The Herschel Documentation Legacy

Mark Kidger, for the HSC, NHSC, ICCs & MOC
30/05/2016

HUG #10

European Space Agency

Outline

1. POPS framework for documentation.
2. POPS objectives and how documentation fits in.
3. What do we mean by “legacy documentation”?
4. Documentation conservation and legacy requirements
5. Implementation of the Legacy Documentation Archive (HELL – Herschel Explanatory Legacy Library).
6. Release schedule (where are we right now?)
7. Other related Legacy documentation tasks.

The POPS Framework for Documentation

1. The top-level objects for Herschel Post-Operations [taken from the Herschel POPS plan] are:

    - Maintain high quality community support for ongoing Herschel data exploitation.
    - **Create, advertise, and eventually deliver, the Herschel legacy for posterity** → *The framework for Legacy Documentation*.

2. The aim is to leave the largest possible Herschel legacy for posterity, including the sum of all knowledge gained by the Herschel mission that will remain available to the scientific community
    - Documentation is an essential part of the legacy of Herschel knowledge.
    - What knowledge is not properly documented is inevitably lost
        - Personnel retire, memories fade, or become corrupted.

Documentation: the POPS Framework

- Experience from previous space missions (e.g. IRAS, or Apollo) is that people will be working with the data in new & innovative ways for several decades.
  - In 30 years time, even the youngest members of the Herschel Operations team will have retired.
  - Documentation will be the only available resource to understand how data was obtained.
    - Nobody will be able to pick up the telephone, or send an e-mail and ask “What happened on OD-324? There is something odd about the PACS data for TT Cyg taken that day.” (A SIAM issue)
    - Collective memories are already fading rapidly 3 years after End of Helium, as the distance from events increases.
      - *The figure of **Mission Archeologist** is becoming increasingly important.*
- We have a duty to identify and preserve the most essential information required to be able to understand the Herschel mission and its data as far into the future as is required.

What do we mean by “Legacy documentation”?

- There are many components of existing Herschel documentation:
  - Millions of words in tens of thousands of documents describing the requirements, design and implementation of Herschel, its instruments and the Ground Segment.
    - Many of these are proprietary to industry.
    - Most are highly technical and meaningless to future users.
  - Dozens of user manuals prepared for Uplink and Downlink users.
    - E.g. HIPE (data processing) guides must remain available and meaningful in a (possibly) HIPE-less future.
    - Is Uplink documentation (i.e. how to prepare observations) still meaningful to Legacy users? Some, probably “yes”.
  - Thousands of Minutes of Meetings and CCBs
    - Is it worth conserving? Probably not, save in exceptional cases!
  - About 50 000 JIRA & Helpdesk tickets in total
    - Full of information that may be nowhere else, but there are major confidentiality issues.

What do we mean by “Legacy documentation”?

- Hundreds (thousands?) of Technical Notes, Reports, Internal Reports and Aide Memoires produced by ICCs, the HSC , NHSC & MOC.
  - Without an overview, we risk losing vital information.
- Thousands of documents of all kinds on the Herschel Twiki pages.
  - Often the only document repository used during Operations.
  - Essential for “quick reaction” documentation. Sometimes the only record of an issue during Flight Operations.
  - No one can guarantee that this information will remain available in the future if not transferred to a safe repository. Save what is really important.
- Logs of all kinds.
  - Will all of these still work in the future? Save what is important!
- Thousands of pieces of information in the Herschel collective memory
  - Many of them never written down.
  - Identify what information needs to be “dumped” to conserve it.

What do we mean by “Legacy documentation”?

- Telemetry is our documentation in bits and bytes of everything that Herschel did in space until passivation.
  - We know that, barring a tiny amount of corrupted data rejected by the on-board Solid State Mass Memory, all TM generated was transmitted to Earth and consolidated at MOC.
  - Did all this telemetry make it into the Archive to be preserved?
  - Are there any remaining gaps in the TM record that need to be filled?
- Uplink products should be conserved too (requested by the ICCs)
  - Calibration deliveries (often with detailed rationale of the calibration observations and their objectives)
  - Mission Configurations (CUS and Calibration files – the full instrument control software record for the mission)
  - MiBs (Mission information Bases)
  - Mission Planning files
- **Legacy Documentation is much more than just WORDS.**

How do we conserve “Legacy documentation”?

1. Take stock of what is available and where it is.
2. Decide what needs to be conserved and what can be neglected.
3. Decide what is missing (and who should provide it).
4. Compile the documentation.
5. Prioritise documentation
    i. Not everything is equally important
    ii. What is the most essential reading for a future users to understand Herschel and its data?
6. Decide how to present the documentation
    i. File format (no platform-dependent formats)
    ii. Tagging? (Highly desirable, but could be a lot of effort)
    iii. Directory structure
        i. It is wasted effort if people cannot find the information that they need and cannot read it when they find it.
        ii. A document should lead a user to other, related documents..

How do we conserve “Legacy documentation”?

7. Decide the requirements for what the Legacy Library should do and how it should do it.
8. Understand the available repositories and infrastructure and their potential issues.
    - Can we satisfy our requirements?
    - If not, are the compromises that we have to make acceptable?
    - Can we adapt our wishes to reality?
9. Design and populate successively:
    - A sandbox library (get a feel for the contents).
    - A draft library (try different structures and see how they work in practice). Semi-public – used “in house”. Plentiful internal feedback.
    - The final user interface. ⟸ **We are here now.**
10. Make the user interface available to the community (initial release) and request feedback.
11. Make further updates and user releases until end of POPS.

User Requirements for the Herschel Legacy Library

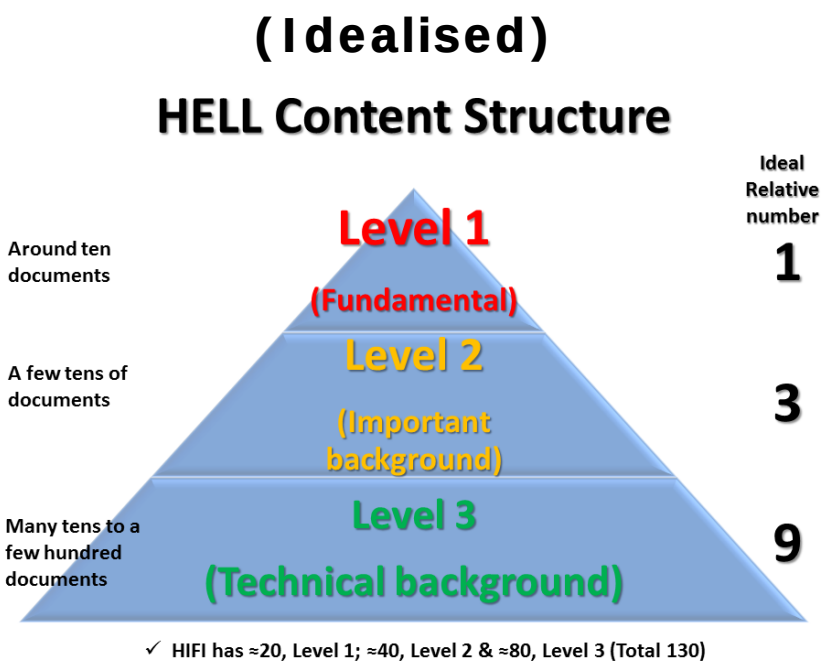
- It should be a library of essential Herschel technical documentation
- It should hang from 4 Explanatory Supplement volumes similar to the ISO Explanatory Supplement
  - One for each instrument and a volume for the spacecraft.
  - Each volume provides a complete top-level overview to the user that points the user to where to find further information, if needed.
- It should be arranged in 3 tiers of importance
- As a first approximation, the top two tiers can be understood as the reference lists for the 4 Explanatory Supplement volumes.
  - Tier 1 is a small number (<20) of the most essential reference documents for each instrument.
  - Tier 2 a rather larger number (a few tens) of more technical documents for obtaining deeper knowledge.
  - Tier 3 is the most technical or historical documentation of strictly limited readership to understand an issue in detail (a few hundred documents)

User Requirements for the Herschel Legacy Library

- It should be under our control, or at least be developable in a direction defined by us.
- It should be maintained and easy to maintain even after HSC has closed.
  - Links should not break (but be fixed if they do).
  - The documentation repository should be guaranteed long-term.
- Links should be transparent, with a simple structure
  - E.g. *\PACS\Performance\Pointing**** that allows users to identify where more documents on pointing can be found.
  - If a document is found through a search engine, it should be easy to find other, associated documentation.
- It should be fully searchable through Google and preferably automatically tagged to identify relevant documents in a search.
- There should be a small enough number of documents in the top two tiers to be easy to use and to find documents.
  - Concerns that if too much documentation is dumped, information will be buried in plain sight.

User Requirements for the Herschel Legacy Library

- Documentation should not be platform dependent.
  - Default format is PDF.
  - No Word (or similar) files.
  - Spreadsheets accepted in exceptional cases, but only if strictly necessary.
- Date and version number essential.
- Informative document titles strongly encouraged (not always possible for older documents).
- Provide a spreadsheet with a directory of all documentation for each instrument.
  - Categorise documents into areas in the spreadsheet to facilitate populating the library.
    - Overview
    - Performance & Calibration
    - Data Products (HPDPs + ancillary data)
    - Data Processing
  - Organise the documents on the page by these topics to make it easy to find what you need.



The actual content turns out to be remarkably close to this “ideal” structure.

Implementation: the HIFI example

- Initial documentation inventory
  - Members of the Herschel Community Support Team trawled for available documentation in the Herschel Twiki.
  - Passed to an instrument contact for each instrument who, with ICC collaboration, identified gaps and added additional documentation.
- Instruments have provided their spreadsheets with the final inventory of documents in their ICC wrap-up.
  - A few documents still TBW – usually papers in preparation – identified and a responsible and target date for completion identified.
  - Links in spreadsheet to all available documents.
- This is the HIFI spreadsheet: as submitted at wrap-up http://herschel.esac.esa.int/twiki/pub/HSC/HIFIGeneralLevel1/HIFI_DOC_v2.xlsx
- And this is the inventory when converted into a draft layout on the Twiki: http://herschel.esac.esa.int/twiki/bin/view/HSC/HIFIlLevel1

Implementation: the Portal

- We wanted a simple, user-friendly Portal to the pages, linking to the instruments, to the Explanatory Supplements and to the different levels of documentation.
- The draft page in the Twiki has allowed us to test different designs:
  - http://herschel.esac.esa.int/twiki/bin/view/HSC/HerschelLegacyDocumentation
  - The Twiki has less functionality than the Web, but has been a good testbed for a definitive design to port to Cosmos.
  - This approach was *endorsed* by the Herschel Users Group #9.

Cosmos Portal Basic Conceptual Design

Satellite	HIFI	PACS	SPIRE
ReadMe			

Click on a Tab in the Portal to open the appropriate instrument. For the active instrument a side-tab allows you to select the level that is being viewed.

	Satellite	HIFI	PACS	SPIRE
Handbook	HIFI Level 2 Content			
Level 1				
Level 2				
Level 3				

Instrument Page

Click on a tab to go to the desired Level and Tag content

HIFI	Handbook	Level 1	Level 2	Level 3
Overview	HIFI Level 2 Performance & Calibration Content			
Performance and calibration				
Products				
Data Processing				

Levels 2 & 3 may have tens or hundreds of documents, so selection of content by tag helps find information.

Implementation: the Cosmos Portal

- The Cosmos Portal has been set up with Orphan pages to allow content to be set up, without being visible to the outside world, until we are ready to reveal it.

- The design will be similar to – but more elegant than – the draft Twiki design.
  - **Visually more attractive.**
  - **We are working on the implementation.**
  - **Only a concept design implemented so far.**
    - Requires dedicated Cosmos support to be assigned to this task to pass the Concept to a visually attractive final version.

- Link to the concept Cosmos Portal (work in progress).

Documentation Repository

- Most documentation is currently in a temporary repository on the Twiki.
  - Transferred from ICC repositories (often password protected).
  - Under HSC control, but…
  - We have no guarantee of long-term preservation of the Twiki.
  - Everything needs to be transferred to a definitive location to give permanent links.
    - Essential for the Handbooks.
    - Required for the Cosmos pages.
    - Several hundred documents have to be linked.
- **Where the final repository is for documentation is not important**, provided that it fulfils our needs and requirements, guarantees long-term preservation of documentation and is maintained by ESA after HSC has closed.

OpenText (Livelink)

- Official ESA documentation repository.
- If, for any reason, it ceases to exist in the future, it will be replaced with an equivalent system.
- Guarantees long-term preservation of documentation.
- Stores Metadata for documents
  - Title
  - Authors
  - Summary
  - Date created
  - Version
  - ESA document code
- Currently all documentation stored in the Herschel directories is password protected, but can be made visible to all users.

OpenText (Livelink) v Requirements

- Most of our requirements can be met by OpenText with some provisos:
  - It takes **23 clicks** to upload each document.
    - Important when you have several hundred documents to upload.
    - Bulk upload is possible using a Windows plug-in, but only saves 3 clicks, as all the Metadata still has to be entered by hand.
  - Tagging is not implemented in the ESA installation.
    - It would make it easier to find Herschel documentation by Google.
  - If the metadata is not entered (e.g. not known or not applicable), OpenText may give a misleading default.
  - Links are not very informative that other content may be available too: http://www.rssd.esa.int/cs/livelink/fetch/2000/3371083/10737/3396562/3396672/HIFI_test_plan_and_test_report_-_SW_modifications_for_OD81_failure_mitigation.pdf?nodeid=3396017&vernum=-2
  - Google searchable????
    - **This is our biggest concern.**

Plan B: Cosmos (LifeRay) – A new ESA system (if Livelink does not meet needs)

- However, since starting to work with Legacy documentation, ESA has provided Cosmos: also an official ESA system that guarantees long-term preservation and maintenance long after HSC has shut down, with its own long-term documentation repository.
- Cosmos's documentation repository offers some very interesting features that would greatly facilitate Legacy Documentation conservation:
  - It takes only **10 clicks** to upload each document.
  - It picks up metadata automatically from documents (e.g. date of creation of document).
    - Less metadata entry required ("wasted" effort).
  - It allows tagging of individual documents.
    - Very simple to do. Facilitates web searches.
  - Documents get stored with the Web pages
    - Less risk than requiring two systems to be up and communicating.
  - We have some on-site Cosmos support at ESAC.

Cosmos Orphan pages

- **The pages are now populated, with a conceptual page design:**
  - **Top level Portal**
    - http://www.cosmos.esa.int/web/herschel/legacy-documentation
  - **Instrument Portals**
    - http://www.cosmos.esa.int/web/herschel/legacy-documentation-observatory
    - http://www.cosmos.esa.int/web/herschel/legacy-documentation-hifi
    - http://www.cosmos.esa.int/web/herschel/legacy-documentation-pacs
    - http://www.cosmos.esa.int/web/herschel/legacy-documentation-spire
- **Now we just need to place the documents in the appropriate repository and make the links.**

Release Schedule & Deadlines

1. **Now** – Pre-release. Everything available has been posted on the Twiki, grouped by instrument, level & keyword (tag). **Version 0**.
    - Pocket guide & re-vamped instrument web page templates needed.
2. **LSP-RR** (May 10) – Formal presentation of HELL, as is, for all 4 instruments and of Cosmos plans. **Version 1.**
3. **HUG** (May 30) – Pre-release of Cosmos version (PACS only) and of Cosmos design. **Version 2**.
4. **DP Workshop** (October) – Explanatory Supplement volumes approved for initial release by Management. Public release of HELL v1 to community. **Version 3**.
5. **Last ICC shutdown** (Dec 31) – Additional content & upgrades go public. **Version 4**.
6. **Deadline for major new content** – End of June 2017 (end of instrument support at HSC). Anything later on a best-efforts basis only.

Release Schedule & Deadlines

1. **HELL** – First public release October 2016 (HELL v1).
    - All instruments must have releasable (but not necessarily final) content on this date, approved by Management.
2. **Instrument Handbooks** – Updates as appropriate.
3. **Pocket Guides to instruments** – June 30th for HIFI & SPIRE.
    - PACS already has PACS Products Explained, so has a waiver for the Pocket Guide until October 31st.
4. **New instrument Flyers** – October 31st for all instruments.
5. **Other new documents** – December 31st (HELL v2)
6. **Final public release** – HELL v3, June 30th 2017 (incremental upgrades expected both before and after – Possible HELL v4, October 31st 2017, if required). This will incorporate the latest refereed papers to be published.

Related Tasks: “Preserving” the Twiki

1. There is a huge amount of information in the Herschel Twiki. ICCs and the NHSC each have their Twiki too.
    - Some is ephemeral (e.g. processing logs and not worth preservation).
    - An important part is high-value (e.g. calibration pages), which should have priority for conservation.
    - ESA’s official wiki is Media wiki
        - Conversion of Twiki to Media Wiki is not straightforward.
        - Conversion of Twiki content to Cosmos appears the best option for long-term conservation.
            - This may turn out to be a bigger task than we were anticipating initially.
        - Have offered to dump ICC Twikis to Cosmos on a best-efforts basis, accepting that not everything will convert perfectly.

Related Tasks: Preserving Uplink products

2. There is a raft of other important Legacy documentation that must be conserved.
    - Uplink products –this consists of
        - Calibration deliveries from ICCs with their documentation
        - Mission Configuration (CUS) deliveries (249 individual MCs were prepared for scheduling)
        - Herschel Planck Satellite Databases (HPSDBs, i.e. MiBs).
        - HSpot properties files (linked to a particular MC to configure the instrument settings in AORs).
        - Manual commanding files
        - Orbit files
    - Mission Planning products –consisting of
        - POS, ICP, SSF files, Planning logs

Compiled and delivered to the Herschel Science Archive as Ancillary Data Products.

Related Tasks: Telemetry & Satellite log

3. Telemetry.
    - Ensuring that the TM database is as complete as possible.
        - A major exercise of gap-filling has been going on since passivation.
        - However, due to way that data is stored, many potential gaps are flagged in checking that turn out not to be real.
            - System Source Counter jumps
            - Observations without telemetry

        Most cases due to SEUs causing instrument anomalies, but each has had to be checked by hand by diving into a wide range of logs and reports to check that there is no genuine missing data (Mission Archaeology).

        Has fed back into a major exercise of revising and updating the Spacecraft & Instrument Events & Activities Log - OD by OD , as it was found to have numerous omissions & inaccuracies.