Space Science Data Handling at the ESAC Science Data Centre

Dr. Bruno Merín (ESA, Head of the ESAC Science Data Centre, SCI-SAA)
ESA_Lab “Space Data Management Workshop”, Venice, 8th of May of 2023
bruno.merin@esa.int
Outline

• Space science data in ESA’s Scientific missions
  • (science-driven) data management
  • International standardisation
• AI in space science data management
  • Citizen science projects
  • Virtual assistants
• The future of space science
Space science data management in Europe today

ESA Coordination with Science Provider and User Communities

Science Operations Systems Studies, Design 
& Requirements

Science Operations Planning

Science Instrument Handling

Science Data Processing

Science Data Archiving

Usually more by ESA

Usually more by Member States

Usually more by ESA

- THE EUROPEAN SPACE AGENCY
Example of data flow for a space science mission
Example data flow for a Planetary mission: JUICE
Example data flow for an Astronomy survey: PLATO

PLATO Data Center
- data processing algorithms
- development support

Input Catalogue generation

L2/L3 processing

follow-up and preparatory database management

PLATO Science Management
- Target/Field Characterisation and Selection
- Follow-Up
- Exoplanet Science
- Stellar Science
- Complementary Science

Payload Calibration/Operation Team (includes members from PDC, PSM, Payload & PPT)
Example data flow for a collaborative mission: SMILE
Example data flow for a Astronomy mission: Euclid
Science is about revisiting the data with new knowledge
Complex data products grow in size and complexity
.. And they are getting very large very soon..
Astronomy is becoming a data-driven field

http://archive.stsci.edu/hst/bibliography/pubstat.html
Hosting science data for decades

24 missions developed, operated and curated for over four decades
International collaboration in standardisation

International Planetary Data Alliance

International Heliophysics Data Environment Alliance

International Virtual Observatory Alliance

ESAC Science Data Centre

Astronomy, Heliophysics and Planetary Archive Users Groups
The traditional scientific method

1. Make hypothesis
   - Motivation
   - Visualization
   - Creativity
   - Knowledge

2. Obtain data to test hypothesis
   - Simple / easy access to reliable and relevant data

3. Analyze data to create new theory
   - Fast computation on new data
   - Easy comparison between data and models/theory

4. Revise existing data with new theory
   - Data completeness and consistency

5. Publish result
   - Reproducible data representation

6. Fame & Glory
ML Lifecycle

The future scientific method
How does IA fit in our datacentre?

- Deep learning classifiers of image contents for data-driven searches or quality control: images with asteroid trails or cosmic rays.
- NLP to automate linking of data to scientific papers. Automatic generation of statistics on Papers on our missions.
- Deep learning classifiers of image contents for data-driven searches: e.g. search images with Interstellar medium bubbles.

**ESAC Science Data Centre Value Chain**

- Data production
- Curation
- Archiving
- Valorisation
- Dissemination
- Support to users in their exploitation

- Noise reduction in astronomical images with neural networks, gaining a 50% in S/N.
- Virtual Assistants to support users on our Web Portals (e.g. Cosmos, ESASky).
AI / ML to enhance science data

Data reception → Data indexing for DB → Data ingestion

Curation → Archiving → Valorisation → Dissemination
AI / ML to enhance science data

Data reception → Data indexing for DB → ML classification of new dataset → Data ingestion

Action → Curation → Archiving → Valorisation
We use citizen science to generate labels
More than 11,000 volunteers classified 19 years of HST images finding a few thousand new trails of asteroids in just a few months!
Trained Google Auto-ML with the labels from volunteers.

<table>
<thead>
<tr>
<th>Images</th>
<th>Train</th>
<th>Evaluate</th>
<th>Test &amp; Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>All images</td>
<td>4,556</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Labelled</td>
<td>4,537</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unlabelled</td>
<td>19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Filter labels**
  - asteroid | 1,302
  - cosmic_ray | 1,331
  - gravitational_lens_arc | 492
  - satellite | 1,613

**ADD NEW LABEL**

- asteroid(2)
- satellite(1)
- asteroid(1), cosmic_ray(1)
- asteroid(1)
Classifying the eHST archive

- 40,151 “composite” HST ACS/WFC and WFC3/UVIS images (x4 cutouts = 160k cutouts)
- 122 computer Node Hours (~ €350)
- Run on multiple nodes for ~3 hours
- Batch classification on Google Cloud: ~7 hours (38 Node Hours, €60)
How will the future look like?

Chris Beaumont - Hackable User Interfaces and The Future of Data Analysis in Astronomy https://www.youtube.com/watch?v=SES7DMqw9Lc
The most advanced ultimate data query system should enable a dialogue with the user, like in the movie “her” (2013).
ESASky already features a virtual assistant!

http://sky.esa.int
Our powerhorse enabling all this: ESA Datalabs

"You can either move your questions or the data. [...] Often it turns out to be more efficient to move the questions than to move the data."

Jim Gray, eScience: A Transformed Scientific Method

BRING YOUR QUESTIONS TO THE DATA

There is a new paradigm, opening completely new opportunities for discovery - a data-intensive approach to science. In many domains, we have entered what could be called the golden age of surveys, with several large-scale projects, spanning decades, between finished, ongoing, and planned activities. ESA is responsible, or is a major partner, in several of these initiatives.

There is, however, a new profound change: data has become a major technological challenge. Increases by multiple orders of magnitude in dataset size means that transferring data to a scientist is often unfeasible.

ESA datalabs gives you a privileged position; bring your code directly to ESA's infrastructure - there is a great set of tools and programming languages are flexible - and execute it with direct access to ESA's archives.
Thanks for your attention!

http://archives.esac.esa.int
http://sky.esa.int
http://datalabs.esa.int
bruno.merin@esa.int