

Keeping Our Databases In Orbit

Holly McIntyre-DeWitt & Matt Pearson,
NASA Goddard Space Flight Center

Background

The Goddard Space Flight Center (GSFC), NASA's first space flight center, is the largest organization of U.S. scientists and engineers dedicated to increasing knowledge and understanding of the Earth, Sun, planets, and the Universe. The Center engages in the development, integration, testing, launch, and operation of unmanned scientific spacecraft and payloads.

Spacecraft and instruments come into being through an evolution of technologies and processes. Configuration Management Databases (CMD) are systems used to manage the documentation, design, integration, and testing activities throughout the project lifecycle, leading to flight readiness. Spacecraft and instruments often require support for several decades beyond development. Once a spacecraft is launched, the CMD enters into the mission operations phase and is typically used for read only/reference purposes. At this stage, the databases are considered a working record, requiring that their context and relational structure remain in-tact and accessible to users on center. The science and development is evolutionary in nature, and therefore the CMD of a decommissioned spacecraft or instrument would remain a relevant and vital resource for future projects.

The National Archives and Records Administration (NARA) has determined that Goddard's CMD are not permanent federal records as a whole and will not be transferred into their custody as such. To meet its community's needs, the Goddard Archives is investigating effective strategies for acquiring, preserving, and managing access to these databases during and beyond the life of the spacecraft or instrument.

Getting started / high-Earth orbit

The Goddard Archives preserves and provides access to key material by facilitating and fostering relationships in order to identify, protect, and retrieve material that provides evidence relating to the Goddard Space Flight Center's legacy.

Retired (legacy) databases are at greater risk due to inactivity, loss of institutional memory and context, less-than-adequate storage and management, and changing software and hardware landscapes. The current strategy of extending access to retired databases falls short of guaranteeing enduring access. Until now, Goddard has not had a clear long-term, on-Center strategy (budget, commitment, plan, etc.) for preserving databases and similar content.

Databases, like other information systems, are comprised of three layers: interface (front end), data, and business rules. While many digital preservation strategies ensure the safekeeping of data, the Goddard concept of "database preservation" considers data, plus aspects of the other layers in order to preserve context and meaning.

Our proposed database preservation process begins with identifying and characterizing our existing systems, including documentation of creating organization, project, function, and points of contact. Risk, requirements, access needs, and use restrictions will be determined during database appraisal. We will devise a process for the physical transfer of a faithful database copy and descriptive metadata will be captured based on provenance and content. Then the systems will enter our digital preservation / OAIS-aligned workflow.

Strategy / medium-Earth orbit

The Goddard Library team will support the Archives in clarifying aims and in careful consideration of factors such as access requirements, restrictions, sustainability, data loss, cost and capability. The Library is investigating a number of popular database preservation strategies:

- migration/normalization (to MySQL),
- emulation (Emulation as a Service),
- Containerization (Docker),
- Open Formats (SIARD with SIARD Suite and related tools),
- CHRONOS, XARCH, and other vendor and partner solutions, and
- Custom, in-house solutions.

The team will perform a landscape analysis and basic research about these strategies, which, combined with an assessment of our organizational goals and needs, will help us identify the most practicable solutions to further evaluate. This process will include testing over several months for each strategy. It will require standing up a test environment, acquiring copies of databases, and applying and testing these preservation strategies. A qualitative and comparative analysis of workflow, quality, performance, and risk will allow the team to make an informed decision on the method that best meets the Archives' needs.

The winning approach will be integrated within our preservation system and workflow, which is currently being developed around Archivematica and Fedora repositories. The Archives aims to leverage this work to better support the rest of the Center by producing internal database transfer guidance at the project level. We will also develop a feedback process for informing the design of future and active systems to better fit with the Archive's preservation capabilities.

Observations and plans / low-Earth orbit

There are a number of challenges that will need to be overcome during our process. We must understand and build an appropriate environment for accepting and managing restricted electronic material. This means working closely with IT staff and our Security Office on infrastructure, policies, and procedures for accommodating sensitive but unclassified (SBU) and ITAR/EAR-restricted content. We will need to determine how to best provide our researchers with the right kind of access to these materials both in

terms of access and use restrictions, as well as finding an interface that will allow the researcher to retrieve the desired records from these complex systems.

At the technical level, even if we are not supporting database front end and software environments, a sustainable approach will be necessary for representing existing data transformations and cross-system dependencies. The form of the preserved database will differ from the original form and it will be important to make sure that relational information is documented and context is not lost.

We will need to work closely with stakeholders from creating offices to better interpret the function of the CMDs and to determine an acceptable level of data loss. Local politics and varying cultures across the center will require clear communication and outreach.

Conclusion

Developing a preservation solution for our center's Configuration Management Databases is a complex, multi-year process that requires careful collaboration with creators, database modeling and querying expertise, a deep understanding of preservation needs, high levels of experimentation, verification, and thorough documentation.