



MEETING

Meeting Date 1-2 October 2020

Ref MoMAAUG#5
ESDC-GEN-MOM-0017

Meeting Place WebEx

Chairperson Beate Stelzer

Minute's Date 20 November 2020

Participants

AAUG Members : Beate Stelzer (incumbent AAUG Chair), Christopher Conselice, Ioannis Georgantopoulos, Søren Larsen, Sandra Savaglio, Eva Villaver, Natalie Webb (new AAUG Chair)

In attendance : Deborah Baines (ESDC Astronomy Archives Science Lead), Guido de Marchi (ESDC Science Lead), Bruno Merín (Head of ESDC), Jan Tauber (Planck Project Scientist and ESA Astronomy Study and Project Scientist's representative), Bruno Altieri (Euclid Archive Scientist), Hector Canovas (Gaia Archive Scientist), Pierre Ferruit (JWST ESA Project Scientist), Raul Gutierrez (Astronomy Archives Technical Lead), Marcos Lopez (JWST Archive Scientist), Alcione Mora (Gaia Archive Scientist), Timo Prusti (Gaia Project Scientist), Elena Puga (ESDC Legacy Archive Scientist), Norbert Schartel (XMM-Newton Project Scientist)

Absent: Hervé Bouy (AAUG member) sent his apologies.

Subject Minutes of the Astronomy Archives User Group (AAUG) meeting #5.

Copy

Description	Action	Status
The UG is interested in seeing the number of users that access the archives via ESASky. D. Baines to report to the ESASky team.	2018-06-12/02	CLOSED
AAUG members to investigate one EU project each, investigate the type of data products from ESA missions that the project will be creating, advise ESDC what to ask the project PI, and if there is enough information at this point, recommend if the data should be added to any of the ESA science archives.	2019-02-13/13	IN PROGRESS
ESDC to contact the relevant EU project PIs.	2019-02-13/14	OPEN
G. de Marchi/ D. Baines to define a timescale for implementation of the visitor programme and the necessary interaction steps with the AAUG.	2019-10-14/01	ON HOLD
The AAUG to produce a report for the Astronomy Working Group (AWG) on the second year of AAUG activities.	2019-10-14/02	CLOSED.



Description	Action	Status
D. Baines to create an in-depth report on the user survey results and send to all stakeholders	2019-10-14/07	IN PROGRESS
D. Baines/G. de Marchi to email the AAUG after action 2019-10-14/03 is closed: send a doodle for the date of the next AAUG; undertake eventual replacements (if any).	2019-10-14/09	CLOSED.
D. Baines and G. de Marchi to provide a flow chart to the AAUG showing the steps in the process to be followed for each EU project evaluation.	2020-04-17/01	CLOSED
D. Baines to send out a doodle in about a month to AAUG members to find next meeting date.	2020-04-17/02	CLOSED
AAUG members to provide names for a new AAUG chair and new AAUG members.	2020-04-17/03	CLOSED
D. Baines and G. de Marchi to undertake the replacement for the 8 th member of the AAUG.	2020-10-02/01	
ESDC to send the Gaia statistics on the number of queries received through the different modes of accessing the archive (UI, cone search, astroquery).	2020-10-02/02	
D. Baines to examine the user survey answers from professional astronomers only and examine how they rated the archive and ESASky, and which preferences they expressed	2020-10-02/03	
D. Baines to update the usage graphs in the user survey report to show only the answers from the participants that use each archive and to add the number of users that responded per archive.	2020-10-02/04	
D. Baines to check the FP7 large projects for any using ESA data that may have been missed.	2020-10-02/05	
ESDC to begin contacting the PIs of the identified EU projects to gauge their interest in ESA preserving their project data.	2020-10-02/06	
D. Baines to circulate a doodle to find dates for the next AAUG meeting, to be held in Spring 2021.	2020-10-02/07	



WELCOME:

B. Stelzer and D. Baines opened the meeting on Thursday 1st October 2020 by welcoming everyone and summarising the main topics to be discussed during the meeting. The meeting was held via telecon due to the global COVID-19 pandemic. A roundtable of introductions was performed with the AAUG members and the upcoming AAUG chair, Natalie Webb.

ADOPTION OF THE AGENDA:

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

REVIEW OF ACTIONS:

The actions from the previous meeting were reviewed. It was agreed that action 2019-10-14/09 is closed, since the new chair has been found, and a new action has been added for the ESDC to find an 8th member of the AAUG (**Action 2020-10-02/01**). Four actions remain open with the following status: **Action 2019-02-13/13**: *AAUG members to investigate one EU project each, investigate the type of data products from ESA missions that the project will be creating, advise ESDC what to ask the project PI, and if there is enough information at this point, recommend if the data should be added to any of the ESA science archives.* This action is in progress and was discussed later in the meeting (see discussion session 4). **Action 2019-02-13/14**: *ESDC to contact the relevant EU project PIs.* Action to be completed based on feedback from Action 2019-02-13/13 and on a project by project case (see also discussion session 4). **Action 2019-10-14/01**: *G. de Marchi/ D. Baines to define a timescale for implementation of the visitor programme and the necessary interaction steps with the AAUG.* G. de Marchi reported the latest news regarding the Archival Research visitor programme. In summary, the launch of the visitor programme has been postponed to spring 2021 due to the COVID-19 pandemic. If this needs to be postponed again, ESA will ask the AAUG to discuss a possible ‘virtual’ version of the programme, to be added as a possible item for discussion in the next AAUG meeting. **Action 2019-10-14/07**: *D. Baines to create an in-depth report on the user survey results and send to all stakeholders.* This action is in progress. A draft of the report has been sent to the AAUG and the ESDC asked the AAUG for actionable recommendations (see discussion session 3).



PRESENTATIONS:

The following presentations and discussion sessions were held on 1st and 2nd October 2020:

1. Feedback from the AWG meeting; ESA Astronomy Archives session at EAS 2020 and plans for EAS 2021 (G. de Marchi, D. Baines, B. Stelzer)
2. The Gaia Archive, present and future (A. Mora)
3. Astronomy Archives User Survey (All)
4. EU Horizon 2020 and FP7 project assessments (All)
5. Demo of the Euclid archive cutout service (B. Altieri)
6. ESASky Statistics and updates (D. Baines)
7. Date and time of next meeting (All)
8. AOB
9. Closed session

The slides of the presentations are available on the ESAC Science Data Centre (ESDC) public web site, under the heading Archives User Groups, Astronomy: <https://www.cosmos.esa.int/web/esdc/aaug/meeting5>

DISCUSSIONS:

Discussions took place during and after each presentation. This section presents a summary of the discussions and is ordered by the above presentations and sessions list.

1. Feedback from the AWG meeting; ESA Astronomy Archives session at EAS 2020 and plans for EAS 2021

B. Stelzer began the session by introducing the slides presented to the Astronomy Working Group (AWG) in June 2020 on the AAUG activities. G. de Marchi then presented a slide summarising the ESA Astronomy Archives lunch session at EAS on Monday 29th June. The EAS 2020 ESA Astronomy Archives lunch session page can be found here: <https://eas.unige.ch/EAS2020/session.jsp?id=LS3> . About 100 EAS participants attended the session and there were a lot of questions about the visitor programme. Since the conference was virtual, the usual interactions at the ESA booth could not take place, where ESDC members normally talk with participants and give them demos and information about the archives and ESASky. An ESA Astronomy Archives lunch session is already reserved for EAS 2021 (28 June to 2 July) and it is planned to have a similar format as the session in June 2020. It is unknown yet if EAS 2021 will be in person or virtual and in either case ESDC hopes a more interactive discussion session can be held.



2. The Gaia Archive, present and future

In a previous AAUG recommendation (recommendation 2019-10-15/05), the AAUG welcomed ‘stronger participation by ESA staff scientists of the individual astronomy missions in the open sessions of the AAUG to share their experiences and ideas on archival developments.’ This session was therefore arranged to hear specifically from one of the ESA astronomy archives, in this case from the Gaia mission archive scientist, Alcione Mora, who gave a presentation on the Gaia archive in its present form and the plans for future developments.

The AAUG asked about the Gaia statistics that were presented, for example, the number of queries per month is not correlated with the number of users per month. A. Mora said he believes it is related to the cone search functionality on top of the current service, and also due to users querying via the Gaia astroquery module. B. Stelzer asked if they have the number or fraction of queries they receive through the different modes of accessing the archive? A. Mora replied that yes, they do have this, it is not however included in the presentation. Therefore, it was agreed that the ESDC will send these statistics to the AAUG (**Action 2020-10-02/02**).

A discussion followed on the desirable level of homogenization for the astronomy archives. A. Mora gave the example that some of the astronomy archives host some of the same catalogues. D. Baines noted this depends on the specific mission science use cases. E. Villaver added that if the uniformity is associated with simplicity, then it is a good move. Uniformity is a much easier way of accessing data for everyone, but if it comes at the expense of complicating the access for users, then it’s not a good thing. If ESA moves in this direction a lot of care needs to be put into the design of the tool. R. Gutierrez added that in this sense, ESASky and the archives are becoming more and more integrated with each other and so the intuitive, easy to use, interface can be achieved in ESASky, which provides the access to data from many archives. And on the other hand, all the specific mission data details are in the individual mission archives which are also connected to ESASky in a number of ways. It’s a two-way communication and both things may be balanced: the speciality of the data in the specific mission archives and the simplicity via ESASky.

Finally, B. Stelzer asked how Gaia uses the third-party service to distribute data. R. Gutierrez explained that this is done via a CDN (Content Delivery Network), for users who want to perform bulk downloads of the main Gaia tables. This service is through a contract with an external company. The CDN provides fast access to the data geographically near to the user, so it’s much faster in terms of data distribution, e.g. if the user is from the US, they get the data from the US. Much of the traffic is absorbed by this network and it reduces the network traffic at ESAC. The CDN is used when the network traffic is particularly high, i.e. it is only used straight after a Gaia data release.

3. Astronomy Archives User Survey

To start the session, D. Baines introduced the 2019 Astronomy Archives User Survey and summarised the results. B. Stelzer suggested that for future surveys, a question asking how the participants had heard about the survey will be very useful in order to know the best channels in which to advertise the survey. A brief discussion followed on how to obtain more responses from archive users for a future survey. Suggestions were given such as adding the survey to the banner of all astronomy archives (the 2019 survey had been



added to the ESASky banner), advertising the survey to national astronomy email lists and sending the survey to more subscribed users (that have agreed to be sent emails from ESA).

A discussion then followed about the survey results. Looking at the type of position or career stage of the survey participants, the AAUG pointed out that there were few PhD students and post-docs responding to the survey with respect to staff astronomers. Ways to enhance the reach of this important group of archive users should be studied (see **Recommendation 2019-10-15/02**). It was also noted that about 34% of the survey participants are not professional astronomers. While this is an interesting result, and shows the outreach potential of the archives and ESASky, it will also be useful to look at the responses from the professional astronomers only and examine how they rated the archives and ESASky, and which preferences they expressed (**Action 2020-10-02/03** and see **Recommendation 2020-10-02/01**). It was also noted positively by the AAUG that 9% of the survey participants were amateur astronomers. This group often performs good science and appealing to this community of users can be seen as a strength.

A discussion followed about the usage of the archives and the AAUG suggested that the usage graphs should be updated with the 'never used' removed, since each answer to the question 'Which ESA astronomy science archives do you use and how often' is archive specific. It was agreed that D. Baines update the usage graphs in the user survey report to show only the answers from the participants that use each archive and to add the number of users that responded per archive (**Action 2020-10-02/04**).

B. Stelzer noted that the results from the question on how the archives are accessed, either via the web interface or machine interface, show a difference between some of the archives. There is a higher percentage of Gaia archive users and XMM-Newton Science Archive (XSA) users who use the web interface, but still a large percentage (75% for all archives collectively) that use the web interface, showing how important the web interface is for users. D. Baines noted that ESDC have learnt that many users first come to the web interface, and some then learn the machine interface and scripting capabilities. Often, the majority of queries to the archives are coming from users running scripts and accessing the archive machine interfaces. Both web and machine interfaces are very important. D. Baines mentioned the high usage of the Gaia astroquery module, and this gave rise to a brief discussion on unifying the astronomy archive astroquery modules into one astroquery.esa module (something currently in progress between ESDC and astroquery).

A discussion followed about visualisation tools within the archives. Advanced web-based visualisation and cutout services came out as the top priority from the user survey question asking users to rate the planned and possible archival developments. Quick visualisation of data in the archives before users download the data, such as the interactive visualisation of spectra and timeseries data in the XSA, is seen by the AAUG as an attractive development for all archives and ESASky. It was noted that, as astronomical data becomes larger, it is becoming more and more unlikely that people will download data to their computer. Therefore, the archives need to have tools to allow users to select their data and download the subset of the data that they want to use. I. Georgantopoulos suggested access to data analysis, such as access to RISA in the XSA, is also very interesting for the archives. G. de Marchi pointed out that this is at the core of the Jupyter notebooks and ESA datalabs, which will allow users to run their own code and analysis close to the data. Regarding cutout services, B. Altieri confirmed that the current version of the Euclid science archive (not yet public) has a cutout service, and it was agreed that he would make a demo of this cutout service feature between the two sessions in the morning (see section 5).



4. EU Horizon 2020 and FP7 project assessments

D. Baines gave a brief overview, background and the current status on the work that has been done on the EU Horizon 2020 and FP7 project assessments. During the second AAUG meeting there was a dedicated discussion session on EU Horizon 2020 and FP7 projects that use ESA archival data and produce high level or user provided products and if these products should be preserved by ESA. 12 relevant large EU projects were identified before and during the previous AAUG meeting, 4 of which are already being followed by the relevant ESA missions (XMM-Newton and Planck). 2 further FP7 projects were later identified related to Herschel (HELP and VIALACTEA), resulting in 10 projects to be looked at by the AAUG. Two actions are open from this meeting (*Action 2019-02-13/13 and Action 2019-02-13/14*). D. Baines took an action to check the FP7 large projects again for any that may have been missed (**Action 2020-10-02/05**).

During the last AAUG meeting, the ESDC was asked to produce a flow chart to show the steps in the process to be followed for each project evaluation. D. Baines went through the flow chart. B. Stelzer summarized some of the difficulties encountered by the AAUG when assessing the projects. For example, within each project there is a wealth of product types, catalogues, simulations, light curves etc. and this is very inhomogeneous across the projects. There are very different numbers of objects addressed by each project, for example one project has only on the order of 80 stars, which arguably is only relevant for a limited part of the archive user community. Some projects are very specific, e.g. the project on solar system bodies. And many of these projects are multi-wavelength studies that involve a lot of non-ESA data, including ground-based data. This raises the question of whether ESA should host this kind of data. Nevertheless, most of the initial assessments by the AAUG members said that these projects look interesting, and others are interesting but only to a limited part of the community. B. Stelzer suggested that a re-evaluation of all the projects, following the flow chart, could help in unifying the assessments by the AAUG and in setting the priorities.

It was also pointed out that in some cases, the AAUG does not have all the information needed to assess the projects, for example when some online documentation is password protected. In these cases, interaction with the PI will need to be made at an earlier stage. A discussion followed on the products from the projects becoming obsolete, especially if they use data from missions in operations. ESDC described that the data, and documentation, from these types of projects will most likely be stored in the ESDC Guest Storage Facility and not in the individual archives themselves. It was also noted that many of the projects have additional tools, such as model simulations and analysis software, and somehow the tools should also be preserved. However, preserving the software brings up the issue of the software needing to have a license that allows it to be openly distributed. In these cases, the initial policy will probably be that ESA will not distribute any software from these projects.

Finally, a discussion rose that the initiative should be on the PIs and projects to contact ESA with the aim of publishing the project data in ESA, rather than the AAUG making an assessment on the projects. At this stage, it is unclear if the PIs or the projects are interested in ESA preserving the data for the community. The AAUG recommended that the ESDC should involve the PI of the EU project from the earliest stage possible in the evaluations (**Action 2020-10-02/06** and see **Recommendation 2020-10-02/03**).



5. Demo of the Euclid archive cutout service

B. Altieri, the Euclid archive scientist, gave a short demonstration of the Euclid Science Archive System, in particular, showing the cutout service in the current development version of the archive, capable of cutting out FITS images from the original Euclid images. Cutouts will also be possible with the archive command line functionality, where users will be able to query, for example, a catalogue with thousands of sources and download the small images around each source. This is a very important feature for Euclid since one single Euclid stacked image will be around 13Gb in size.

A discussion followed on the various types of cutout services that are provided, and could be provided in the future, by the archives and ESASky. For example, the Planck Legacy archive has a cutout service on their healpix data. ESASky may want to provide both types of cutout services: on healpix (HiPS) data and on the archival FITS images.

6. ESASky Statistics and updates

D. Baines presented the latest statistics for ESASky, to fulfil **Action 2018-06-12/02**, where the AAUG asked to see the number of users that access the archives via ESASky. The presentation also included a status update with the latest releases, the ESASky roadmap with future plans and a demo of the latest features and data in ESASky.

The AAUG asked if ESASky has a way of tracking scientific publications that use ESASky. Currently, ESASky can only do this if the authors have acknowledged ESASky in their paper. How to acknowledge ESASky is given in the ESASky credits web page and the team has begun to add this message at the end of each ESASky video. This is also a common problem seen with the archives, very few users acknowledge in their scientific publications the archives that were used. The AAUG suggested adding a popup or another window when users connect to ESASky, to make the users more aware of acknowledging ESASky in their scientific publications. Finally, the AAUG suggested that the ESASky team could give webinars to users since there are many different functionalities available in the tool and in the Jupyter widget pyESASky.

7. Date and time of next meeting

It was agreed that a doodle will be sent to the AAUG to fix the dates for the next meeting for some time in March or April 2021 (**Action 2020-10-02/07**).

8. AOB

There were no AOBs raised in this meeting.



RECOMMENDATIONS

The following recommendations were formulated by the AAUG:

User Survey

Recommendation 2020-10-02/01: The AAUG has examined the exhaustive report on the 2019 ESA Astronomy Archives User Survey Results prepared by D.Baines (ESDC). A very high satisfaction level of the users emerged from this report (across all missions only between 0-2% of users declared not to be satisfied with the archives) certifying that the attention given to the archive developments by the ESDC over the last years is valued by the community. The AAUG, therefore, thinks that the ESDC is on the right path to ensure the attractiveness and usefulness of the astronomy archives long into the future.

In earlier meetings this UG has urged the ESDC to enhance the usage of ESASky in the scientific community (see *Recommendation 2018-06-12/01* and *2018-06-12/04*). The high fraction of survey participants (~70%) that declare to know ESASky shows that the advertising by ESDC has been successful, and the AAUG invites ESA to take further steps to reach those parts of the astronomy community that are not yet aware of the existence of this powerful tool. It would be good to differentiate – as far as possible - the answers given in the user survey from professional astronomers and other users of the archives (school students, amateur astronomers and media/outreach staff make ~ 19%). The archives thus constitute an important channel to enhance the awareness and acceptance of ESA’s astronomy activities - and science in general - in our society. While the interest of the overall society in ESA’s astronomical activities is an asset for the agency, it is essential that the ESDC understands the needs of scientists towards which the technological developments should be primarily addressed. The AAUG, therefore, invites the ESDC to examine how the subgroup of scientists have rated the archives and ESASky, and which preferences they have expressed.

The AAUG asks the ESDC to be guided by the priority list expressed in the user survey in terms of new archive developments (see *Recommendation 2020-10-02/02* for more details).

Future Developments of the archives and ESASky

Recommendation 2020-10-02/02: An extensive list of new features have been implemented in ESASky. The AAUG is particularly pleased to see among those the direct access to many external (non-ESA) data archives that was suggested in *Recommendation 2018-06-12/01*, the multi-lingual aspect of ESASky, and the ongoing development of other user options requested by the AAUG, like the possibility of defining the sky portion through a cone search. All these new features make ESASky an indispensable tool for astronomical research.

A key point to keep in mind for the future is the overall architecture of the astronomy archives. One result of the user survey was that a vast majority of users are interested in having a single interface to all ESA astronomy archives. Discussions on the desirable level of homogenization of the astronomy archives have started in the AAUG and are expected to continue in future meetings of this group. As a specific example, “advanced web-based visualizers & cutout services” stand out as one of the top priorities in the wish list of



users. ESDC has already started developing the corresponding tools for individual archives (EUCLID), and the AAUG suggests that the ESDC applies this novel feature to all missions and to implement it in ESASky.

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

Recommendation 2020-10-02/03: Following discussions in previous AAUG meetings on the potentially high value of including user-supplied data into the ESA Astronomy Archives (see esp. *Recommendation 2018-06-12/01*), astronomy related EU FP7 and EU Horizon 2020 projects have been identified. The AAUG is currently reviewing these projects based on their websites with the goal to advise ESA on whether the data products (provided from these EU projects) are valuable enough for the astronomy community to justify that ESA takes steps to ensure their long-term preservation and accessibility to a wider community.

The preliminary study of the AAUG shows that there should be a better defined process for the evaluation and the respective responsibilities, and that the ESDC should involve the PI of the EU project from the earliest stage possible.

Related to this, the AAUG suggests that the ESDC also looks into different possibilities of storing the products from such projects (e.g. in a guest storage facility or directly in the archives), depending on the work-power required for their ingestion and maintenance.