Data Access Policy in the Einstein Telescope Era

Contents

[Introduction to Data Access Policy 2](#_Toc203744301)

[Global Scenario 2](#_Toc203744302)

[LIGO-Virgo-KAGRA (LVK) Gravitational Wave Detector Network 3](#_Toc203744303)

[LISA (Laser Interferometer Space Antenna) 3](#_Toc203744304)

[ALMA - Atacama Large Millimiter Array 4](#_Toc203744305)

[Hubble Space Telescope 5](#_Toc203744306)

[James Webb Space Telescope 6](#_Toc203744307)

[ESA XMM-Newton 7](#_Toc203744308)

[NASA Chandra X-Ray Observatory 8](#_Toc203744309)

[The Vera Rubin Observatory (formerly LSST) 8](#_Toc203744310)

[The European Southern Observatory (ESO) Science Archive Facility (SAF) 10](#_Toc203744311)

[Comments on the Data Access Policy survey 11](#_Toc203744312)

[Definition of ET Data 11](#_Toc203744313)

[ET raw data 11](#_Toc203744314)

[ET preprocessed data: “h-reconstructed” 11](#_Toc203744315)

[ET alerts 11](#_Toc203744316)

[ET early warnings 11](#_Toc203744317)

[Strategies on the Data Access Policy in the Einstein Telescope Research Infrastructure 11](#_Toc203744318)

[Maximization of the science return with ET data. 12](#_Toc203744319)

[Data production sustainability 12](#_Toc203744320)

[Access to ET Data 12](#_Toc203744321)

[ET Raw Data 12](#_Toc203744322)

[ET “h-reconstructed” data access 12](#_Toc203744323)

[ET Alerts 12](#_Toc203744324)

[ET early warnings 12](#_Toc203744325)

# Introduction to Data Access Policy

Bla Bla

## Global Scenario

In this section we make a short and slightly arbitrary survey of the Data Access Policies adopted by different projects mainly in the astrophysical and astroparticle sector. The aim of this survey is to offer a larger view of the policies adopted by communities that are in the neighboring of the Gravitational Wave sector. This panorama had as initial seed the presentation[[1]](#footnote-1) and the discussion that the ET Collaboration had at the XIV ET Symposium (Maastricht, 2024)

|  |  |  |
| --- | --- | --- |
| **Project** | **Proprietary Period [months]** | **comment** |
| ALMA (Atacama Large Millimeter Array) | 12 |  |
| Hubble Space Telescope | From 0 to 6 | Depending on the proposal |
| James Webb Space Telescope | Usually 12 | PIs can consider to specify 6 months, 3 months or zero proprietary period at proposal submission time. |
| ESA Integral | 12 | A proposer can never ask for “data rights for the entire FOV” or “data rights for all sources (known/unknown) to be found in the FOV”. |
| ESA XMM-Newton | 12 |  |
| NASA Chandra X-Ray Observatory | 12 |  |
| RUBIN-LSST | 24 |  |
| Ice Cube | 24 - 36 |  |
| LISA | 6-12 | See presentation at the CE symposium |

### LIGO-Virgo-KAGRA (LVK) Gravitational Wave Detector Network

[note: AI Generated]

The LIGO-Virgo-KAGRA (LVK) Collaboration follows an open science policy, making its gravitational wave data publicly available after a proprietary period. This data, recorded by the detectors, is released through the Gravitational Wave Open Science Center (GWOSC). The proprietary period, typically 18 months after a data run, allows the LVK collaboration to conduct initial analyses before wider public access. This approach aligns with the FAIR Guiding Principles for scientific data management.

Here's a more detailed breakdown:

* Open Science Commitment: The LVK Collaboration is dedicated to open science principles, emphasizing data sharing and collaboration.
* Data Release: Data is released publicly through the GWOSC after a defined proprietary period.
* Proprietary Period: This period, usually 18 months after a data run, allows the LVK collaboration to perform their initial analyses.
* GWOSC Access: The Gravitational Wave Open Science Center (GWOSC[[2]](#footnote-2)) serves as the central platform for accessing and downloading the released data. GWOSC data are also used for dissemination and training, through the “GWOSC GW Open Data Workshop” series.
* FAIR Principles: The data release process adheres to the FAIR Guiding Principles, ensuring data is Findable, Accessible, Interoperable, and Reusable.
* Community Engagement: The public data release encourages broader scientific exploration and contributions to gravitational wave research.

### LISA (Laser Interferometer Space Antenna)

[note: AI Generated and the latest presentation of the LISA DAP has been made at the last CE symposium, but the slides aren’t available]

The LISA (Laser Interferometer Space Antenna) data access policy is still under development, but the general principle is to ensure broad and timely access to the data for the scientific community. This involves releasing lower-level data products (L1, L2, L3) with appropriate documentation and support, while also making algorithms, software, and models used in processing available. Discussions are ongoing about the specifics of data release, including timing, frequency, and the level of support provided.

Here's a more detailed breakdown:

* Public Release: The guiding principle is to enable the scientific community to reproduce analyses from at least L1 (TBC) to L3 products.
* Data Flow and Processing: The stages of data processing (L0 to L1 to L2 to L3) are subject to ongoing research and evolution as the understanding of the instrument and data improves.
* Software and Documentation: Along with the data, all algorithms (like TDI, searches), software, and models used in processing need to be made available.
* Protected Periods and Alerts: The policy will also need to address protected periods (e.g., around massive black hole merger events) and low-latency alerts for multi-messenger astronomy.

NASA's Role:

* NASA is studying how US scientists can participate in the science ground segment and access LISA data through public release. [[3]](#footnote-3)

Community Engagement:

* The policy aims to facilitate broad participation and diverse thinking within the scientific community.

In essence, LISA's data access policy is designed to be open and inclusive, promoting scientific discovery while ensuring the integrity and usability of the data.

### ALMA - Atacama Large Millimiter Array

[note: AI Generated]

The Atacama Large Millimeter/submillimeter Array (ALMA)[[4]](#footnote-4) is an astronomical interferometer of 66 radio telescopes in the Atacama Desert of northern Chile, which observe electromagnetic radiation at millimeter and submillimeter wavelengths.

ALMA data access is governed by policies that balance open access with the need to protect the rights of proposers and observers. In general, data becomes public after a proprietary period, typically 12 months, but exceptions can be made.

Key aspects of ALMA data access policies:

* Proprietary Period: Data is typically proprietary for 12 months after the completion of Quality Assurance (QA2) for each observing cycle, meaning only the Principal Investigator (PI) and their team can access it during this time.
* Public Data: After the proprietary period, data becomes publicly available through the ALMA archive.
* Proposal Information: Information about accepted proposals (title, abstract, PI/co-Is) becomes public soon after the proposal review process, but scientific and technical justifications, reviews, and figures remain confidential.
* Calibration Data: Calibration data are generally made public immediately after observation.
* Registered Users: Anyone can register for an ALMA user account to access data and other resources.
* Exceptions: Extensions to the proprietary period or restrictions on data access can be granted in special cases, requiring justification and approval.
* Data Delivery: PIs receive data delivery emails with information about downloading data, the proprietary period, available support, and publication requirements.
* Data Usage: ALMA data is copyrighted by ESO and distributed under the Creative Commons Attribution 4.0 International license (CC BY 4.0).

In summary, while ALMA data has a proprietary period to protect proposers, it eventually becomes publicly available through the archive, fostering open science and collaboration.

### Hubble Space Telescope

[note: AI Generated]

Hubble Space Telescope[[5]](#footnote-5) data is generally public, but proposers of observations have a period of exclusive access to their data before it becomes publicly available. This exclusive access period is typically 6 months for Small and Medium General Observer (GO) proposals. For Treasury and Large Programs, data is typically made public immediately or after a short proprietary period. Users can access the data through the Mikulski Archive for Space Telescopes (MAST).

Key Aspects of Hubble Data Access:

* Exclusive Access Period: Proposers of Small and Medium GO programs have a 6-month exclusive access period. This means they are the only ones who can access the data during that time.
* Public Access: After the exclusive access period, the data becomes publicly available to anyone through the MAST archive.
* Data Types: Both raw and processed data are archived. Enhanced products for non-exclusive access observations may also be available through the Hubble Legacy Archive (HLA).
* Accessing Data: Data can be retrieved from the MAST archive via the MAST Portal or through programmatic access via HTTP requests.
* Special Cases: Exceptions to the standard access rules can be requested, such as extending the exclusive access period or restricting access to header information.
* Data Rights: Specific data rights and duplications are detailed in the HST Call for Proposals. For example, proposers should justify any duplications of previous observations when proposing new ones.

In essence, the Hubble Space Telescope data access policy balances the need for proposers to analyze their data and make scientific discoveries with the goal of making the data publicly available for further research and exploration.

### James Webb Space Telescope

[note: AI Generated]

James Webb Space Telescope data is generally released publicly after an exclusive access period (EAP), usually 12 months after data is archived. During the EAP, the principal investigators and their teams have exclusive rights to analyze the data. After the EAP, the data becomes freely available to the public through the Mikulski Archive for Space Telescopes (MAST).

Data Access Details:

* Exclusive Access Period (EAP): Investigators who propose and are awarded time on JWST have an exclusive access period for their data, typically 12 months.
* Public Release: After the EAP, the data becomes publicly available through the MAST archive.
* Authentication: To access data, including during the EAP, users need to authenticate their identity with a STScI MyST account.
* MAST Portal: The MAST archive is the central location for accessing JWST data.
* API Access: Researchers can also access data through the MAST API, often using an API token for authenticated access to data within the EAP.
* Proposal Categories: Different types of proposals (e.g., Small, Medium, Survey, Treasury, Calibration, Large Programs) may have different EAP policies, with some having no EAP by default.
* Data Rights: The EAP is designed to allow the scientists who proposed and executed the observations to analyze their data and publish their findings before the general public gains access.

In essence, the policy balances the need for scientific discovery with the principle of open data access. The EAP allows the scientists who invested significant time and effort in obtaining the data to analyze it first, while also ensuring that the data eventually becomes available to the broader scientific community and the public.

### ESA XMM-Newton

[note: AI Generated]

ESA's XMM-Newton data access policy is designed to promote open science. Data from XMM-Newton observations are generally proprietary for a period of time (typically 12 months) after the observation, but become publicly available after that period. ESA retains the right to use the data for instrument evaluation, calibration, and public relations purposes, with proper acknowledgment of the Principal Investigators (PIs).

Here's a more detailed breakdown:

Data Access:

* Proprietary Period: XMM-Newton data is initially proprietary, meaning it's not immediately available to the public.
* Public Release: After a set proprietary period, the data becomes publicly available through the XMM-Newton Science Archive (XSA).
* Targets of Opportunity (ToOs): ToO data often has a shorter proprietary period, typically six months, but this can be linked to other observations or determined by the Project Scientist.
* Discretionary Time (DPS): Data obtained in Discretionary Time is also subject to the standard proprietary period and public release through XSA.

ESA's Data Rights:

* Instrument Evaluation and Calibration: ESA has the right to use the data for instrument evaluation and calibration purposes.
* Public Relations: ESA can use the data for public relations activities, with proper acknowledgment given to the PIs.
* Scientific Confidentiality: ESA maintains scientific confidentiality during the proprietary period.

Publication Guidelines:

* Acknowledgement: Any publication using XMM-Newton data must acknowledge the observation with the target name, date, and observation ID.
* Footnote: A footnote on the first page of the publication should acknowledge that the research is based on XMM-Newton observations.
* Catalogue Usage: Publications using XMM-Newton catalogues should also include the catalogue version and the specific source names.

Data Access and Tools:

* XMM-Newton Science Archive (XSA): The primary interface for accessing XMM-Newton data and catalogues.
* Astroquery and TAP: Tools for accessing XMM-Newton data and catalogues programmatically.
* Observation Data Files (ODF): Raw, uncalibrated data files.
* Pipeline Processing System (PPS) Products: Calibrated, scientifically usable data products.

In summary, ESA's XMM-Newton data access policy balances the need for proprietary research with the broader goals of open science. Data is made publicly available after a proprietary period, and ESA retains specific rights for instrument evaluation, calibration, and public relations, while emphasizing the importance of proper acknowledgement of the source data in any publications.

### NASA Chandra X-Ray Observatory

[note: AI Generated]

NASA's Chandra X-ray Observatory data access policy dictates that data is publicly available in an archive one year after the observation, with some exceptions for data obtained from sources with more restrictive policies. This policy ensures broad access to valuable scientific data for researchers and the public, promoting open science principles.

Here's a more detailed breakdown:

* Public Archive: Chandra data, including processed observations, is stored in a public archive accessible to anyone after a one-year proprietary period.
* NASA Policy: The policy emphasizes no exclusive access to data by individuals or science teams, allowing for a brief period for calibration and quality assessment before public release.
* Exceptions: Some exceptions may exist for data obtained under more restrictive policies or from inherently sensitive sources.
* Purpose: The policy aims to facilitate scientific research and public engagement with Chandra's findings, promoting transparency and collaboration.
* Data Access Tools: The Chandra X-ray Center (CXC) provides tools like ChaSeR, find\_chandra\_obsid, and download\_chandra\_obsid to help users locate, access, and download publicly available data.
* Open Source Software: NASA also encourages the use of open-source software for analyzing Chandra data, further enhancing accessibility.
* Community Engagement: Chandra's data is used by professional astronomers, amateur astronomers, and the general public for various purposes, including research, education, and public outreach.

### The Vera Rubin Observatory (formerly LSST)

[note: AI Generated]

The Rubin Observatory's data access policy[[6]](#footnote-6) dictates that Rubin data rights holders have access to the Rubin Science Platform (RSP) and its data, including proprietary data during the initial two-year period after release. Public data, after this proprietary period, will be available to everyone. Rubin data rights are automatically granted to scientists and students affiliated with institutions in the US and Chile, as well as to international contributors who are part of designated in-kind contribution programs.

Here's a more detailed breakdown:

* Rubin Data Rights: These rights are essential for accessing and analyzing the data within the RSP. They encompass the ability to access, analyze, and publish results based on proprietary data products and services.
* Rubin Science Platform (RSP): This is the primary platform for accessing and analyzing Rubin data. It provides access to both public and proprietary data, as well as computational resources for analysis.
* Proprietary Period: The initial two-year period after data release is considered proprietary. During this time, only individuals with Rubin data rights can access the data and use the RSP.
* Public Data: After the proprietary period, the data becomes public and accessible to anyone.

Data Rights Holders:

* Scientists and students affiliated with institutions in the US and Chile are granted Rubin data rights.
* Individuals on international teams who have made in-kind contributions to the Rubin Observatory also receive data rights.
* The specific individuals with data rights from in-kind contributions are listed as part of the data rights agreements with the Rubin Observatory.

International Contributions:

* International programs can earn data rights for their members through in-kind contributions to the Rubin Observatory.

Data Access for Everyone:

* After the proprietary period, anyone can access the public data products.

Data Access Methods:

* The RSP is the main way to access and analyze Rubin data, but alert brokers and independent data access centers also provide access methods

### The European Southern Observatory (ESO) Science Archive Facility (SAF)

The data access policy of the ESO Science Archive Facility is described in this public web page: <https://archive.eso.org/cms/eso-data-access-policy.html>. To allow an easy reading of this document, the first part of that web page is copied hereafter:

*The European Southern Observatory (ESO) Science Archive Facility (SAF) has been operational at ESO headquarters in Munich since 1991. It contains data from ESO telescopes at La Silla Paranal Observatory, including the APEX submillimeter telescope on Llano de Chajnantor. All raw data from the La Silla Paranal Observatory are stored together with the corresponding calibrations, as well as selected products both contributed by the community or generated at ESO. In addition, it contains the UKIDSS/WFCAM data obtained at the UK Infrared Telescope facility in Hawaii, as well as the science processed data from the ESO large programmes.*

*The Principal Investigators (PIs) of successful proposals for time on ESO telescopes, and their delegates, have exclusive access to their scientific data for the duration of a proprietary period, after which it is accessible to all users from the international community. Processed data distributed via the SAF retain the same proprietary protections as the raw data they were derived from. The length of the proprietary period is set by the Director General and communicated in the Phase 1 Call for Proposals. The proprietary period starts when the raw data is made available for download from the ESO SAF to the PI, or their delegates. It is typically of one year and may depend on the observing programme type (e.g. Public Survey raw data is not covered by proprietary protection).*

*The information contained in the headers of all the data files in the ESO SAF is in normal circumstances immediately and publicly available. The abstracts of successful proposals are made public at the time the proprietary period expires. All calibration data are public immediately after the observations.*

*Exceptions to the above rules, such as the extension of the proprietary period or restrictions to the access to header information, may be granted in special cases. The corresponding requests have to be justified in detail and submitted to the Director General of ESO for approval. Deviations from the general policies may also occur for Guaranteed Time Observations (GTO). Please check the dedicated information at the GTO policy page.*

*All data in the ESO archive retain ESO’s copyright, and are distributed according to the terms of the Creative Commons Attribution 4.0 International license (CC BY 4.0).*

*ESO subscribes to the FAIR Guiding Principles for scientific data management and stewardship (see https://www.go-fair.org/fair-principles).*

*When downloading data from the SAF, all users are requested to limit to maximum five concurrent downloads.*

***Requirements for third parties distributing ESO data***

*ESO data files (both raw and processed data) may be distributed by third parties, and disseminated via other services, according to the terms of the Creative Commons Attribution 4.0 International license. Credit to the ESO provenance of the data must be acknowledged, and the file headers preserved.*

## Comments on the Data Access Policy survey

In the previous section, the Data Access Policies of a set of physics and astrophysics large scientific projects have been summarized. Obviously the list of projects analyzed is limited and probably arbitrary; surely have been neglected relevant projects having completely different approaches can be found, like, for example, the Fermi Gamma-ray Space Telescope, that after a first period of about one year after the launch, is releasing immediately its data, or the expected High Luminosity LHC ESFRI project where the DAP will implement a more restricted data access. In any case, from the set of described project it is possible to extract a similar approach in the DAP: the produced data are openly accessible after a proprietary period that ranges from few months to two years.

# Definition of ET Data

## ET raw data

## ET preprocessed data: “h-reconstructed”

## ET alerts

## ET early warnings

# Strategies on the Data Access Policy in the Einstein Telescope Research Infrastructure

The Einstein Telescope Data Access policy must be elaborated keeping in mind several objectives and constraints.

## Maximization of the science return with ET data.

This requires:

* Open Access to ET data
  + Bla bla
* Data reliability and stability
  + Bla bla
* Data preprocessing and reduction
  + Bla bla
* “FAIRness” of the ET data
  + Bla bla

## Data production sustainability

Instrument scientist “preservation »

## Access to ET Data

### ET Raw Data

Open Access description and infrastructures

### ET “h-reconstructed” data access

Proprietary period

Open access description and infrastructures

### ET Alerts

Open Access description and infrastructures

### ET early warnings

Open Access description and infrastructures

1. <https://indico.ego-gw.it/event/710/contributions/6375/attachments/3578/6405/Nando_Patat_EinsteinTelescopeMaastrict.pdf> [↑](#footnote-ref-1)
2. <https://gwosc.org/> [↑](#footnote-ref-2)
3. Obviously the impact of the financial cuts of the current American Administration on the NASA’s budget will impact on these aspects. [↑](#footnote-ref-3)
4. <https://www.almaobservatory.org/en/home/> [↑](#footnote-ref-4)
5. <https://science.nasa.gov/mission/hubble/> [↑](#footnote-ref-5)
6. <https://rubinobservatory.org/for-scientists/data-products/data-policy> [↑](#footnote-ref-6)