

ALMA Science Portal Policies

Draft version 0.1

Felix Stoehr, Archive/SciencePortal/UserPortal working group

Introduction

This document presents the policies for the set-up, distribution, evolution and content-access of the ALMA Science Portal.

The Science Portal represents the main entry point for the interaction of users with ALMA. It is intended to provide a one-stop access gateway to all ALMA web resources, documents and tools relevant to users for proposal preparation, project assessment, project tracking, project data access, data discovery and retrieval as well as to the ALMA Helpdesk.

The Science Portal provides access to restricted content or tools for registered and authenticated users depending on the user's role.

The term User Portal was a name for the Science Portal once a user has authenticated but is no longer used.

[Comments: the term "Science Portal" was coined during discussions in the ASA/UP working-group to merge "Science Web and User Portal". The main reason for having a new name was to indicate that there is actually just one single portal/web site which offers different functionality depending on whether users are logged-in or not. Any other name is possible.]

Policy for the access of content and tools

All public resources including public tools, public web content, public metadata and public science data shall be accessible anonymously.

Authentication and authorization shall be restricted to services and data access that by construction require it. This includes the Helpdesk, data access of proprietary data as well as tools that are available only as a function of the role of the user e.g. assessor tools, project tracker and Phase 1 Manager.

Alternative 1:

All users must be registered and authenticated to access all the contents of the Science Portal. Authenticated users will -- in addition to the public content and the Helpdesk -- have access to their proprietary data and depending on their role to specialized tools e.g. assessor tools, project tracker and Phase 1 Manager.

Alternative 2:

Allowing only for some TBD part of the content and functionality to be used anonymously.

The retrieval of science data from the ALMA archive by external archive users should be per preference over the network and preparation of media should be regarded as an exceptional case for special or very big requests. PIs shall always get their data delivered on media. This policy can change in the future if for PIs the network transfer of all their data becomes easily feasible.

[Comments: The practice of allowing anonymous access of data is becoming more and more the

standard for astronomical archives these days (VLA, Spitzer, HST, Gemini, SDSS, 2MASS, ...). However, there are also exceptions (VLT, ISO). Anonymous access is not only a requirement for the Virtual Observatory (VO) but also reduces the entry barrier for novice users. This seems important as the declared goal for ALMA is to make it easy for optical astronomers to use it. The main reason for requiring authentication for the download of public data seems to be able to compile better statistics. It turns out, however, that the relevant statistics can be obtained to rather high accuracy by analysing the download traffic. That functionality is already built into the ALMA request handler.

But see also COMP-70.50.00.00-001-F-DSN:

9.1.7: "Access to datasets in general is only granted to authorized users, but publicly available datasets are available to any authenticated user."

9.3: "Some external user operations like data retrieval requests or proposal submission require that the person who is executing them is registered and authenticated"

My feeling is, that anonymous download should be considered very seriously.]

Policy for the Science Portal deployment

There shall be three instances of the Science Portal publicly accessible: One at each ARC. This set-up is chosen to provide fast data access to the local user communities, spread of the load of web-, database- and data-access as well as to provide redundancy of the content and functionality for maximal uptime of the system as a whole.

Each Science Portal instance is connected to the local installations of the database, request handler and user-authentication server. This assures high availability even if one instance should be temporarily unavailable.

There shall be one single entry point to the ARC Science Portal instances located at JAO (e.g. almaobservatory.org/science or similar). From there users are automatically redirected to their geographically closest ARC Science Portal. Web links placed on each instance allow the users to switch to a different ARC Science Portal instance e.g. if some portal functionality is temporarily down at one ARC. The Science Portal instances at the ARCs must be reachable directly (and bookmarkable in the web-browser) so that the single entry point at JAO is not a single point of failure.

There shall be an instance of the Science Portal available at JAO allowing for for development (see Science Portal evolution) and possibly for JAO-staff access.

Alternative 1:

There shall be four instances of the Science Portal publicly accessible. One at each ARC with full capabilities and one at JAO with reduced functionality (i.e. no Helpdesk, user support, data download, data request on media).

[Comments: The more the I have thought about this, the more I believe having only three Science Portal instances at the three ARCs is overall the cleanest way forward. If there was an publicly accessible instance at JAO this would lead to a very inconsistent user experience as there will be no user support, no data download, no data delivery on media and no Helpdesk at the JAO as far as I know. However, if JAO had a Science Portal instance running, experience shows that many users would connect there as users will associate JAO (rightly!) with the ALMA science centre. The cleanest solution with the maximum of consistency is reached -- at least my eyes -- with a single science entry point at JAO which redirects to identical ARC portal instances. This also fits in with

the choice that the interaction with the users of ALMA is task of the ARCs and the Science Portal is that interaction mechanism.]

Policy for the Science Portal user experience

The user-experience of the Science Portal should ideally be at the state-of-the-art of scientific archives and technology to assure a maximum efficiency of the use of ALMA.

The instances of the Science Portal running at the different locations shall provide exactly the same user experience to assure ALMA-wide consistency of the user-interactions with the system. This includes the look, feel, layout, behaviour, content and functionality.

Localizations of the different instances other than connecting to the local technical architecture like database or NGAS servers are limited to a different additional logo on the portal as well as translations into additional languages (e.g. Japanese).

Content that is local to each ARC (e.g. visitor information, local news, ...) shall be reachable though a web link from each Science Portal instance. The local content will follow the web-policy of the respective ARC.

Alternative 1:

Content that is local to each ARC shall be placed into one point of the menu of Science Portal. The local content shall follow the ALMA Science Portal policies where they apply.

[Comments: the advantage of the proposed policy is that each ARC has full control over the content that is local. It allows to modify the local content frequently and without having to touch the operational Portal, provides a natural clear separation of the responsibilities and makes it easier to comply with local ARC web-policies (e.g. layout). Furthermore the ARC content can be placed in the local CMS instead of having to live in the Plone of the portal.

The advantages (depending on the viewpoint) of Alternative 1 is that the ARC content is forced to the same look and feel of the Science Portal providing perhaps an even more consistent user experience at least from the ALMA point of view (not from the ARC point of view). A second distinct technical advantage is that local content that needs authentication could be used with the single-sign-on provided by the portal.

Technically Alternative 1 could be implemented by having all local content of the ARCs in the global Science Portal distribution and when an instance is installed locally, the local content of the other two ARCs is removed (or hidden) in Plone. That is very easy to do and could be done at the same time when the technical adaptation of the portal to the local environment is done.]

Policy for the Science Portal evolution

In order to assure the possibility of adapting the Science Portal to technological advances, feature requests from ALMA users or the invention of new ways to ease the interaction of users with ALMA, it is envisaged that the Science Portal will evolve over time.

There shall be a single central development instance (e.g. located at JAO) of the Science Portal available where new content and features are added and tested. Upon release of a new version the Science Portal instance shall be distributed to the ARCs. This process assures the consistency of the Science Portal instances and a homogeneous user-experience throughout ALMA by at the same

time minimizing development effort.

It is envisaged that new functionality and content if it is relevant for more than one ARC it will be added to the Science Portal instead of the local ARC web sites.

The responsibility for the evolution of the portal content lies at JAO together with the ARCs. The responsibility of the Science Portal functionality lies at ESO. TBD!

[Comments: Part of this policy is due to the fact that the ALMA Science Portal is based on Plone. The current understanding is that it is impossible to e.g. add a new feature to a local Plone service and then distribute it to the other ARCs. This is true for all content or functionality that are stored in the Plone database file Data.fs. In order to improve this situation, the archive group now uses Plone "products" which can more easily be added and removed from the system. This allows also some (limited?) development directly at the ARCs without preventing the result of the development to be rather easily integrated into the central development/integration system.

As for the responsibility separation, the relevant parties (JAO, ARCs and ALMA construction) should comment please.]

Policy for User Registration

For policies about registration and password reset: see Mark Lacy's document "ALMA Archive Policies" archive_policy_a1p3.doc

A single-sign-on system assures that all registered users can authenticate at each of the three instances with the same credentials. The user registration, however, is done centrally at JAO and the user databases are replicated to the ARCs on very short timescales.

The ARCs are encouraged to provide single-sign-on systems so that registered and authenticated users of the local ARC portals are automatically registered and authenticated at the ARC instances of the ALMA Science portal.

Policy for Virtual Observatory access

There shall be access to Virtual Observatory services available at at least one but potentially all three ARCs. The kind services that will be offered will evolve over time. It is envisaged that the first services to be offered will be TAP (and/or ObsTAP) as well as SIAPv2 for reduced data cubes. The Virtual Observatory access can be part of the Science Portal or be a separate service depending on the practicality of both possibilities.

[Comments: As the VO services need querying and data delivery capability they are probably best situated at the ARCs. Should we have more than one service serving the same data available? The advantage would be increased redundancy, the inconvenience that users will see several times the same data offered in the VO tools.

VO access to data is dealt with in "ALMA Archive Policies" archive_policy_a1p3.doc already.]

Policy for the Science Portal Look&Feel&Functionality

[Comment: Do we need policies for this? Probably that's all implementation.]