Astronomical DAta Model for Single-dish), by using the ObsDM for VO stan-
dardised metadata. RADAMS implements for the first time for radio astronomical
observations, irrespective of the data product being generated, parts of the Ob-
sDM such as Curation, Policy, Provenance, and Packaging, and defining how to
calculate the radio astronomy specific parts of the CharDM.

Figure 1 shows the main composition of the RADAMS, with the origin of the
different set of classes, and shows greyed out the classes which are defined by the
RADAMS for the first time.

RADAMS' defining classes and sub-models are:

- *Observation:* Abstract root class for the data model. Works as a hub to which
both actual observational data (ObsData) and metadata are linked together.
- *ObsData:* Represents the actual data being described by all the RADAMS'
classes.
- *Target:* Describes the target of the observation, providing as much information
as available for already known targets, and only pointing information for not
known targets.
- *Characterisation:* Corresponds to the CharDM classes, describing where to lo-
cate the observation in scientific parameter space (spatial bounds, temporal

---

[5] Incidentally, there is no existing data model yet for images or for more complex data within
the VO, beyond what is available from FITS conventions.
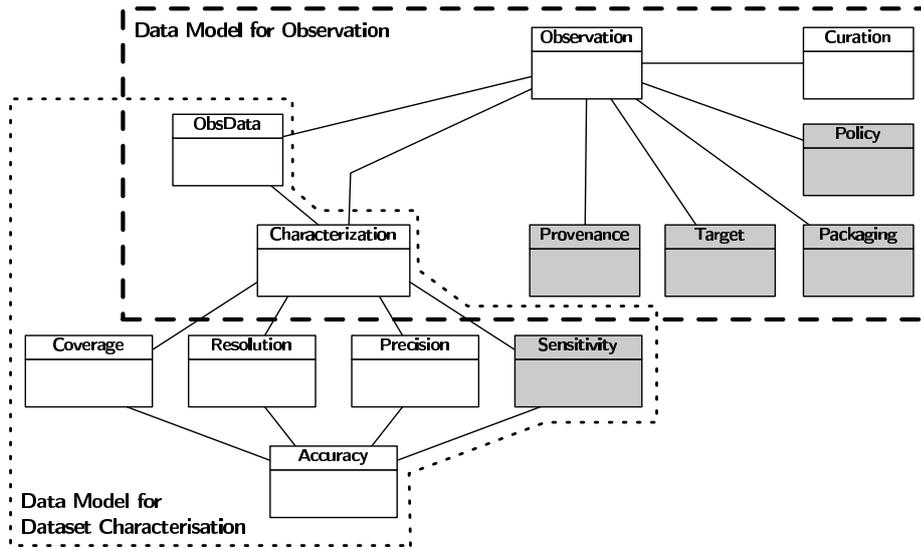
Fig. 1 The complete RADAMS data model, showing the inspiration of their classes, with classes newly specified by the RADAMS in grey.

bounds, spectral bounds, and even observed flux bounds or polarisation), and resolution along those axes.

In practice, a different class exists for each different kind of axis, as specialising per axis allows for different axes having different metadata attached.

– *Provenance:* Binds all the information regarding how the observation was originated, both from technical and scientific point of view.
– *Curation:* Describes who is responsible for maintaining a particular observation, a set of observations, or all of the archive. It also contains the information identifying who originated the observation (i.e., the PI of an observation program).
– *Policy:* Is used to specify the access rights for different systems and persons accessing the archive. Information from the Curation class is used in order to assess the accessibility of datasets.
– *Packaging:* Describes the way an observation, or a set of observations, are actually delivered when the archive is queried. Enables component reuse, recursive embedding, and sub-package selection.

RADAMS attributes are assigned their semantic meaning using a flat table that relates each attribute to its definition. UCDs and UTypes are used to relate RADAMS attributes to common semantic definitions in existing IVOA data models. These mappings collectively constitute the RADAMS data model.

All of the metadata classes listed before are implemented by the RADAMS. In the following sections we will describe in detail only the main contributions of the RADAMS, namely Policy, Packaging, and specially observation Provenance. The description includes tables describing the specific metadata, semantics, and UCDs assigned to each metadata entity.

The medatata for each class will be shown in Tables 1 to 17, and referred to appropriately in the following sections. For each table, the RADAMS attribute name is listed together with its description, UCD, and where available the correspond-

ing FITS keyword either from the FITS standard (FITS Working Group, 2008), or IRAM Multi-Beam FITS (Muders et al., 2005) or NRAO extensions (Prestage and Clark, 2004). When a clear candidate for a FITS keyword was not found, the table will show `N/D` (Not Defined). For information encoded as FITS comments, the `COMMENT` keyword will be used.

4 Specifying access policy

Policy is defined as *a course or principle of action adopted or proposed by a government, party, business, or individual* (OUP, 2005). For the VO, policies we are concerned with are *data access policies*, and they can be defined as the granting or denial of access to data (or metadata) following some principles proposed by data providers[6]. Those principles take into account the agents (people, groups of people,or systems) responsible for data generation or data curation, and the relationship between these agents and the user trying to access the data.

Moreover, those principles (policies) change from institution to institution, and also for different datasets curated by those institutions. Hence, the way to specify those different policies must be general enough, and must take into account the different users' relationships with the datasets, and the different ways to implement policies, from *everything is accessible*, to *PI's eyes only*, and going through *metadata for everyone, data only for PI and CoIs*, among other possibilities[7].

The Policy data model must be able to allow for very complex policies to be applied to the data. In the case of a VO archive where the data are not instantly available as part of an operational workflow, or for data which are intended to be public from the moment of insertion, Policy becomes simpler[8], as in *everything in the archive is available for everyone as soon as it is in the system.*

The proposed Policy data model implements a Role-Based Access Control (RBAC) system (see Ferraiolo and Kuhn, 1992; Ferraiolo et al., 1995), in order to simplify the administration of access permissions. RBAC greatly simplifies the management of access, and at the same time provides greater flexibility. Users can be assigned to roles following the logic of their relationship with the data to be accessed (usually corresponding to the organisational relationships, i.e., whether users are PI, CoIs, observers, operators, delegates, etc. with respect to the data). When needed, roles can be granted new permissions (accessing new kinds of metadata, for instance), or permissions revoked, in a way which is independent of the logic rules assigning roles to the users.

One of the most successful RBAC-based data systems is the Integrated Rule-Oriented Data System (iRODS; Weise et al., 2008), which uses rules not just for data access, but also for data storage and deployment.

Role-based policies are more flexible within the astronomy domain than user or group related policies: as the actual file or metadata access or modification permissions will be the same for users with the same role, systems administrators

---

[6] Other policies of interest for an astronomical archive, but outside of the scope of the VO, might involve data erasure, replication, or billing.

[7] An example of very user oriented policy can be found at ESO, where certain datasets are only accessible by users who are registered as belonging to an ESO member state.

[8] In the case of the Robledo Archive, the policy implemented was the standard NRAO policy: 18 months since the end of the observations.
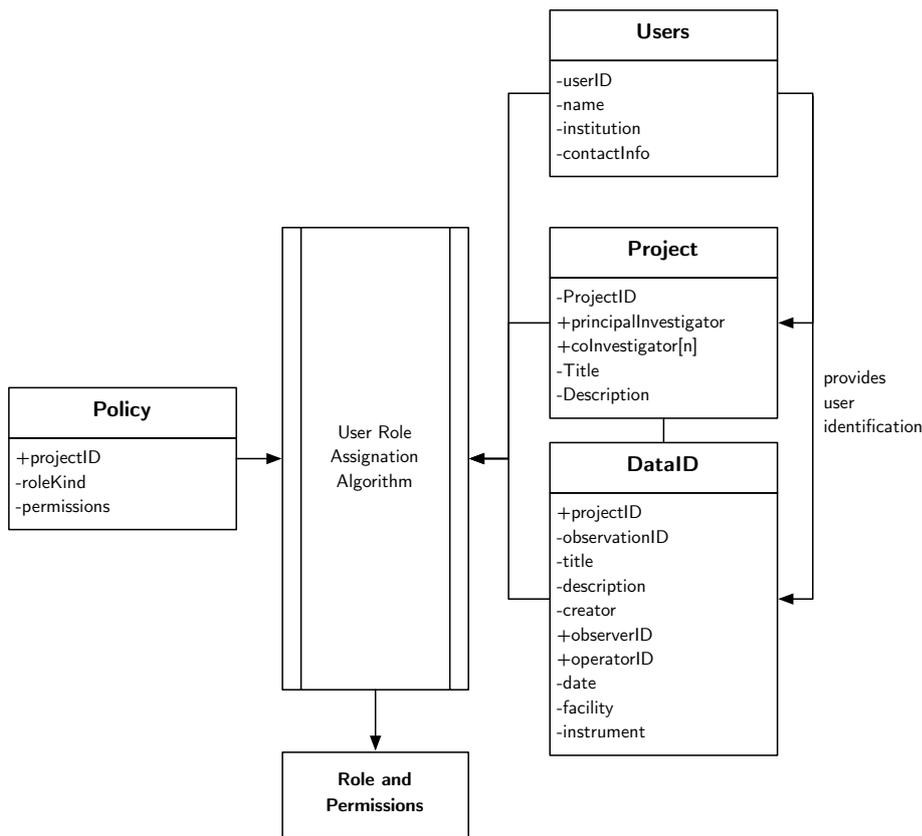
Fig. 2 Policy data model, with role and permissions.

have less administrative burden. Roles only need to be assigned (or calculated) for users; the corresponding permissions are already predefined for each role.

Figure 2 shows the different classes needed to characterise the archive policy in a more generic setup.

In RADAMS, roles are chosen from a controlled vocabulary (`principalInvestigator`, `observer`, `coInvestigator`, `observatoryStaff`, `none`); roles are dynamically derived from user/project relationships, so that people not belonging to the observatory, and who have nothing to do with the project, would default to `none`. The `none` role ensures that there is always at least one role to be matched, with the most restrictive access permissions to be implemented.

New roles can be created, with their own permissions set, but would only be assigned to a user if logic for the user-data-role matching is provided.

Each project in the archive should have, at least, an explicit policy of what is allowed for someone with no relationship with the Principal Investigator, and people with `principalInvestigator` roles have all access rights to the archive; people with roles other than `principalInvestigator` would fallback to the `none` role, if the permissions for their role are not explicitly declared.

Table 1 Policy metadata.

Attribute	FITS Keyword	UCD	Description
projectID	PROJID	`meta.curation.project;` `meta.id`[a]	Project identifier.
roleKind	N/D	`meta.policy.role;` `meta.code; meta.id`[b]	Role kind for which permissions will be provided.
permissions	N/D	`meta.policy.permissions;` `meta.code`[b]	Permissions provided for roleKind users of this project. The particular permissions to be provided are to be discussed by IVOA.

[a] There is no `meta.curation.project` UCD, but we propose the inclusion of one.

[b] There are no `meta.policy.*` UCDs, but we propose, at least, the inclusion of the `meta.policy` atom.

Table 2 Policy related Users metadata.

Attribute	FITS Keyword	UCD	Description
userID	COMMENT	`meta.id`	User identifier for all user related operations in the archive.
name	COMMENT	`meta.name`	Real name of the user. Any known user of the archive has to be registered, or be anonymous.
institution	COMMENT	`meta.name`	Name of the institution to which the user belongs.
contactInfo	COMMENT	`meta.note`	Contact info (probably e-mail) for this user.

Therefore, we use Policy, Users and ObsData metadata in order to select the corresponding role for the agent (a user, or a software agent on their behalf) just logged in. Figure 3 shows the flow diagram for the role selection.

Although specialised RBAC description and enforcement systems such as the eXtensible Access Control Markup Language (XACML; Moses, 2005), or the aforementioned iRODS exist, and could be used, the RADAMS implements a much simpler mechanism, as the number of foreseeable roles is limited.

Policy metadata and attributes are specified in Table 1. We will also need at least a subset of Curation attributes for successful Policy attribution. These are indicated in Tables 2, 3, and 4.

5 Provenance

Provenance metadata provides support for the description of how data have been generated, for the purposes of attribution, data evaluation, and transformation, and in some cases even for data access. Provenance metadata within the RADAMS is further classified as:

– *Instrumental:* Has to do with the instrumental setup, i.e., the configuration of all observation elements between the original photon source and the photon detection equipment.

Table 3 Policy related Project metadata.

Attribute	FITS Keyword	UCD	Description
projectID	PROJID	`meta.curation.project;` `meta.id`[a]	Project identifier.
principalInvestigator	COMMENT	`meta.id`	User identifier for the principalInvestigator of the project.
coInvestigator[n]	COMMENT	`meta.id`	User identifier for the nth co-investigator.
title	COMMENT	`meta.curation.project;` `meta.title`[a]	Project title.
description	COMMENT	`meta.curation.project;` `meta.note`[a]	Project description.

[a] There is no `meta.curation.project` UCD, but we propose the inclusion of one.

Table 4 Policy related DataID metadata.

Attribute	FITS Keyword	UCD	Description
projectID	PROJID	`meta.curation.project;` `meta.id`[a]	Project identifier.
observationID	OBSID	`obs; meta.dataset;` `meta.id`	User identifier for the principalInvestigator of the project.
coInvestigator[n]	COMMENT	`obs; meta.id`	User identifier for the nth project co-investigator.
title	COMMENT	`meta.curation.project;` `meta.title`[a]	Project title.
description	COMMENT	`meta.curation.project;` `meta.note`[a]	Project description.
creatorID	AUTHOR	`meta.curation; meta.id`	User identifier for the creator of the data entry.
observerID	OBSERVER	`obs.observer; meta.id`	User identifier for the person performing the observation generating these data.
operatorID	OBSERVER	`obs.operator; meta.id`[b]	User identifier for the person performing operator duties while performing the observation.
date	DATE-OBS	`time.obs.start`	Date of observation.
facility	TELESCOP	`instr.obsty`	Facility (observatory) where the telescope/instrument resides in.
instrument	INSTRUME	`instr.tel`	Instrument performing the observation[c].

[a] There is no `meta.curation.project` UCD, but we propose the inclusion of one.

[b] We propose the inclusion of either `obs.operator` or `instr.operator` as new UCDs to characterise operator-related data. However, `obs.observer` can be used when providing both observer and operator at the same time.

[c] In principle, this metadata item should contain an instrument-backend pair, but it might be useful for some future facilities to separate attributes for both.

– *Environmental:* Has to do with the elements in the path of the photon source which cannot be controlled by the instrumental setup, but that nonetheless affect the photon collection (i.e., by causing turbulence, changing the refraction index, or absorbing photons). We will register measurable environmental parameters in order to estimate possible effects and/or defects in the detections.

– *Processing:* Once raw data are recorded, many different processing steps need to be performed in order to provide science-ready data, or as it is sometimes said, data are provided with the instrument signature removed as much as possible. Processing Provenance records the different processes and their inputs performed in order to achieve the result being offered by the archive.

As stated previously, the medata related to the observation proposal, being arguably part of its provenance, will be kept as part of the Curation class.

5.1 Instrumental provenance

Instrumental provenance in the RADAMS was initially inspired by the IVOA note by Lamb and Power (2003). The aim was to provide a framework that could be adapted to single-dish telescopes, but also to radio interferometers, and could be used to specify all instrument configuration data using the beam as the unit by which we build up feeds, antenna, and global telescope configuration.

Figure 4 shows the classes associated with the instrumental configuration for the observation.

– *InstrumentConf:* Each observation is associated to a particular *instrumental configuration*, which is the result of the particular setup of the instrument + antenna + feed system. InstrumentConf instances group those settings.

– *Instrument:* Instances of this class specify the instrument-specific part of the configuration, as part of the observation setup.

– *Instrument.Location:* This is an instance of a Location class, used for specifying the location for the instrument[9].

– *AntennaConf:* In the same way InstrumentConf allows the grouping for all the instrumental settings, AntennaConf instances group together Antenna and Feed metadata (regarding polarisation), plus BeamConf —another aggregator class—. Several AntennaConf instances can provide information for different antenna + feed combinations. Each possible antenna configuration will be labelled by a name. The relationship between AntennaConf and Feed instances allows the specification of different beams, formed by the combination of different feeds. We can use this association for a single-dish, multiple-feed configuration where each feed went to a different receiver. For single-dish, single-feed configurations, there is only one beam associated to a given receiver.

– *Antenna:* Instances of this class specify the general properties for each given antenna. We will also use these instances to specify the type of scan being performed on the source from a controlled vocabulary.

– *Feed:* Instances of this class —one or more per AntennaConf— specify each of the feed horns used for the observation, and their corresponding polarisation

---

[9] In a generalisation of the RADAMS to radio interferometry, aditional locations would have to be applied to each individual Antenna, or an array of locations applied to the AntennaConf.

Table 5 Provenance instrument metadata.

Attribute	FITS Keyword	UCD	Description
name	INSTRUME	meta.id; instr; meta.main	Instrument name.
description	N/D	meta.note; meta.main	Instrument description.
shortName	N/D	instr; meta.id	Short name for the instrument.
locationName	N/D	instr; pos; meta.id	Instrument site identification.
URL	COMMENT	instr; meta.ref.url	URL for the instrument (website for the instrument, documentation, or any other type of instrument description).

Table 6 Instrument location metadata.

Attribute	FITS Keyword	UCD	Description
locationName	N/D	instr; pos; meta.id	Name of a particular instrument location.
latitude	SITELAT	instr; pos.earth.lat	Instrument location latitude.
longitude	SITELONG	instr; pos.earth.lon	Instrument location longitude.
altitude	SITEELEV	instr; pos.earth.altitude	Instrument location altitude.

from a controlled vocabulary: L, R, X, Y. We cannot make use of the Stokes polarisation parameters, as they cannot be directly measured via the feed configuration: instead, they have to derived by means of data processing steps.

– *BeamConf:* This class is used to group multiple beam+receiver pairs for a given feed.
– *Beam:* Metadata for this class specify the actual beam for the telescope, associated to a given spectral band.
– *Receiver:* These metadata are used to describe the most relevant properties of the receiver, such as receiver type, intermediate frequency —in the case of heterodyne stages—, et cetera.
– *Spectrum:* In case of spectroscopic observations, the spectral analyser that has been used is specified by instances of this class.
– *Velocities:* This class mirrors the Spectrum class, and is preferred for those cases where velocities are used, instead of frequency.

The detailed Instrumental provenance metadata recorded within the RADAMS is compiled in Tables 5, 6, 7, 9, 10, and 11. In addition, the description of spectral features (also to be reused by the CharDM in the Spectral axis) is provided in Table 12, with a mirror for velocity-based metadata in Table 13.

The definition of each element of the scanType vocabulary (Table 7) can be found in Table 8.

The vocabulary in use for Table 12 reflects the current status of radio astronomical observations, where only a line is selected as the target for the observation. However, broader spectrometers, such as those available with EMIR, or in the future with ALMA, might ask for a list of available lines, instead of specifying a single one.

Table 7 Antenna configuration metadata.

Attribute	FITS Keyword	UCD	Description
name	N/D	`instr.telescope; meta.title; meta.id`	Name of the particular antenna.
scanType	N/D	`instr.setup; meta.code`	Type of scan being performed by this antenna, from a limited vocabulary (Lamb and Power, 2003): `beamSwitch`, `cal`, `cross`, `dopplerTrack`, `dwell`, `focus`, `frequencySwitch`, `holography`, `mosaic`, `onOff`, `onTheFly`, `point`, `positionSwitch`, `pulsar`, `raster`, `skyDip`, `tiedArray`, `track`, `wobblerSwitch`.
mount	N/D	`meta.note`	Mount type for the telescope from a limited vocabulary: `azimuthal`, `equatorial`, `altazimuthal`, `dobson`, `german equatorial`.
majorAxis	N/D	`instr; phys.size.smajAxis`	Major axis dimensions.
minorAxis	N/D	`inst; phys.size.sminAxis`	Minor axis dimensions.
effectiveArea	N/D	`instr; phys.area`	Effective instrument area.

Table 8 Meaning of the different valid values for the `scanType` attribute in Table 7.

scanType	description
`beamSwitch`	Scan where the change of reference position is accomplished by means of changes in the beam optical path.
`cal`	Power or gain calibration scan.
`cross`	Pointing scan across azimuth and elevation.
`dopplerTrack`	Tracking with doppler correction for the channels being automatically applied. Compare with`track`.
`dwell`	A tracking scan for which the duration of the actual observation time is prescribed, apart from any time needed to reach the source.
`focus`	Scan used to focus the system.
`frequencySwitch`	On-off scan performed by means of frequency switching.
`holography`	Surface precision scanning by means of raster-scanning the beam through a well-known point source.
`mosaic`	Interferometry-specific scan type, for creating data cube mosaics.
`onOff`	On-off scan, without specifying the switching technique.
`onTheFly`	Mapping scan performed while the antenna follows a pre-defined path, with some periodical movement to a reference (off) position.
`point`	Power or spectral observation on a fixed antenna position.
`positionSwitch`	On-off scan performed by means of changing the antenna reference position.
`pulsar`	Pulsar-tuned tracking scan.
`raster`	Power or spectral observations along a raster line.
`skyDip`	Sky-dip, or tipping of the antenna to get a T_A or T_{sys} profile over different elevations.
`tiedArray`	Interferometry-specific scan type, where the complete array produces phased data to VLBI observations.
`track`	Power or spectral observation while the antenna tracks a particular celestial point, or follows a solar system body.
`wobblerSwitch`	Similar to `beamSwitch`, but using a *wobbler* as secondary mirror.

Table 9 Feed configuration metadata.

Attribute	FITS Keyword	UCD	Description
polarisation	POLTY	`phys.polarization; meta.code`	Polarisation value from a controlled vocabulary: L, R, X, Y

Table 10 Beam configuration metadata.

Attribute	FITS Keyword	UCD	Description
beamMajor	BMAJ/HPBW	`instr.beam; phys.size.smajAxis`	Major axis HPBW of the main lobe of the beam.
beamMinor	BMIN/HPBW	`instr.beam; phys.size.sminAxis`	Minor axis HPBW of the main lobe of the beam.
sensitivity	BEAMEFF	`instr.beam; instr.sensitivity`	Beam average sensitivity.
mainBeamSolidAngle	N/D	`instr.beam; pos.posAng; meta.main`	Main lobes beam solid angle.
totalBeamSolidAngle	N/D	`instr.beam; pos.posAng; stat.max`	Total beam solid angle, including secondary lobes.
directivity	N/D	`instr.beam; instr.setup; arith.factor`	Directivity percentage.
gain	ANTGAIN	`instr.beam; instr.setup; arith.factor`	Beam gain[a].

_____

[a] We still have to clarify if the gain attribute is related to the directivity concept or not, and if it is related with the receiving stages or not.

Table 11 Receiver metadata.

Attribute	FITS Keyword	UCD	Description
type	BACKEND	`instr.setup; meta.note`	Receiver type (HEMT, Bolometer, SIS, et cetera).
skyCentreFreq	N/D	`src; em.radio; em.freq`	Antenna tuning frequency.
ifCentreFreq	N/D	`instr.setup; em.freq`	Heterodyne receiver intermediate frequency (or list of frequencies).
bandwidth	BANDWID	`instr.bandwidth`	Filter-bank total bandwidth.

5.2 Environmental provenance

Environmental provenance is the part of the provenance dealing with ambient conditions, and as such the main class is called Provenance.AmbientConditions: it encompasses all metadata needed to specify weather conditions, air mass, opacity, et cetera. Figure 5 shows the corresponding classes and their relationships.

Table 12 Spectrum metadata. It might be necessary to change the `MOLECULE` and `TRANSITI` keywords to `LINE`, for better CLASS compatibility.

Attribute	FITS Keyword	UCD	Description
numChannels	NAXISn	`spect; em.freq; meta.number`	Number of spectral channels.
chanSeparation[a]	N/D	`spect; em.freq`	Mean channel separation (in frequency units), or channel frequency separation array.
freqResolution	FREQRES	`spect.resolution; em.freq`	Frequency resolution.
velRefFrame	SPECSYS	`spect; phys.veloc; pos.frame; meta.id`	Identification of the reference system used for the velocity.
refChanNum	N/D	`spect; em.freq; meta.number; meta.ref`	Spectral reference channel.
refChanFreq	OBSFREQ	`spect; em.freq; meta.code; meta.ref`	Spectral reference frequency (observed frequency).
restFreq	FREQn or RESTFREQ	`spect.line; em.freq`	Observed spectral line rest frequency.
molecule	MOLECULE[b]	`spect; phys.mol; meta.id`	Molecule name.
transition	TRANSITI[c]	`spect; phys.atmol.transition; meta.id`	Transition.

[a] There is a certain redundancy between the Provenance.Spectrum.chanSeparation attribute and the Coverage.Spectral.Resolution attributes.

[b] It might be necessary to change the `MOLECULE` keyword by `LINE`, for better CLASS compatibility.

[c] It might be necessary to change the `TRANSITI` keyword to `LINE`, for better CLASS compatibility.

– *AmbientConditions:* Holds all metadata related with weather conditions for the observation, such as humidity, wind speed, opacity at zenith, et cetera.
– *OpacityCurve:* Includes the opacity curve (linked as a VOTable file) associated to the observing term where the data were observed (which we will derive from the nearest two skydip scans performed before and after the observation). We propose the inclusion of an array of `[elevation, Tsky]` pairs, together with the azimuth and the starting time of the skydip. Calculation of the opacity curve is different for bolometric or heterodyne observations. It is also necessary to include information on the atmospheric model and/or software used for opacity fitting. For instance, the atmospheric model used by MOPSIC and MIRA, the data reduction packages at the IRAM 30m antenna, is the *Atmospheric Transmission at Microwaves* (ATM; Pardo et al., 2001).

By providing a unified Ambient Provenance model, we allow for querying global ambient variables across the VO, which can allow for better selection of observations, but also for establishing a multidimensional quality metric that takes into account weather, and also CharDM resolution, and that can even be used for weather science and forecast, including applications such as weather-prediction assisted scheduling (Alvarez et al., 2010), but using data across observatories.

Table 13 Velocity metadata.

Attribute	FITS Keyword	UCD	Description
numChannels	N/D	`spect; phys.veloc; meta.number`	Number of velocity channels.
chanSeparation[a]	N/D	`spect; phys.veloc`	Mean channel separation (in velocity units), or velocity channel separation array.
velResolution	N/D	`spect.resolution; phys.veloc`	Velocity resolution.
velRefFrame	N/D	`spect; phys.veloc; pos.frame; meta.id`	Identification of the reference system used for the velocity.
refChanNum	N/D	`spect; phys.veloc; meta.number; meta.ref`	Velocity reference channel.
refChanVel	N/D	`spect; phys.veloc; meta.code; meta.ref`	Frequency for the velocity reference channel.
restFreq	`RESTFREQ`	`spect.line; em.freq`	Observed spectral line rest frequency.
molecule	`MOLECULE`[b]	`spect; phys.mol; meta.id`	Molecule name.
transition	`TRANSITI`[c]	`spect; phys.atmol.transition; meta.id`	Transition.

[a] There is a certain redundancy between the Provenance.Velocity.chanSeparation attribute and the Coverage.Spectral.Resolution attributes.

[b] It might be necessary to change the `MOLECULE` keyword to `LINE`, for better CLASS compatibility.

[c] It might be necessary to change the `TRANSITI` keyword to `LINE`, for better CLASS compatibility.

Table 14 AmbientConditions metadata.

Attribute	FITS Keyword	UCD	itemize
timeStamp	`DATE-OBS`	`time.obs.start`	Moment at which the weather snapshot is taken.
opacity	`TAUZEN`	`phys.absorption.coeff`	Opacity at zenith estimated at the observation frequency.
airMass	`AIRMASS`	`obs.airMass`	Air mass at zenith at the observing site.
temperature	`TAMBIENT`	`phys.temperature`	Ambient temperature.
humidity	`HUMIDITY`	`obs.atmos; phys.columnDensity`	Ambient humidity.
waterVapour	`TAU_WPATH_RD<freq>`[a]	`obs.atmos; phys.pressure`	Equivalent pressure of the water vapour column.
tauFrequency	`TAU_WPATH_RD<freq>`[a]	`obs.atmos; em.freq`	Tau radiometer frequency.
wind	`WINDSPEE`	`obs.atmos; phys.veloc`	Wind speed.

[a] This keyword is defined within the IRAM MB-FITS bidimensional table. The first column corresponds to `waterVapour`, the second to `tauFrequency`.

Table 15 Opacity metadata.

Attribute	FITS Keyword	UCD	itemize
opacity	TAUZEN	phys.absorption.coeff	Opacity at zenith estimated at the observation frequency.
skydipStart	DATE-OBS	time.obs.start	Skydip starting time.
azimuth	AZIMUTH	pos.az	Skydip azimuth.
elevation[n]	ELEVATIO	pos.el	Skydip scan elevation.
tsky[n]	N/D	instr.skyTemp	Sky temp at n^{th} skydip.
atmosModel	N/D	meta.modelled; obs.atmos; meta.id	Atmospheric model identification.

RADAMS' ambient metadata are listed in Tables 14 and 15.

5.3 Processing provenance

The Provenance.Processing class enables the specification of processing processes applied to the data before archival, including some processes necessary for the actual observation, such as the determination of the background signal via `frequencySwitching` or `positionSwitching` for background/source data comparison.

The RADAMS will make use of just two classes, Processing and Calibration —this is a subclass of processing—. Order is relevant, and it should be possible to reconstruct the pipeline by the ordering of Processing and/or Calibration instances. Figure 6 shows these classes.

- *Processing:* It holds information specifying the type of processing applied to data before archival. This includes pseudo-observational techniques such as position switching or frequency switching, as well as the type of data averaging, data weighting, et cetera. Table 16 provides minimal initial metadata, using arrays of parameter keywords for extensibility at the expense of complexity.
- *Calibration:* Is a subclass of Processing, where the type of processing is `data-Calibration`. This class provides additional attributes to specify the type of calibration, and the axes to where this calibration will be applied. Table 17 provides minimal initial metadata, using arrays of parameter keywords for extensibility at the expense of complexity.

We still have to develop a calibration and/or pointing model; maybe based upon IRAM-Multi-Beam-FITS, or GBT FITS calibration tables.

6 Packaging of observations

Single datasets in the VO are typically delivered in one of three forms[10]: VOTable tabular data, FITS tabular data, or FITS observation data (i.e., images, spectra, data cubes) with a VOTable description.

---

[10] Other formats, such as JPEG previews (possibly including Astronomical Visualization Metadata, http://www.ivoa.net/Documents/latest/AOIMetadata.html), or CSV values, are technically possible.

Table 16 ProcessingStep metadata.

Attribute	FITS Keyword	UCD	Description
timestamp	DATE-RED	`obs.param; time.epoch`	Timestamp for the processing step being performed.
kind	N/D	`obs.param; meta.code`	Type of processing applied to source data; comes from a controlled vocabulary: `unprocessed`, `noiseWeightedAverage`, `nonWeightedAverage`.
softwarePackage	N/D	`meta.software; meta.id`	Software package used for data processing; should come from a controlled vocabulary: `CLASS`, `AIPS`, `AIPS++`, `CASA`, `MOPSIC`, `GILDAS`, `MIRA`, `MIR`, `other`. In the case of `other`, the actual package that was used should be added as a parameter, with parameter.name as `softwarePackage` and the parameter.value as the package name.
parameter[n].name	N/D	`obs.param; meta.code`	Additional processing parameter name, whose value will be in parameter.value; eventually, we will have a controlled list of possible parameter.name values.
parameter[n].kind	N/D	`obs.param; meta.code`	From a controlled vocabulary: `integer`, `float`, `string`, et cetera. At least all of FITS data types should be present.
parameter[n].value	N/D	`obs.param`[a]	Value for the parameter indicated by parameter.name.

[a] The final UCD to mark parameter[n].value will be calculated when writing the VOTable, as it depends on parameter.kind; it will be `obs.param; meta.number` most of the time, but it could be `obs.param; meta.name` or `obs.param; meta.code`, depending on the context.

Table 17 Calibration metadata[a].

Attribute	FITS Keyword	UCD	Description
timestamp	DATE-RED	`obs.param; time.epoch`	Timestamp for the calibration step being performed.
parameter.name	N/D	`obs.calib; obs.param; meta.id`	Keyword defining the parameter that we will characterise with the remaining attributes.
parameter.kind	N/D	`obs.calib; obs.param; meta.code`	Type of calibration parameter used, from a controlled vocabulary: `additive`, `factor`, `polynomial`, `exponential`, `logarithmic`.
parameter.value	N/D	`obs.calib; obs.param; meta.number`	Value for the main calibration parameter, where parameter.kind is not `polynomial`.
parameter.sigma	N/D	`obs.calib; obs.param; meta.number`	Value of sigma, for `exponential` calibrations.
parameter.calCoeff.[n]	N/D	`obs.calib; obs.param; meta.number`	n^{th} degree coefficient for a `polynomial` calibration parameter; polynomial degree is derived from the maximum n.

[a] It is mandatory that at least one [`parameter.name, parameter.kind, parameter.value`] triplet appears, with `fluxScale` as parameter.name, and one of `antennaTemperature`, `mbBrightnessTemperature`, or `S_nu` as the parameter.value, with a parameter.kind of `string`.

However, many datasets need the delivery of more than one data product: for instance, OTF with heterodyne instruments can produce many different spectra in individual scans, and the final reprocessed data cube, and users might want to access all of them at the same time. Having an organisational unit that binds together those different data types, and keeps information on the Characterisation of them, their provenance, etc., enabling offline evaluation and selection.

The Packaging class is used to specify how data from different sources will be presented together. For instance, if we wanted to retrieve data belonging to a particular survey subset, a VO system could reply with a Multi-Beam FITS file containing all the scans and sub-scans that conform the On-Off patterns for a single issued observation, or with a `.zip` file with all the FITS files belonging to the survey, or with just a single On-Off pair, et cetera.

An instance of the Packaging class describes the contents of the data retrieved, in terms of project organisation, and of the particular files being actually delivered.

As there is no Packaging class defined at the VO level, we have resorted to other packaging description mechanisms available in other archiving tools.

We believe a standard VO packaging scheme is needed in order to facilitate distribution of datasets. The packaging scheme that we propose, called VOPack —Virtual Observatory Package; `.vopack` file extension— is a tarred and gzipped folder (`.tgz` file) with an XML content descriptor, which acts as VO-aware manifest of contained assets.

In particular, a VOPack describes the relationships between the grouped units (VOTable, FITS, and/or ancillary files, or included VOPacks), and also provides CharDM metadata for the whole VOPack, which allows to picture the package in observational parameter space.

The VOPack, then, is a way of distributing VO-compliant content, in a way that makes it easy to reuse and point to existing content, either remote or locally.

VOPacks consist of a compressed file that contains at least a `voPack.xml` file —following the VOPack XML Schema— that describes all the additional content of the compressed VOPack, and their relationships between them. Figure 7 shows the structure of a VOPack from the VOPack schema.

In that diagram, the `voPack` element is the root for the XML document. It includes a description, the originating query, and one or more `packUnits`, which actually point to the information being retrieved. The `originatingQuery` element contains the string with the URI that allows the retrieval of the voPack. Additional `characterisation` elements, following the Characterisation schema, can be used to further specify properties on the data being delivered with the VOPack.

The `packUnit` corresponds to a single piece of data, or to another `packUnits`, in case of more structured data. The depth of inclusion is arbitrary.

`packUnits` have a `type` attribute that can be one of: `votable`, `fits`, `vopack`, `compressedFolder`, `folder`, `otherXML`, `otherNonXML`. Table 18 specifies the meaning of this attribute.

For the `vopack`, `folder` and `compressedFolder` values, a new `vopack.xml` file has to be provided for their description. This allows for meta-packaging of ready-made VOPacks.

For the first three types, the `informationPath` attribute gives an XPath to the actual data being pointed, just in case the `packUnit` contains several tables, and not all of them are to be considered. In the case of FITS files, the informationPath looks XPath-like, but points to the HDU or Image holding the data.

Table 18 Meaning of the different valid values for attributes of the `packUnit` data type.

packUnit	description
votable	The packed unit is a VOTable.
fits	The packed unit is a FITS file.
vopack	The packed unit is itself a VO pack. The characterisation of the referencing VO pack encompasses all packed units, while each particular one will have its own, *narrower* characterisation.
compressedFolder	The referenced packed unit is a compressed directory.
folder	The referenced packed unit is a directory in the same file-system as the referencing VOPack.
otherXML	An XML representation, other than a VOTable, is used.
otherNonXML	A non-XML representation, also different from a FITS file, is used. This kind of representation should be avoided, but would be useful for packing instrument specific raw data formats, which are correctly characterised.

The VOPack XML Schema has been inspired by the concepts of Digital Items, Digital Item Containers, and Digital Item Components from MPEG-21 (Bormans and Hill, 2002), with the aim of being able to reuse the CharDM for direct package and sub-package inspection and selection.

The VOPack concept is orthogonal to the rest of RADAMS' classes, and can be used independently.

7 RADAMS Uptake and Implementations

The RADAMS has been implemented, so far, in two radio astronomical archives: the TAPAS archive for the IRAM 30m radio telescope[11], and the DSS-63[12].

TAPAS is a complete archive implementation for the IRAM 30m, with special emphasis in keeping observation metadata as modelled by the RADAMS, and the capability of storing (and delivering) data for large programs in the future. It covers both heterodyne and bolometer observations for both single-pixel and multi-pixel instruments, and has shown its modularity by the recent incorporation of the EMIR heterodyne backends[13], substituting the old ABBA ones.

Figure 8 shows TAPAS output for a given observation, with the relationship of the different sections to the RADAMS highlighted.

The DSS-63 archive contains the observations in K-band (18 to 26 GHz) of the DSS-63 antenna taken during 2006. The system is prepared for future extensions to other ranges, in particular the Q-band (40 to 50 GHz) and Ka-band (32 GHz), and forms part of the work developed through a collaboration between the AMIGA group (IAA-CSIC) and CAB (INTA-CSIC) for the integration of radio astronomical archives in the Virtual Observatory.

---

[11] https://tapas.iram.es/tapas/

[12] http://sdc.cab.inta-csic.es/robledo/

[13] http://iram.fr/IRAMFR/ARN/sep09/node4.html

8 Conclusions & Future Work

We have performed a review of the VO enhances astronomer's interaction with astronomical archives, and how data in those archives become interoperable thanks to a unified description and semantics, and a shared data model.

The IVOA has proposed several data models pertaining to astronomical observations, specially the description of the spatial, temporal, and energy parameters defining observations. However, there are no complete instantiation of the complete IVOA data model for observations, and many radio astronomical specifics are not taken into account in existing models, hence the need for the RADAMS.

Using IVOA principles drafted by the IVOA Data Modelling WG (McDowell et al., 2005), and the data model for astronomical data Characterisation (IVOA Data Modelling WG, 2008), we have developed a data model which includes Provenance, Policy, Curation, and Packaging classes.

Those added classes are completely modular, both in their expression in database form as in their XML serialisation, as they use their own UTypes, and as such they can be safely ignored by applications not able to understand such metadata. Radio astronomical specificities are captured in the Provenance class, and the rest of the classes remain unchanged. In that way, the RADAMS does not need to be an IVOA standard, but can be used as a guide on how to provide the missing observation information from IVOA's ObsDM framework for radio astronomy, but also for additional bands outside of the optical domain.

The RADAMS has been validated by being used as the basis for the development of two radio astronomical archives, those for the DSS-63 and IRAM 30m antennas. Those archives back-up the feasibility of using IVOA data models as the basis for the development of new astronomical archives, helping with their interoperability.

These archives have needed a complete set of metadata for their XML serialisations, consisting of target FITS keywords for the Data-filler, and Unified Content Descriptors (UCDs) for all attributes. Many UCDs have been built by means of juxtaposition of existing ones, so that the resulting UCDs are more specific than their individual atoms. A few UCDs have been proposed for addition to the UCD1+ vocabulary.

We have presented the different modules making up the RADAMS for approval to the IVOA Data Modelling Working Group, in order to provide a reference implementation of an IVOA-based observational data model for radio astronomy.

We plan to further generalise the RADAMS with support for radio interferometry, and time-series data (i.e., for pulsar observations).

Finally, we intend to formulate the Provenance relationships in terms of the Open Provenance Model (OPM; Moreau et al., 2010), so that OPM based tools can be used to derive attributions, lineage, and provide quality metrics.

The SQL source code, and the JSON file describing the RADAMS meta-model are available from the authors on request.

References

Alvarez, C., Cuevas, O., Curé, M., Dougnac, C., Marín, J., Pozo, D., Ortiz, J., and Oyanadel, A. (2010). Project: Study of Precipitable Water Vapor (PWV) at Llano de Chajnantor. Final report, Universidad de Valparaíso, Facultad de Ciencias, Departamento de Física y Astronomía.

Bormans, J. and Hill, K. (2002). MPEG-21 Overview. Technical Report v.5, ISO/IEC JTC1/SC29/WG11/N5231.

Christian, D., Abney, F., Comeau, T., Hanisch, R., Harrison, J., Imhoff, C., Kidwell, R., Kimball, T., Levay, K., Padovani, P., Postman, M., Richon, J., Smith, M., and Thompson, R. (1999). The Multimission Archive at Space Telescope. In D. M. Mehringer, R. L. Plante, & D. A. Roberts, editor, *Astronomical Data Analysis Software and Systems VIII*, volume 172 of *Astronomical Society of the Pacific Conference Series*, pages 233–+.

Derriere, S., Gray, N., Mann, R., Preite Martinez, A., McDowell, J. C., Mc Glynn, T., Ochsenbein, F., Osuna, P., Rixon, G., and Williams, R. (2004). UCD (Unified Content Descriptor) - Moving to UCD1+. IVOA Working Draft 1.03, International Virtual Observatory Alliance.

Ferraiolo, D. and Kuhn, D. (1992). Role based access control. In *Proceedings of the 15th Annual Conference on National Computer Security*, pages 554–563, Gaithersburg, MD. National Institute of Standards and Technology.

Ferraiolo, D. F., Cugini, J. A., and Kuhn, D. R. (1995). Role-Based Access Control (RBAC): Features and Motivations. In *Proceedings of the 11th Annual Conference on Computer Security Applications*, pages 241–248, Los Alamitos, CA, USA. IEEE Computer Society Press.

FITS Working Group (2008). Definition of the Flexible Image Transport System (FITS). FITS Standard Version 3.0, Commission 5: Documentation and Astronomical Data-International Astronomical Union, http://fits.gsfc.nasa.gov/iaufwg/.

Hanisch, R. J., Arviset, C., Genova, F., and Rino, B. (2009). IVOA Document Standards. IVOA Proposed Recommendation Version 1.2, International Virtual Observatory Alliance.

Hanisch, R. J. and Quinn, P. (2003). The International Virtual Observatory. Technical report, International Virtual Observatory Alliance.

Hirschheim, R., Klein, H. K., and Lyytinen, K. (1995). *Information Systems Development and Data Modeling: Conceptual and Philosophical Foundations*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press.

IVOA Data Modelling WG (2008). Data Model for Astronomical DataSet Characterisation. IVOA Recommendation v1.13, International Virtual Observatory Alliance.

Kamp, I., Thompson, R., Conti, A., Fraquelli, D., Kimball, T., Levay, K., Shiao, B., Smith, M., Somerville, R., and White, R. L. (2005). MAST in the Context of VO Activities. In P. Shopbell, M. Britton, & R. Ebert, editor, *Astronomical Data Analysis Software and Systems XIV*, volume 347 of *Astronomical Society of the Pacific Conference Series*, page 208.

Lamb, P. and Power, R. (2003). IVOA Data model for raw radio telescope data. IVOA Note, International Virtual Observatory Alliance.

Leon, S., Espigares, V., Ruíz, J., Verdes-Montenegro, L., Mauersberger, R., Brunswig, W., Kramer, C., Santander-Vela, J., and Wiesemeyer, H. (2012). TAPAS, a VO radio archive at the IRAM-30m telescope. *Experimental Astronomy, in press.*

Louys, M., Bonnarel, F., Schade, D., Dowler, P., Micol, A., Durand, D., Tody, D., Michel, L., Salgado, J., Chilingarian, I., Rino, B., Santander-Vela, J. D., and Skoda, P. (2011). Observation Data Model Core Components and its Implementation in the Table Access Protocol Version 1.0. IVOA Recommendation, International Virtual Observatory Alliance.

McDowell, J., Bonnarel, F., Giaretta, D., Lemson, G., Louys, M., and Micol, A. (2005). Data Model for Observation. IVOA Data Modelling WG Internal Draft v0.23, International Virtual Observatory Alliance.

McDowell, J., Tody, D., Budavari, T., Dolensky, M., Kamp, I., McCusker, K., Protopapas, P., Rots, A., Thompson, R. W., and Valdés, F. (2007). IVOA Spectral Data Model. IVOA Recommendation 1.03, International Virtual Observatory Alliance.

Moreau, L., Clifford, B., Freire, J., Gil, Y., Groth, P., Futrelle, J., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Simmhan, Y., Stephan, E., and Van den Bussche, J. (2010). The Open Provenance Model — Core Specification (v1.1). *Submitted to Future Generation Computer Systems.*

Moses, T. (2005). eXtensible Access Control Markup Language (XACML). OASIS Standard v 2.0, OASIS.

Muders, D., Polehampton, E., and Hatchell, J. (2005). Multi-Beam FITS Raw Data Format. Technical report, Max-Planck-Instituts für Radioastronomie.

Murphy, T., Lamb, P., Owen, C., and Marquarding, M. (2006). Data Storage, Processing, and Visualization for the Australia Telescope Compact Array. *Publications of the Astronomical Society of Australia*, 23:25–32.

Ochsenbein, F. and Williams, R. (2009). VOTable Format Definition. IVOA Recommendation Version 1.2, International Virtual Observatory Alliance.

Osuna, P., Guainazzi, M., Salgado, J., Dubernet, M.-L., and Roueff, E. (2010). Simple spectral lines data model. IVOA Proposed Recommendation v1.0, International Virtual Observatory Alliance.

OUP (2005). *The New Oxford American Dictionary.* Oxford University Press, second edition.

Pardo, J. R., Cernicharo, J., and Serabyn, E. (2001). Atmospheric Transmission at Microwaves (ATM): an improved model for millimeter/submillimeter applications. *IEEE Transactions On Antennas And Propagation*, 49(12):1683–1694.

Preite Martinez, A., Derriere, S., Delmotte, N., Gray, N., Mann, R., McDowell, J., Mc Glynn, T., Ochsenbein, F., Osuna, P., Rixon, G., and Williams, R. (2007). The UCD1+ controlled vocabulary. IVOA Recommendation Version 1.23, International Virtual Observatory Alliance.

Prestage, R. M. and Clark, M. H. (2004). Device and Log FITS Files for the GBT. Technical report, NRAO Green Bank.

Rots, A. (2007). Space-Time Coordinate metadata for the Virtual Observatory. IVOA Recommendation 1.33, International Virtual Observatory Alliance.

Ruíz, J., Santander-Vela, J. D., García, E., Espigares, V., Leon, S., and Verdes-Montenegro, L. (2008). Virtual Observatory activities in the AMIGA group.

ArXiv e-prints.

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

Tody, D., Dolensky, M., McDowell, J., Bonnarel, F., Budavari, T., Busko, I., Micol, A., Osuna, P., and Vales, F. (2008). Simple Spectral Access Protocol. IVOA Recommendation Version 1.04, International Virtual Observatory Alliance.

Tody, D. and Plante, R. (2009). Simple Image Access Protocol. IVOA Proposed Recommendation 1.0, International Virtual Observatory Alliance.

Verdes-Montenegro, L., Sulentic, J., Lisenfeld, U., Leon, S., Espada, D., Garcia, E., Sabater, J., and Verley, S. (2005). The AMIGA project. I. Optical characterization of the CIG catalog. *Astronomy & Astrophysics*, 436:443–455.

Viallefond, F. (2006). The Alma Science Data Model. In Gabriel, C., Arviset, C., Ponz, D., and Enrique, S., editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, pages 627–+.

Weise, A., Wan, M., Schroeder, W., and Hasan, A. (2008). Managing groups of files in a rule oriented data management system (iRODS). In Bubak, M., van Albada, G., Dongarra, J., and Sloot, P., editors, *Computational Science - ICCS 2008*, volume 5103 of *Lecture Notes in Computer Science*, pages 321–330. Springer Berlin / Heidelberg.

Williams, R., Hanisch, R., Szalay, A., and Plante, R. (2008). Simple cone search. IVOA Recommendation Version 1.03, International Virtual Observatory Alliance.
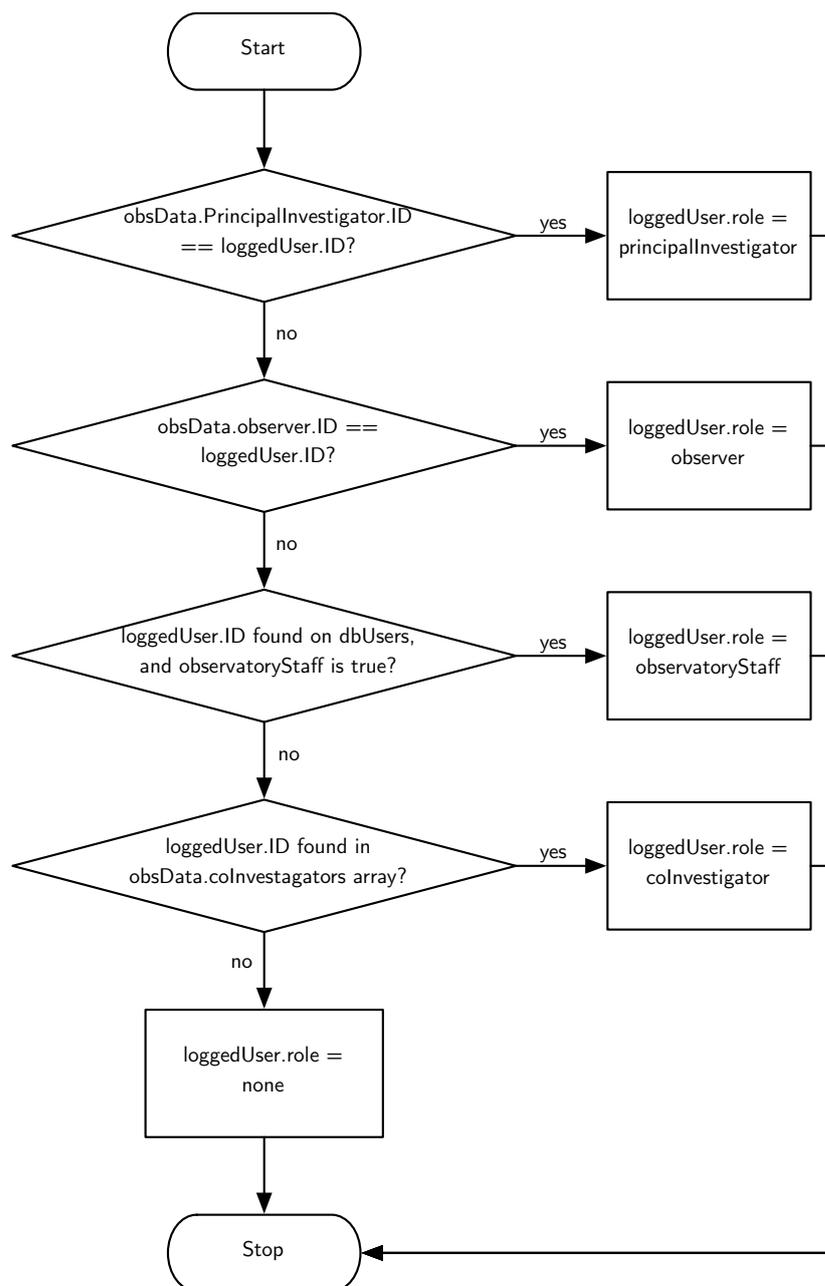
Fig. 3 Flow diagram for the role determination algorithm.
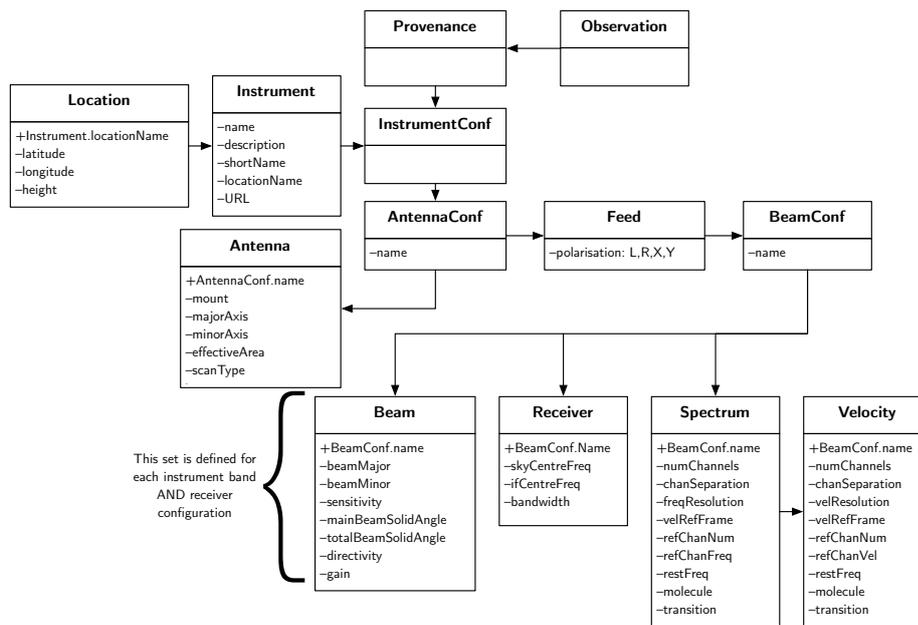
Fig. 4 Provenance.Instrument data model.
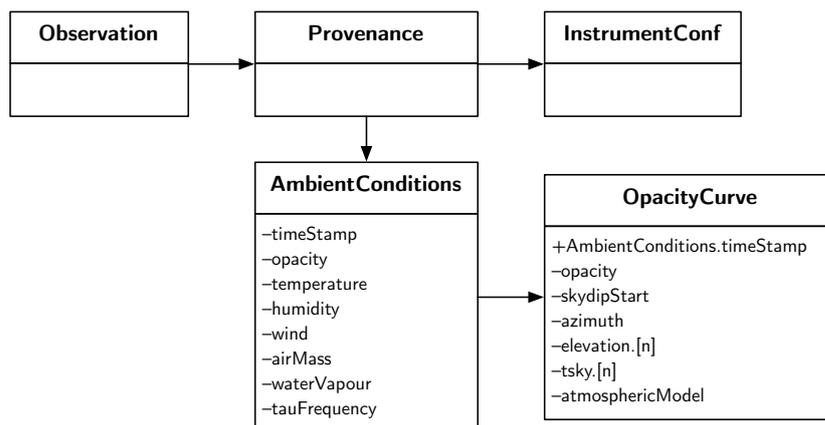
Fig. 5 Provenance.AmbientConditions data model.
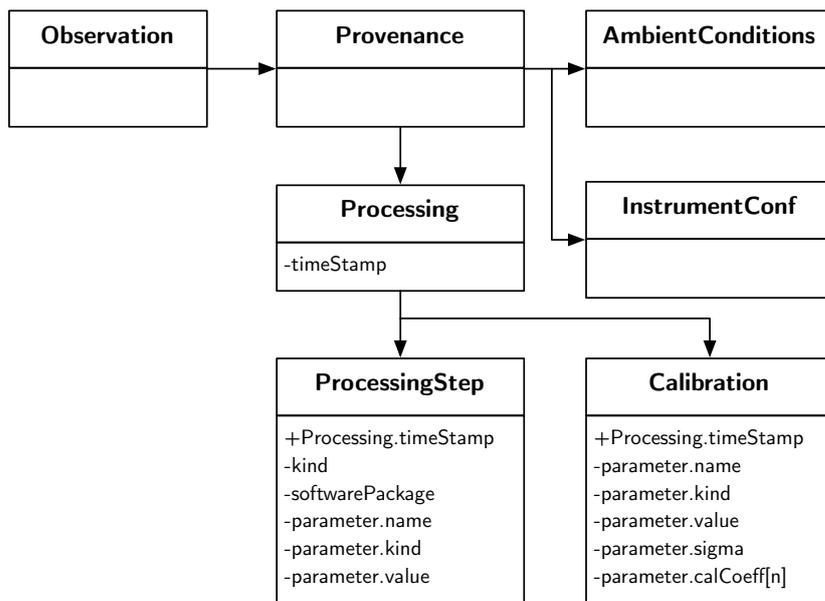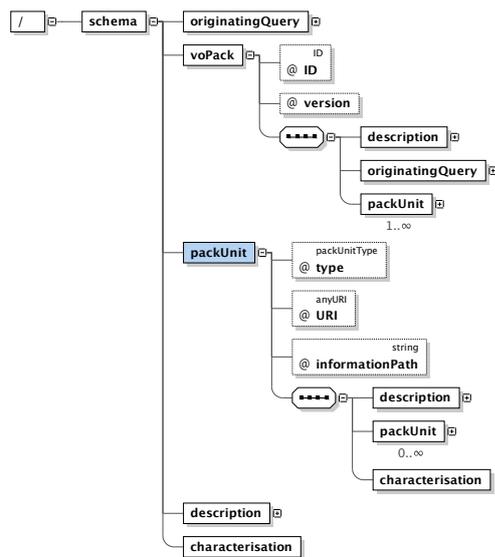
Fig. 6 Provenance.Processing data model.

Fig. 7 VOPack structure. Diagram generated by **Oxygen** from the XML schema.

Fig. 8 Details for a scan belonging to an OTF map, indicating the relationship of each metadata element with RADAMS submodels.