



MEETING

Meeting Date 11-12 June 2018

Ref MoMAAUG#1

Meeting Place ESAC, D52 /Videoconf

Chairperson Beate Stelzer

Minute's Date 24 August 2018

Participants

AAUG Members : Beate Stelzer (Chair), Hervé Bouy, Ioannis Georgantopoulos, Søren Larsen, Marc-Antoine Miville-Deschenes, Sandra Savaglio, Eva Villaver

In attendance : Deborah Baines (ESDC Astronomy Archives Science Lead), Guido de Marchi (ESDC Science Lead), Jan Tauber (Planck Project Scientist and ESA Astronomy Study and Project Scientist's representative), Bruno Merin (Head of ESDC), Jesus Salgado (Astronomy Archives Technical Lead), Marcos Lopez (Planck Archive Scientist and User Community Scientist), Alcione Mora (Gaia Archive Scientist).

WebEx: Eva Verdugo, Pedro Osuna, Guillaume Belanger, Daniela Coia,.

Absent: Christopher Conselice (AAUG member) sent his apologies.

Subject Minutes of Astronomy Archives User Group (AAUG) meeting #1.

Copy

Description	Action	Due Date
Edited by Deborah Baines Approved by AAUG members on 2 nd October 2018		

WELCOME:

B. Stelzer (Chair) and D. Baines (ESDC Astronomy Archives Science Lead) opened the meeting on Monday 11th June 2018. Since this was the first Astronomy Archives User Group (AAUG) meeting, all members were welcomed to the group and a roundtable of introductions was performed. The role of the AAUG and its main tasks were presented and a brief background on the ESAC Science Data Centre (ESDC) was given.

ADOPTION OF THE AGENDA:

The agenda of the meeting was presented and adopted by the AAUG members.

PRESENTATIONS:

The following presentations were given and discussion sessions were held on 11th and 12th June 2018:



1. ESA Project Scientists (Jan Tauber)
2. ESDC website and discussion: feedback on archives from members (All)
3. EHST description (D. Baines)
4. ESASky demo and discussion (D Baines)
5. Archives long-term strategy (G de Marchi)
6. Enhancing visibility and usage of ESA archives
7. Date and time of next meeting (All)
8. AOB
9. Closed session

The slides of the presentations are available on the ESDC public web site, under the heading Archives User Groups, Astronomy:

<https://www.cosmos.esa.int/web/esdc/aaug/meeting1>

DISCUSSIONS:

Discussions were held during both days of the meeting, during and after presentations, and in dedicated discussion sessions. This section gives a summary of the discussions and is ordered by the above presentations and discussion sessions.

1. ESA Project Scientists

A presentation was given by J. Tauber, introducing the role of the ESA Project scientists, their relationship with the mission archives and important points to note. After the presentation, AAUG members asked how the experience that the Project Scientists acquire on one mission regarding the archives and solutions they find are eventually fed to ongoing or future missions. It was explained that every archive is very specific, but the process of going through the development and eventually operations of an archive is similar for all the missions, and there are a lot of things that can be learned from one mission to the next, including the integration of the archives in the mission. Much is done to see synergies but often one size does not fit all, however, the lessons learned are a very important product of a mission.

The question of how to preserve all the data and associated materials of a mission in the long-term was raised, e.g. are users going to be able to access and interpret current mission data in 100 years' time. This led to a discussion on standards continuing into the future, e.g. the FITS standard can already be called 'old' but it is being updated and still used today and the AAUG were told that the planetary missions are currently converting their data from the PDS3 standard to PDS4. Data centres need to periodically assess what is in their archives and ensure that everything is still accessible, and if needed, update data to the latest standards. It was also explained that one of the most difficult things to keep up to date is the mission analysis software, and it might be the first thing that becomes obsolete. ESAC are trying to look into solutions using a system called the Science Exploitation and Preservation Platform (SEPP) that offers the ability to have Virtual



Machines, in which the science user can run code and software in old Operating Systems. More on the SEPP project was discussed in the long-term strategy session (#5).

The AAUG asked how missions handle updating the requirements on an archive and it was explained that there are formal reviews of the missions, where the archive is a part of the review, and also there are continuous internal processes within the teams looking at what's possible. The UG also asked if the Project Scientist exists in the legacy phase of a mission. It was explained that Project Scientists never lose their title whilst working at ESA and do keep providing information, however, more than half of all the ESA missions collecting space data are in the legacy phase, and this fraction is only going to grow. At the moment ESA try to cope with this by having people who were formally associated to a certain mission, contact scientists, that are still on site and can provide information, but in around 10 years from now this will be very difficult to achieve because most will be retired. This is an issue that has not yet been solved. A discussion then followed on the importance of the mission documentation and explanatory documents.

2. ESDC website and discussion: feedback on archives from members

It was noted by the AAUG chair that the time evolution of sources is still missing in most of the astronomy archives and that there are no time series data in most of the archives. The inclusion of this data into the archives is strongly encouraged and is included in recommendation 2018-06-12/01.

A lengthy discussion followed on the topic of EU funded projects producing high-level science products from ESA mission data and whether, and how, the data is going to be ingested into the relevant archives. It was explained that the ESDC is interested in receiving these higher level science products and a team at ESAC is setting up a proof of concept system that would allow users to have their own space, in a place close to the archive, to upload their own high level data products and software and when ready make this data publically available via the archives. It is a system that wasn't initially conceived for this purpose but can easily be used for it. ESDC is looking for any inputs from the AAUG regarding adding these high level science products from EU projects into the archives and are proposing to prepare a handbook on how the data should be prepared in order to most easily ingest it into the archives. How many projects will be interested in providing their data is the next step to find out. The UG were told about two H2020 projects producing high level products from Planck data (RADIOFOREGROUNDS and Beyond Planck) that have already been contacted by ESA and are both interested in adding their data to the Planck archive. UG members gave examples of two more EU funded projects producing high level data products from ESA mission data: ARCHES and EXTraS, both FP7 projects using XMM-Newton data. Both of these projects have finished and in the meantime XMM-Newton continues to take data, therefore the UG suggested in these cases it may be more important to preserve the software and pipelines created by these projects (see **Recommendation 2018-06-12/03**)

The UG asked if ESA have statistics on the papers that are produced using archival data. It was explained that ESA collect this for many of the astronomy archives, in particular XMM-Newton, Herschel and the Hubble Space Telescope. The Planck mission also have a record of this. The example of the number of HST publications versus year was shown, where the number of archival papers can clearly be seen to be increasing



over time. We see this trend with the other missions but it takes a number of years. There is an ongoing project called the Science Archives Publication System which when finished will be able to provide this information for all missions. For many Planetary and Heliophysics missions, and for survey missions, there is no such concept as an observing program which is associated to a specific proposer. A general discussion followed on the archives: the need for one portal to all the data (one of the concepts behind ESASky); limitations with the archives (such as the lack of proprietary HST data); and the need to include added value functionalities to attract users to the ESA archives (if the mission data is also being provided elsewhere by external data providers).

UG members brought up the topic of programmatic access to the archives and believe it is going to be more and more important as a growing fraction of the community need programmatic access. As an example, within a few months almost all users working with Gaia data in the stellar community have learnt ADQL. The UG like that `astropy/astroquery` can be used to access some of the ESA archives. It was noted that in a few decades we do not know what language users will be using to access the data, therefore the archives need to adapt to be accessible from the latest programming languages used by the community. It was also noted that it is very useful for users to have simulated data.

Other points that were discussed were that the planning and development of the archives now start during the development phase of a mission and in most cases, by the time the mission is launched the archive is ready. The ESDC use VO protocols wherever possible within the archives. The UG were interested in having external catalogues available in the relevant archives and in ESASky. The UG agreed that it is very important users know when using ESASky, that they are looking at a selection of data (high level science products) from each mission, and that far more data is available via the individual archives.

A discussion around the pros and cons of other multi-mission portals developed by external data centres followed. Examples were given, such as `W3browse` from HEASARC, which is used by most of the x-ray community because users can find all high-energy data via the platform; MAST at STScI, etc. Further efforts to enlarge the accessible data base by adding other major non-ESA missions and astronomical catalogues to ESASky are recommended.

Finally, a discussion on interactive data analysis tools within the archives was raised. The XMM-Newton Science archive now links to the interactive data analysis tool called RISA, users can perform some analysis online before deciding to download data. This is the concept behind the SEPP project (users bring their code to the data). Any example use cases for SEPP from the UG would be very welcome. The AAUG find the development of interactive data analysis tools a high priority.

3. EHST description

The presentation on the EHST was followed by questions from the AAUG members regarding why there are separate Hubble Space Telescope (HST) archives. The history of the reason for this was given, it made sense to have archives located around the world when the internet broadband was very slow. Now this isn't an issue



but it still makes sense to have the data in different locations. And the HST is not just a NASA mission, it's a joint NASA and ESA mission and it's important to serve the needs of the European astronomical community. The pros and cons of having different HST archives around the world were discussed and it was suggested to include features in the EHST that are not available in any other archive providing HST data. Example functionality in the EHST was shown where users can very easily search for all observations that have no known associated publications. The distribution of EHST users from around the globe accessing the EHST was shown, and the majority of users are from the US and Europe (at roughly the same percentage depending on the month).

ESASky was briefly discussed as it does contain access to HST data and links to the EHST. It would have been more difficult to include HST data in ESASky if there was no EHST archive. Also to note, ESASky does not access proprietary data, users need to go to the individual archives for this.

4. ESASky demo and discussion

Questions were raised and discussed during and after the presentation and demo of ESASky. Related to the accesses from around the globe, UG members asked why there are so many users from Spain and Italy. Reasons for this are partially because the accesses from ESAC have not been cleaned from the Google analytics, which affect the Spanish numbers. However the reason is also to do with the tool being publicised more in Spain and Italy, and the press within both countries have reported on the tool. The UG asked what the number of users is. On a typical day there are around one hundred users; peaks have been around a few thousand.

The UG commented that currently ESASky appears to be trying to fulfill two goals: both a public outreach and education tool; and a tool for science users and scientific projects. Currently, to a scientific user the tool appears more like an outreach and education tool, and the science community at large is not very aware of ESASky. To address this, the UG suggested that the functionalities serving the needs of these distinct communities be identified and possibly be separated in the configuration of the user interface.

The UG asked if it would be easy to upload an image into the tool. It was explained that currently it is not possible but it is interesting and the ESASky team have received this request from other users. Finally, the UG asked for the possibility of adjusting the contrast of the background images, and a discussion followed around the possible solutions to do this. See **Recommendation 2018-06-12/01** and **Recommendation 2018-06-12/04** for the recommendations regarding ESASky.

5. Archives long-term strategy

The presentation on the ESAC Science Archives long-term strategy was given by G. de Marchi. A long-term strategy document was written in 2013 which is currently being reviewed and updated. The results and ideas coming from an internal review were shown to the UG members and discussed. The ESAC Science Archives



strategy falls into 3 pillars: 1) Enable maximum science exploitation of data sets; 2) Ensure efficient long-term preservation of data, software and knowledge, using modern technology; 3) Build cost-effective archives, also through integration in, and across, projects.

A discussion followed around the ideas in the first pillar, in particular the idea to engage early career scientists through research proposals to present their research on a recent publication using archival data at conferences, and funding their trip to a conference. The UG was very welcome to this idea (see **Recommendation 2018-06-12/02**). Proposals should be evaluated on the science, and for the first round, perhaps the call should be for recent usage of the archives (rather than waiting for researcher to write their paper). The committee that evaluates these proposals should try and balance out the number of proposals using different astronomy archives, and not accept too many for one archive. Currently, the numbers being discussed are to support around 50 early career scientists per year (covering travel costs, registration fees, etc.). The UG commented that it is important to keep the turn around time short (i.e. for PhD students). Possibly there should be two calls per year, or it could be a constant open call, this needs to be explored. One idea would be to release the call in conjunction with EWASS and put the call on the EWASS site. This would make a link with the EAS strengthening the call as a European initiative. The UG noted, however, that other more specialized conferences might be more attractive and have higher visibility among young researchers such that a broader advertising seems also useful. The UG asked if the plan is to open the call to the whole world, or only to ESA member states; and whether the funding is only to conferences in Europe, or anywhere in the world. It was commented that both points are very interesting, have not been investigated yet, and need to be followed up.

The UG suggested holding summer schools for ESA archive usage, also for early career scientists, and a tandem between the proposals idea and workshops would be good. Another comment was made on outreach, and that it would be interesting to develop science projects using the archives for high-school students, and possibly amateur astronomers. It was commented that this idea will be passed on to ESA education.

The UG asked about funding one ESA Research Fellow specifically for archival research. It was explained that this had been looked into and requires considerably more money than funding to conferences. The question is what is more effective, funding one RF or funding 50 scientists to conferences.

For the second pillar, the long-term preservation of data, software and knowledge, it was explained that a legacy manual will be prepared that addresses what, and in which format, the team can most easily ingest data. This will be aimed at groups producing high-level science data, and also for 'offsite' archives, e.g. small missions. The UG ask if there are any plans for the future to make this more efficient, and not have offsite archives. It was explained that the ESA contribution to some small missions is limited, and not intended to be used for operations, therefore during operations there will be an offsite archive, but ESA is interested in hosting the data once the mission is in the legacy phase.

Slides were shown on the ranking of all the proposed ideas on the long-term strategy for the ESA archives. These were ranked internally by about 35 people (scientists and software engineers). For the first group (high impact, low cost), the UG commented that a User Survey might be very useful, and would give feedback in the short-term (see **Recommendation 2018-06-12/04**). The UG noticed that social media is missing in the



ranking. It was commented that the archives team currently have a twitter account. The UG suggested putting announcements and news on social media astronomer groups. It was explained that the ESDC also coordinates with the ESA science communications on articles and images related to the archives. The second group of long-term strategic ideas (considerable development needed, higher costs) include: collaborative areas (proof of concept); analysis tools integrated in the archives; DOIs to improve linking to papers etc.; access to some housekeeping data; and defining metrics and housekeeping. The UG commented that competing analysis tools already exist, and it will take a lot of effort to develop new tools. It was commented that working in collaboration with external groups could be a solution for this. The UG commented that access via python for all archives is of high interest. Another comment was to look at CERN who has a very interesting platform for users to share their data. The UG commented that it would be very useful to have the bibliometrics for all the archives (such as the example shown for the HST in the earlier session). It was commented that the Science Archives Publication System project, once finished, should be able to provide this information for all mission publications.

Finally, it was asked if the UG members have heard of the European Open Science Cloud (EOSC). One UG member had (via the SKA). This project is sponsored by the European Commission and is a federation of clouds for research data in Europe. Groups can add data and run processes in this system. The only astronomy project currently involved, that we know of, is LOFAR. The project is still in an early stage, the kick-off was January 2018, and the idea is it will be a collaboration environment where users can share their work, processes, data and papers. See **Recommendation 2018-06-12/02** for all recommendations regarding the long-term strategy for the ESA archives.

6. Enhancing visibility and usage of ESA archives

After the presentation on enhancing the visibility and the usage of the ESA astronomy archives, UG members asked if the statistics given to ESA's Science Programme Committee are cleaned for internal ESA users. It was explained that these statistics are cleaned for internal ESA users (for example, to remove usages related to internal testing on the archives), but users from archive consortiums, external to ESA, are not removed. The UG asked, in terms of network load, which is the most bandwidth consuming archive? It was explained that currently this is Gaia, then the Heliophysics and Planetary physics archives. The reason for Gaia is because of the large number of users, for June 2018 there were around 2 million queries sent to the archive. Gaia DR3 will be at least 10 times larger than DR2.

The UG members asked if they can receive these SPC reports, raising the following action:

Action 2018-06-12/01: On G. de Marchi and D. Baines to ask if the AAUG can receive a copy of the SPC reports. If the answer is positive, send the AAUG the recent reports and subsequent triannual reports.

The UG commented that they would be interested in seeing the number of users that access the archives via ESASky. **Action 2018-06-12/02:** On D. Baines to give this feedback to the ESASky team.



7. Date and time of next meeting

12-13 February 2019, in ESTEC, The Netherlands. Details to be finalised, and it was suggested to have some telecons before the February meeting.

Action 2018-06-12/03: On G. de Marchi and D. Baines to ask for a room to be preliminarily booked at ESTEC for these two days.

RECOMMENDATIONS

The following recommendations, grouped by topic, were formulated by the AAUG:

ESASky

Recommendation 2018-06-12/01: The UG is impressed by the progress made in the development of ESASky, the web interface to the astronomy archives. Further efforts to enlarge the accessible data base by adding other major non-ESA missions and astronomical catalogues are recommended.

The planned inclusion of high-level science data products (such as lightcurves), along with the generation and ingestion of such data products in the individual mission science archives, is strongly encouraged.

Several features have been identified as potential tasks to be implemented in the future development of ESASky, among the most pressing being the possibility to tune the contrast of the images.

The UG noted that the approach to combine data access for science users and public outreach features in a single tool may be potentially deterring in the goal of establishing ESASky as a prime instrument for scientific data exploitation. It is recommended that the functionalities serving the needs of these distinct communities be identified and possibly be separated in the configuration of the user interface.

Action 2018-06-12/04: D. Baines to pass the ESASky requirements (and those given in recommendation 2018-06-12/04) to the ESASky team, include them in the project backlog and prioritise them in the ESASky roadmap. DB will provide feedback on the progress in the next AAUG meeting.

Long-term strategy for the ESA archives

Recommendation 2018-06-12/02: The UG recognizes that the long-term preservation of the astronomy archives, well beyond the mission lifetimes, is of paramount importance for the full exploitation of the missions' science potential. The decision of ESA to develop a strategy and the individual steps identified within ESA for the realization of this task are strongly endorsed.

Among the strategic steps the development of interactive data analysis tools and the possibility of programmatic access to the archives is considered to be of high priority.

The UG particularly welcomes the idea to promote the use and publication of archival data by funding archive-based science projects as part of the plan of maximizing the scientific output of the ESA astronomy



archives. A potential way of implementing such a scheme includes funding the active participation of junior scientists, including those from non-ESA member countries, in conferences with high visibility such as EWASS. The UG looks forward to seeing detailed plans for the implementation of such programs in the near future.

Action 2018-06-12/05: D. Baines to feedback to the astronomy archives teams the AAUG recommendation on the development of interactive data analysis tools and the possibility of programmatic access to the archives as high priority.

Action 2018-06-12/06: G. de Marchi and D. Baines to draw up detailed plans for the implementation of a programme to fund the active participation of junior scientists, including those from non-ESA member countries, in conferences with high visibility such as EWASS.

Inclusion of non-ESA data products / pipelines in the archives

Recommendation 2018-06-12/03: The UG recommends to explore ways to include in a systematic way higher-level science data products and pipelines developed from ESA data based on EU funded programs (e.g. FP7 and H2020) and other externally funded research teams in the astronomy archives. The purpose of this effort should be, next to preservation on long timescales, to make these data and analysis tools accessible through ESA science interfaces such as ESASky.

Action 2018-06-12/07: D. Baines to identify and produce a list of EU FP7 and H2020 projects producing high-level science data products from ESA astronomy mission data.

“Publicity” for the archives

Recommendation 2018-06-12/04: A User Survey to better understand the needs of the community of the ESA astronomy archives and to increase awareness for new archive developments, including ESASky, is encouraged. The UG ensures its support in the eventual realization of such a survey, preferably to be defined and carried out in short timescales.

In order to increase the visibility among scientists, ESASky should be promoted more aggressively as a tool for scientific data exploration.

The potential of ESASky as a public outreach tool is recognized. The development and implementation of science projects for high-school students is regarded as an interesting addition to the current and planned initiatives by ESA-education to enhance the visibility of its astronomy archives among the general public.

Action 2018-06-12/08: B. Stelzer and all AAUG members to define the plans for a user survey and questions to be asked to the community.

Action 2018-06-12/09: On D. Baines to send the results from the previous archives survey (in 2011), along



with the survey questions.