DAXA: Traversing the X-ray desert by Democratising Archival X-ray Astronomy

David J. Turner[*1,2], **Jessica E. Pilling**[2], **Megan Donahue**[1], **Paul A. Giles**[2], **Kathy Romer**[2], **Agrim Gupta**[1], **and Toby Wallage**[2]

1 Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan, USA **2** Department of Physics and Astronomy, University of Sussex, Brighton, UK

DOI: DOIunavailable

Software

- Review
- Repository
- Archive

Editor: Pending Editor

Reviewers:

- @Pending Reviewers

Submitted: N/A
Published: N/A

Summary

We introduce a new, open-source, Python module for the acquisition and processing of archival data from multiple X-ray telescopes, Democratising Archival X-ray Astronomy (hereafter referred to as DAXA). The aim of DAXA is to provide a consistent, easy-to-use, Python interface to the disparate X-ray telescope data archives, and their processing packages. We provide this interface for the majority of X-ray telescopes launched within the last 30 years. This module will enable much greater access to X-ray data for non-specialists, while preserving low-level control of processing for X-ray experts. The package is useful for identifying relevant observations of a single object of interest, but it excels at creating and managing multi-mission datasets for serendipitous or targeted studies of large samples of X-ray emitting objects. Once relevant observations are identified, the raw data can be downloaded (and optionally processed) through DAXA, or pre-processed event lists, images, and exposure maps can be downloaded if they are available. As we may enter an 'X-ray desert', with no new X-ray missions coming online, during the next decade, archival data is going to take on an even greater importance than it already has, and enhanced access to those archives will be vital to the continuation of X-ray astronomy.

Statement of need

X-ray observations provide a powerful view of some of the most extreme processes in the Universe, and have had a profound impact on our understanding of many types of astrophysical objects; from in-solar-system objects, to supernovae, to galaxies and galaxy clusters. As such, access to X-ray data should be made as simple as possible, both for X-ray experts and non-specialists whose research benefits from a high-energy view; organisations such as the European Space Agency (ESA) and the High Energy Astrophysics Science Archive Research Center (HEASARC) have gone to great lengths to enable this access, and our software builds on their success. Through DAXA, most X-ray observatory archives are accessible through a single unified interface available in a programming language that is ubiquitous in astronomy (Python); locally searching for data relevant to a particular sample gives us the opportunity to better record and share the exact search parameters, through a Jupyter notebook for instance. X-ray data can also be particularly intimidating to those astronomers who have not used it before, which acts as a barrier to entry, limiting the reach and scientific impact of X-ray telescopes; it is in our interest to maximise the use of these data, both to support X-ray astronomy through the 'X-ray desert', and to persuade funding bodies of the great need for X-ray telescopes. DAXA is particularly powerful in this regard, as it provides a normalised, simple, interface to different backend software packages (some of which can be difficult for new users), allowing for the easy processing of X-ray data to a scientifically useful state; this is in addition to the ability to download pre-processed data from many of the data archives.

*turne540@msu.edu

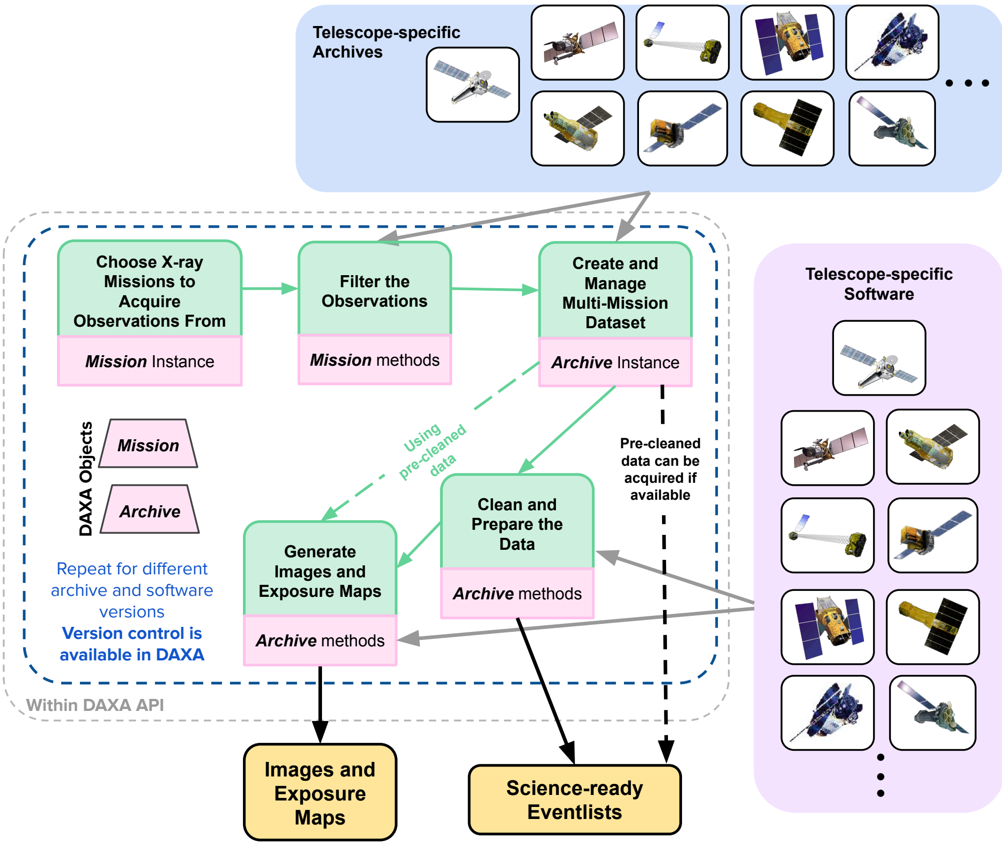

Figure 1: A flowchart giving a brief overview of the DAXA workflow.

Almost every sub-field of astronomy, astrophysics, and cosmology has benefited significantly from X-ray coverage over the last three decades; calibrating weak-lensing mis-centering for galaxy cluster studies (Zhang et al., 2019), even probing the irradiation of exoplanets (Poppenhaeger et al., 2021). The current workhorse X-ray observatories (*XMM*-Newton (Jansen et al., 2001) and *Chandra*; other telescopes are online but are more specialised) are ageing however, with *Chandra* in particular experiencing a decline in low-energy sensitivity that might limit science cases; these missions cannot last forever. If we are to enter an X-ray desert, where the astrophysics community has only limited access to new X-ray observations from specialised missions like *Swift* (Gehrels et al., 2004), *NuSTAR* (Harrison et al., 2013), and *XRISM* (XRISM Science Team, 2020), then archival data (and serendipitous studies) take on an even greater value than they already hold. DAXA is part of an ecosystem of open-source software designed around the concept of enabling serendipitous studies of X-ray emitting objects, and can download and prepare X-ray observations for use with tools like 'X-ray: Generate and Analyse' (XGA; D. J. Turner et al. (2022)). X-ray observations are uniquely well suited for the kind of archival study facilitated by DAXA and XGA, as they generally record the time, position, and energy of each individual photon impacting the detector (true for all missions currently implemented in DAXA); this means that we can create images, lightcurves, and spectra for any object detected within the field-of-view, even if it was not the target. With this software, we enable the maximum exploitation of existing X-ray archives, to traverse the X-ray desert and ensure that we are fully prepared for future X-ray telescopes such as *Athena* (Nandra et al., 2013) and *Lynx* (Gaskin et al., 2019). Having easy access to the whole history of X-ray observations of an object can provide extra context as to its astrophysics, and comes at no extra cost.

Finally, DAXA can be used to further one of the tenets of open-source science, reproducibility.

Its management features both allow the user to keep track of their dataset, but also to version control it. If more data become available, or existing data need to be reprocessed, then the version of the dataset can be automatically updated. Research publications can thus reference an exact version number of a dataset, which can be reproduced without offering the whole dataset for download.

Features

DAXA contains two types of Python class, mission classes and the archive class. Mission classes directly represent a telescope or survey (for instance there are separate classes for pointed and survey observations taken by *ROSAT* (Truemper, 1982), as the characteristics of the data are quite different), and exist to provide a Python interface with the current telescope observation database. Such mission classes allow the user to easily identify data relevant to their objects of interest with various filtering methods (it is also possible to download the entire archive of a telescope); these include filtering on spatial position (determining whether a coordinate of interest is within the field-of-view), filtering on the time of the observation (also filtering on whether a specific coordinate was observed at a specific time, for whole samples with different coordinates and times of interest), and filtering on specific observation identifiers (ObsIDs) if they are already known. Each mission class has some knowledge of the characteristics of the telescope it represents (such as the field-of-view) to make observation filtering easier. The user can also select only a subset of instruments, if the telescope has more than one, to exclude any that may not contribute to their analysis.

Once a set of relevant observations have been identified, for either a single mission or a set of missions, a DAXA data archive can be declared. This will automatically download the selected data from the various telescope archives, and proceeds to ingest and organise the data so that it can be managed (and if necessary, updated) through the DAXA interface. We have also implemented user-friendly, multi-threaded, data preparation and cleaning routines for some telescopes (*XMM* and *eROSITA* in particular, though this will expand); fine control of the parameters that control these processes is retained, but default behaviours can be used if the user is unfamiliar with the minutiae of X-ray data preparation. Another key benefit of reducing data with DAXA is the easy access to data logs through our interface, in case of suspected problems during the reduction processes. The module is also capable of safely handling processing failures, recording at which processing step the failure occurred for a particular ObsID.

All of this information is retained permanently, not just while the initial DAXA processes are running. Any DAXA archive can be loaded back in after the initial processing, once again providing access to the stored logs, and processing information. At this point the archives can also be updated, either by searching for new data from the existing missions, adding data from a different mission, or re-processing specific observations to achieve more scientifically useful data. Any such change will be recorded in the archive history, and processed observations version controlled, so that the data archive can have a specific version that refers to its exact state at any given time; this version can be referred to in published work using the data archive. Each data archive is also capable of creating a file that other DAXA users can import, and which will recreate the data archive by downloading the same data, and processing it in the same way; this renders making fully processed, and large, X-ray data files available with a piece of research unnecessary.

Existing software packages

There are no direct analogues to our module, though we must acknowledge the many pieces of software (and data archives), that greatly helped the development of DAXA. Data access is made possible primarily by the HEASARC data archive, though the Astroquery (Ginsburg et al., 2019) package is also used. HEASARC provides an online interface to query their data archive, which has similar functionality to some of the filtering methods of mission classes in DAXA (though we provide slightly more functionality in that regard), and they provide Python SQL examples to access the data, but none of the data management and cleaning functionality that we include.

Also worthy of mention are the various telescope-specific software packages that underpin DAXA's ability to perform data preparation and cleaning. Particularly important are the *XMM* Science Analysis System (SAS; Gabriel et al. (2004)) and the complementary extended SAS (eSAS; Snowden & Kuntz (2011)) packages, which allow us to provide simple Python interfaces to the complex, multi-step, processes that are required to prepare raw *XMM* data for scientific use. The analogous *eROSITA* Science Analysis Software System (eSASS; Predehl et al. (2021)) must also be mentioned, as it provides the tools needed to reduce and prepare *eROSITA* data. In this vein we must also acknowledge the HEASoft package, which is almost ubiquitous in X-ray data analyses, and is used by both SAS and eSASS.

Another related software package is the other module in our open-source X-ray astronomy ecosystem, X-ray: Generate and Analyse (XGA; D. J. Turner et al. (2022)). It has none of the same features as DAXA, as it exists to analyse large sets of X-ray data, we created DAXA to create and manage the kind of dataset required for XGA to attain maximum usefulness (though such datasets do not *have* to be analysed with XGA).

DAXA is greater than the sum of its parts, but is only possible because of the existing software packages it builds upon; we hope that it only enhances the value that astrophysicists derive from the other software we have mentioned.

Research projects using DAXA

DAXA has been useful for several research projects by different collaborations, and we anticipate that the number of research projects benefiting from its features will increase significantly. The primary use to which DAXA has been put is to assemble the multi-mission X-ray dataset (*XMM*, *Chandra*, *eROSITA*, *Swift*, and *ROSAT*) for the X-ray follow-up component of the Local Volume Complete Cluster Survey (LoVoCCS; Fu et al. (2022), David J. Turner, Donahue, Evrard, et al. (2024, in prep.)). It was used to identify the relevant observations, download, and process them, as well as to organise the significant number of files and make it easier for the dataset to be served to the X-ray community. Construction and administration of such large, complicated, multi-mission datasets is rendered quick and easy.

The X-ray Cluster Science (XCS; formerly known as the *XMM* Cluster Survey) collaboration now uses DAXA to create and manage their processed X-ray archive; particularly useful is DAXA's support for telescopes other than *XMM*, which has allowed the serendipitous science undertaken by XCS to expand to the use of different telescopes. These telescopes are complementary to *XMM*, and also increase the sky coverage, which in turn increases the likelihood that an object of interest has an accessible X-ray observation.

As DAXA now supports XCS, it has contributed to a research project that has measured X-ray properties (spectral, time-series, and photometric) for every LOFAR source that falls on an *XMM* observation (a DAXA generated dataset was used for an XGA analysis). This kind of bulk analysis is trivial when our software packages are utilised, and will result in a comprehensive catalogue that is invaluable to the radio astronomy community.

Finally, DAXA has been used to identify *XMM* and *Chandra* (alongside other telescopes, though they play only a supporting role) observations of a series of galaxy groups that appear in the foreground of UV bright quasars (David J. Turner, Donahue, & Stocke, 2024, in prep.). Absorption features that indicate the presence of Oxygen VI were identified in the spectra of several of the quasars, and the data that DAXA identified and retrieved allowed for an exploration of the hot-gas properties of these groups.

Future Work

The most significant new features implemented in DAXA will be new mission classes added when new X-ray telescope archives become available, or one of the existing missions that we have not yet implemented is added (for instance *XMM* observations taken whilst slewing). We will also seek to include support for more telescope-specific cleaning methods taken from their backend software; additionally we wish to implement our own generic processing and cleaning techniques where possible, applicable to multiple missions. We also aim to include source detection capabilities; specifically techniques that are generally applicable to multiple missions whilst taking into account instrument-specific effects.

Acknowledgements

DT and MD are grateful for support from the National Aeronautic and Space Administration Astrophysics Data Analysis Program (NASA80NSSC22K0476).

KR and PG acknowledge support from the UK Science and Technology Facilities Council via grants ST/T000473/1 and ST/X001040/1. JP acknowledges support from the UK Science and Technology Facilities Council via grants ST/X508822/1.

References

Fu, S., Dell'Antonio, I., Chary, R.-R., Clowe, D., Cooper, M. C., Donahue, M., Evrard, A., Lacy, M., Lauer, T., Liu, B., McCleary, J., Meneghetti, M., Miyatake, H., Montes, M., Natarajan, P., Ntampaka, M., Pierpaoli, E., Postman, M., Sohn, J., … Wilson, G. (2022). LoVoCCS. I. Survey Introduction, Data Processing Pipeline, and Early Science Results. *933*(1), 84. https://doi.org/10.3847/1538-4357/ac68e8

Gabriel, C., Denby, M., Fyfe, D. J., Hoar, J., Ibarra, A., Ojero, E., Osborne, J., Saxton, R. D., Lammers, U., & Vacanti, G. (2004). The XMM-Newton SAS - Distributed Development and Maintenance of a Large Science Analysis System: A Critical Analysis. In F. Ochsenbein, M. G. Allen, & D. Egret (Eds.), *Astronomical data analysis software and systems (ADASS) XIII* (Vol. 314, p. 759).

Gaskin, J. A., Swartz, D. A., Vikhlinin, A., Özel, F., Gelmis, K. E., Arenberg, J. W., Bandler, S. R., Bautz, M. W., Civitani, M. M., Dominguez, A., Eckart, M. E., Falcone, A. D., Figueroa-Feliciano, E., Freeman, M. D., Günther, H. M., Havey, K. A., Heilmann, R. K., Kilaru, K., Kraft, R. P., … Zuhone, J. A. (2019). Lynx X-Ray Observatory: an overview. *Journal of Astronomical Telescopes, Instruments, and Systems*, *5*, 021001. https://doi.org/10.1117/1.JATIS.5.2.021001

Gehrels, N., Chincarini, G., Giommi, P., Mason, K. O., Nousek, J. A., Wells, A. A., White, N. E., Barthelmy, S. D., Burrows, D. N., Cominsky, L. R., Hurley, K. C., Marshall, F. E., Mészáros, P., Roming, P. W. A., Angelini, L., Barbier, L. M., Belloni, T., Campana, S., Caraveo, P. A., … Zhang, W. W. (2004). The Swift Gamma-Ray Burst Mission. *611*(2), 1005–1020. https://doi.org/10.1086/422091

Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., Cowperthwaite, P. S., Craig, M. W., Deil, C., Guillochon, J., Guzman, G., Liedtke, S., Lian Lim, P., Lockhart, K. E., Mommert, M., Morris, B. M., Norman, H., Parikh, M., Persson, M. V., Robitaille, T. P., Segovia, J.-C., Singer, L. P., … a subset of the astropy collaboration. (2019). astroquery: An Astronomical Web-querying Package in Python. *157*, 98. https://doi.org/10.3847/1538-3881/aafc33

Harrison, F. A., Craig, W. W., Christensen, F. E., Hailey, C. J., Zhang, W. W., Boggs, S. E., Stern, D., Cook, W. R., Forster, K., Giommi, P., Grefenstette, B. W., Kim, Y., Kitaguchi, T., Koglin, J. E., Madsen, K. K., Mao, P. H., Miyasaka, H., Mori, K., Perri, M., … Urry,

C. M. (2013). The Nuclear Spectroscopic Telescope Array (NuSTAR) High-energy X-Ray Mission. *770*(2), 103. https://doi.org/10.1088/0004-637X/770/2/103

Jansen, F., Lumb, D., Altieri, B., Clavel, J., Ehle, M., Erd, C., Gabriel, C., Guainazzi, M., Gondoin, P., Much, R., Munoz, R., Santos, M., Schartel, N., Texier, D., & Vacanti, G. (2001). XMM-Newton observatory. I. The spacecraft and operations. *365*, L1–L6. https://doi.org/10.1051/0004-6361:20000036

Nandra, K., Barret, D., Barcons, X., Fabian, A., den Herder, J.-W., Piro, L., Watson, M., Adami, C., Aird, J., Afonso, J. M., Alexander, D., Argiroffi, C., Amati, L., Arnaud, M., Atteia, J.-L., Audard, M., Badenes, C., Ballet, J., Ballo, L., … Zhuravleva, I. (2013). The Hot and Energetic Universe: A White Paper presenting the science theme motivating the Athena+ mission. *arXiv e-Prints*, arXiv:1306.2307. http://arxiv.org/abs/1306.2307

Poppenhaeger, K., Ketzer, L., & Mallonn, M. (2021). X-ray irradiation and evaporation of the four young planets around V1298 Tau. *500*(4), 4560–4572. https://doi.org/10.1093/mnras/staa1462

Predehl, P., Andritschke, R., Arefiev, V., Babyshkin, V., Batanov, O., Becker, W., Böhringer, H., Bogomolov, A., Boller, T., Borm, K., Bornemann, W., Bräuninger, H., Brüggen, M., Brunner, H., Brusa, M., Bulbul, E., Buntov, M., Burwitz, V., Burkert, W., … Yaroshenko, V. (2021). The eROSITA X-ray telescope on SRG. *647*, A1. https://doi.org/10.1051/0004-6361/202039313

Snowden, S. L., & Kuntz, K. D. (2011). *Cookbook for analysis procedures for XMM-newton EPIC MOS observations of extended objects and the diffuse background*.

Truemper, J. (1982). The ROSAT mission. *Advances in Space Research*, *2*(4), 241–249. https://doi.org/10.1016/0273-1177(82)90070-9

Turner, David J., Donahue, M., Evrard, A., Fu, S., Dell'Antonio, I., Chary, R.-R., Clowe, D., Cooper, M. C., Lacy, M., Lauer, T., Liu, B., McCleary, J., Meneghetti, M., Miyatake, H., Montes, M., Natarajan, P., Ntampaka, M., Pierpaoli, E., Postman, M., … Wilson, G. (2024, in prep.). *X-LoVoCCS. I. Initial X-ray properties of LoVoCCS galaxy clusters*.

Turner, David J., Donahue, M., & Stocke, J. (2024, in prep.). *An X-ray view of galaxy groups in the line of sight of UV-bright quasars - comparing X-ray properties to absorption detection of warm-hot gas*.

Turner, D. J., Giles, P. A., Romer, A. K., & Korbina, V. (2022). XGA: A module for the large-scale scientific exploitation of archival X-ray astronomy data. *arXiv e-Prints*, arXiv:2202.01236. http://arxiv.org/abs/2202.01236

XRISM Science Team. (2020). Science with the X-ray Imaging and Spectroscopy Mission (XRISM). *arXiv e-Prints*, arXiv:2003.04962. http://arxiv.org/abs/2003.04962

Zhang, Y., Jeltema, T., Hollowood, D. L., Everett, S., Rozo, E., Farahi, A., Bermeo, A., Bhargava, S., Giles, P., Romer, A. K., Wilkinson, R., Rykoff, E. S., Mantz, A., Diehl, H. T., Evrard, A. E., Stern, C., Gruen, D., von der Linden, A., Splettstoesser, M., … DES Collaboration. (2019). Dark Energy Surveyed Year 1 results: calibration of cluster mis-centring in the redMaPPer catalogues. *487*(2), 2578–2593. https://doi.org/10.1093/mnras/stz1361