Digital Object Identifiers and the ESA Science Archives

Guido De Marchi

Head Scientist

ESDC

European Space Agency

Better linking data and papers

- At ESA we track monthly number of papers based on ESA missions and the volume of data downloaded from archives

- Directly linking papers to data is difficult, no standards exist, some missions do it manually (e.g. XMM, Herschel), very time consuming

- Monitoring which data are used more/less tells us if calibration is not sufficient or data analysis tools are missing

- Information might also help ESA advisory bodies make informed decisions about payload on future missions (extension)

European Space Agency

Two avenues

- Artificial Intelligence: natural language processing and machine learning to extract the syntactic dependencies between instruments and measurements in papers text and linking them to the ESA databases
  - Just started exploratory study with National Observatory of Athens on analysis of Mars Express, Planck, Cluster papers
  - Longer term, we will keep you updated

- Digital Object Identifiers: assign unique identifier to (collections of) data to help authors/readers easily link them to the archives
  - Short term, happening now

DOIs help data to be FAIR

- Science community at large promotes open science

- Easier if data are FAIR:
  Findable, Accessible, Interoperable, Reusable

- Funding agencies encourage this, including EU

- Soon persistent identifiers might become mandatory

- Good time to adopt

European Space Agency

Some journals are already getting strict…

- American Geophysical Union journals (e.g. JGR) reject papers if data are not available to referee

- Since 1 August JGR and A&A accepted papers must make data available to all via persistent repositories (not just institute's websites)

- With DOI standard, authors make their results more transparent, verifiable, reproducible, reusable, interconnected → **visibility**

- Helps advancement of science

DOIs at ESDC

- We want our data to be easy to find and cite

- Goals: help scientists track and reuse data
    - simplify linking papers to data
    - improve visibility of our data

- ESA has DOI capability, through agreement with CrossRef:
prefix **10.5270** = **ESA**

- Already used for publications and Earth Observation data, now also starting with Space Science data (Herschel, Mars Express, Planck)

European Space Agency

Keep it simple

- Recall goal: make ESA data easy to find, reuse, and cite

- Need uniform solution across all archives, yet flexible

- Must apply to "collections" defined both by missions and by users

- Define collections at right level to keep system lean (granularity)

- Avoid versioning whenever possible (DOIs refer to current data)

- DOIs get random names, "ESA–h2o7nvx", to avoid second-guessing

Collection level – Missions

- Astronomical observatories: one DOI per proposal, ~20,000

- Other astronomy missions: DOI at level of catalogue (e.g. Gaia DR) or map (e.g. Planck), ~100

- Planetary science missions: one DOI per dataset/bundle (e.g. one per instrument/experiment/phase), ~500

- Heliophysics missions: one DOI per set of data (could be one per experiment/instrument/[sensor]), ~100

- Less than 1,000 new DOIs per year (manageable costs)

European Space Agency

Collection level – Example for Heliophysics

- SOHO: per sensor plus one for experiment (CELIAS)

- CLUSTER: per combination of experiment/detector (for CIS)

- Double Star: per experiment

- Ulysses: per experiment (or per detector when needed)

- Solar Orbiter: per instrument (or per detector when needed)

- Proba 2: per experiment

Collection level – Users

- We want DOIs to help scientists track and reuse ESA data discussed in papers

- Let scientists assign DOIs to specific collections of data they used in their papers

- Paper defines data used, scientists select them from archive, put them in a "basket" and mint DOI on the fly (already offered by STScI)

- Also community-generated Higher Level Science Products delivered to ESA will receive DOIs in the same way

Coordinate with journals

- Once DOI assignment (automatic and on-the-fly) is ready and tested, we will contact editorial offices

- Starting with European journals

- We expect journals to welcome the initiative: encourage persistent links (and discourage institutes' websites)

- Ask for feedback and to advertise service in instruction to authors

- Any comments or suggestions by **IHDEA** members very welcome!

Backup slides

Next steps   ???REMOVE SLIDE???

- Missions (PS/AS/MM) to discuss/decide granularity level and collect metadata for landing pages

- Staggered approach, starting with Herschel and Planck as trailblazers

- ESDC to progressively automate DOI and landing page generation for each mission/archive

- Currently DOIs for ESA assigned manually by EOP at ESRIN; following our request EOP will automate the process (we will need thousands)

- ESDC to later develop on-the-fly DOI minting tool for archive users

Landing page template

- Automatically-generated landing pages will populate templates with metadata in archives

- ESDC will provide standard template, missions welcome to modify it

- For now, landing pages hosted on cosmos.esa.int

- Landing pages to be edited and kept up to date
  (AS to coordinate)

European Space Agency

Some picture/logo

DOI number

Metadata

Description of the data

To facilitate future maintenance:
- Better stick to static text/images (extracted from headers?)
- Limit number of hardcoded URLs
- If a paper exists, link to its DOI

Guidelines for
landing page template

DOI metadata
- DOI number
- Short description
- Mission/Instrument
- Producer (ESA/partners)
- Format
- (Version)
- DOI creation date
- Link to the data

Examples from Sebastien/Pedro/Marcos

Discussion

DOI systems

- Managed system* for persistent identification of content on digital networks

- Provides unique identification, persistence, resolution, metadata

- Information about object can change over time, including where to find it, but its DOI name will not change

*DOI Foundation (doi.org) is a federation of registration agencies
  (e.g. datacite.org, crossref.org, etc)

European Space Agency

Landing pages

- DOIs are just links to "landing pages"

- Landing page hosts information (metadata) about the "object"

- Landing page points to persistent "home" of the object (museum, URL, etc)

- If anything changes, landing page must be updated

European Space Agency

DOI

10.5270/esa-0j79yk8

landing page

Data Set / Bundle	UCL-MSSL_iMars_HRSC_V1.0
Description	A high spatial resolution (50m) Digital Terrain Model (DTM) and orthorectified Image (ORI) have been produced for the Martian South Polar Residual ice-Cap (SPRC)for 33 HRSC strips and the associated ORIs at 12.5m. In addition, a 50m DTM mosaic has been created alongside a 12.5m ORI mosaic. For the ORI mosaic, Individual HRSC image strips have been corrected for different surface scattering properties prior to mosaicing.
Contact Point	Alfiah Rizky Diana Putri (alfiah.putri.15@ucl.ac.uk) and Jan-Peter Muller (j.muller@ucl.ac.uk)
Data Access	ftp://npsa01.esac.esa.int/pub/mirror/Guest-Storage-Facility/UCL-MSSL_iMars_HRSC_V1.0/
Data browse	www.i-mars.eu/web-gis and press "S" to show South polar view
Product User Guide	Link to the PUG
DOI	10.5270/esa-0j79yk8
Version History	V1.0 First version of this data set
Citation Guidelines	European Space Agency, 2019, UCL-MSSL_iMars_HRSC_V1.0, https://doi.org/10.5270/esa-0j79yk8
Associated Publication(s)	Putri, A. R. D., Sidiropoulos, P., Muller, J. P., Walter, S. H., & Michael, G. G. (2019). A New South Polar Digital Terrain Model of Mars from the High Resolution Stereo Camera (HRSC) onboard the ESA Mars Express. *Planetary and Space Science*. DOI: 10.1016/j.pss.2019.02.010
Dataset credit	When publishing any works relating to this dataset, please cite the afore-referenced "Associated Publication". When reproducing any of the datasets on web-pages or publications you should include the following credit line ESA/DLR/FU-Berlin/UCL
Mission(s)	Mission : Mars Express Instrument : HRSC

data

collection

noun

UK

🔊 /kəˈlek.ʃᵊn/ **US**

🔊 /kəˈlek.ʃᵊn/

---

collection *noun* (THINGS TOGETHER)

B1 [C]

a group of objects of one type that have been collected by one person or in one place:

• *a private art collection*

• *a valuable stamp collection*

Granularity

- DOIs are assigned at collection level

- Collections are defined based on area of research

- Collections can contain data from more than one mission

- Different collections can contain (part of) same data