Cross-mission and archival papers

Guido De Marchi ESDC Head Scientist

28/10/2021

Cross-mission and archival papers

- Archives make it easy to find the data for individual missions and across areas
- Statistics on data access on ESDC webpage



How are data used?
 To write papers!



Publications over last 5 years









Publications over last 5 years



Publications in last 5 years



Citations to papers of last 5 years



In summary

- About 10% of papers use data from more than one ESA mission: Hundreds of papers per year!
- We invest about 10% of our time/resources in crossmission archive activities
 Great match with paper outputs!
- Why include multi-mission data in papers?
 Increased impact, more citations!

Publication statistics

- Science Programme Committee asked for statistics on Archival papers
- What exactly are they? Definition depends on community
- Why is the SPC asking about them? No reasons given...

Finding a definition

- Must apply in similar way to all missions, in all areas of space science
- Keep in mind that only some of ESA missions are observatories, most are PI-lead missions
- Must be based on information available for all missions
- Application as automated as possible (no time to read all papers)

In practice

- Fully archival: the list of authors does not include the Principal Investigator of the proposal, instrument, or experiment producing the data used in the papers
- Non archival: the list of authors includes one or more PIs
- Partly archival: data come from different proposals, instruments, or experiments and at at least one of the PIs is in the list of authors

To PI or not to PI?

- Some NASA missions (HST) also take Cols into account. Others (Chandra) only consider PIs
- Tough to track Cols for Solar System missions, PI teams change over time, members not always known
- But considering Cols might not be needed (see later...)

Observatory-type missions



For these we know which observation from which proposal goes into which paper

Survey-type/PI-lead missions



For these, Natural Language Processing will help us identify which data went into which paper, but for now statistics are based on matches between names of authors and of PIs

Survey-type/PI-lead missions



For these, Natural Language Processing will help us identify which data went into which paper, but for now statistics are based on matches between names of authors and of PIs

Observatory-type: Pls vs. Pls & Cols



The differences are truly marginal: +3% non-archival, – 5% fully-archival. Results are reliable even when we can only base statistics on Pis.

Observatory-type: HST vs. XMM



For XMM, already 8 years after launch fully-archival papers were more than nonarchival. For HST this only happened 26 years into operations...