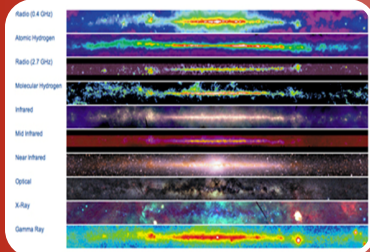


ESA Archives Long-Term Strategy

Guido De Marchi

12.6.18

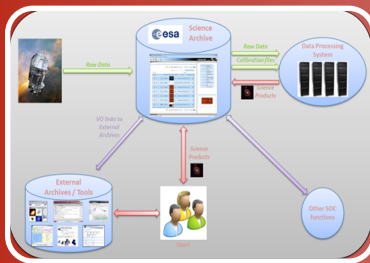
ESAC Science Archives Strategy



Enable maximum ***science exploitation*** of data sets



Ensure efficient ***long-term preservation*** of data, software, knowledge, using modern technology



Build cost-effective archives, also through ***integration in, and across, projects***

Brainstorming on strategy document

- First long-term strategy discussed and defined in 2013, time to update it
- Internal ESA forums agree vision with 3 pillars is valid
 - We want to enable best science
 - We want to preserve data, software, knowledge
 - We want efficient preservation, hence integrate archives with operations (when possible)
- Need to revise the way these topics are addressed in the 2013 strategy document, too much emphasis on IT, VO and multi-mission tools, not enough on legacy and community

1. Enable maximum science return

- Ensure data quality and completeness
 - Ensure complete, calibrated data; also ancillary data?
 - High-level science products from community (H2020)
 - Link to related ground based data
- Improve data access and ease data search
- Widen ESDC users base

1. Enable maximum science return

- Ensure data quality and completeness
- Improve data access and ease data search
 - Enable fast, reliable access to all data (e.g. CDN like ESASky)
 - Develop Science Exploitation and Preservation Platform with strong community interaction
 - Develop collaborative research areas (jupyter notebook)
 - Search/mining tools based on science values/themes, across space, time, energy
- Widen ESDC users base

1. Enable maximum science return

- Ensure data quality and completeness
- Improve data access and ease data search
- Widen ESDC users base
 - Engage young generations through research proposals
 - Support conferences, summer schools, town halls
 - Encourage and support citizen science projects
 - Community updates through ESDC Newsletter
 - Feedback through surveys, promote helpdesks, webinars
 - Monitor exploitation metrics

2. Preserve data and knowledge for the long term

- Maintain data ingestion capabilities in legacy phase
- Maintain science support to users of legacy archives
- Search and include missing legacy data
- Make legacy data part of multi-mission tools

This means that ESDC will have to

- Work with mission teams and community to prepare legacy transition
- Agree with mission teams on what to keep/ignore
- Prepare a legacy manual (also for “off-site” archives)

3. Encourage archives integration within/across projects

- Goal is to avoid duplications and ease transition to legacy, but recall that “one size does hardly fit all”
- Work early on with SOCs to plan mission archives
- Agree early on with consortia on ESA core areas
- Agree early on with PIs on calibrated data
- Define minimum set of metadata across archives?
- Regularly review and assess technical framework
- Foster dialogue across archives on functionalities
- Interoperability: standards take time, new features?

Internal call for priorities

- Many ideas collected internally (scientists and software engineers)
- Some are ready for implementation
- Ideas ranked internally, two groups:
 - High impact, low cost, mostly organisational work
 - High impact, substantial development work
- Inputs and comments by UGs sought

Group 1

High impact, low cost, mostly organisation needed (no development)

Actively enable data exploitation [first choice, 31 points]

- [Research proposals to fund young scientists (need reviewers)]
- [Visitor programme linked to data exploitation]
- [Citizen science projects (needs tools for analysis)]

Advertisement [third choice, 19 points]

- [Support/organise summer schools (also at ESAC)]
- [Newsletters, town hall meetings]

Assist and feedback [second choice, 20 points]

- [Improved helpdesks, with videos, demos, chats]
- [User surveys (feedback on what does/does not work)]

Archival Research Programme

1. Goal is to support early-career scientists present their science from archival data (ok if combined with other data)
2. Postdoc/student writes short application (2 pages) outlining science idea
3. Committee selects x most promising ideas, ESA allocates funds (travel, hotel, conference registration; max 800 euro?)
4. When paper is ready (draft? submitted? accepted?) successful proposer contacts ESA for travel authorisation
5. Proposer asked to mention Archival Research Programme in her/his presentation (one slide)
6. After conference, proposer sends tickets/receipts to ESA for reimbursement
7. First year will accept proposals for recently published papers if based on archival data

Group 2

Considerable development required, higher costs

Collaborative areas (proof of concept) [second choice, 19 points]

- Offer collaborative research areas, JupyterLab example
- Code-to-data, move storage & tools to “cloud”
- Science Exploitation & Preservation Platform

Analysis tools integrated in archives [first choice, 27 points]

- Data mining and analysis tools as part of the archives
- Cross-mission data selection in space, energy, time
- Archive interfaces based on science themes
- Integrate metadata, with source details/links and quality flags; encourage/allow users to do that too?

Other ideas [third choice, 13 points]

- DOI to improve linking to papers, identify unused data
- Access to some housekeeping data, priority for some missions
- Define metrics and monitoring (for R&D products)

Your opinion

1. Are we missing something important?
Any comments or ideas?
2. Are archival research proposals useful?
If so, would you help or suggest names for reviewing the applications?
3. Do you see a science use for the SEPP?
Or should the European Open Science Could be used instead? (what do you/your colleagues say about that?)