A Tale of Two Models: Comparing OAIS and the USGS Science Data Lifecycle Model

Data collection occurs in a variety of environments. An individual on a road trip may make note of how frequently they see license plates from various states or makes and models of vehicles. A company may track sales, expenses, buying habits, profits, and a host of other information. A library may keep tabs on what services customers use most or what books are most popular, to say nothing of the data about each item in its collection. Whatever the information, they need some means of keeping track of everything they collect and ensuring its accuracy. For simple lists, tallies on a sheet of paper or even mental notes may suffice. For more complex data sets, data lifecycle models help to understand, conceptualize, and standardize information collecting, such as the Open Archival Information System (OAIS) and U.S. Geological Survey Science Data Lifecycle Model (SDLM). These models serve different purposes, so dissimilarities between them are to be expected. However, a comparison of the two also shows a number of functional similarities.

Open Archival Information System (OAIS)

The OAIS Functional Model was developed by the Consultative Committee for Space Data Systems (CCSDS). This model defines the flow of information from a producer to an archive and the archive's subsequent interactions with "potential consumers who should be able to understand a particular set of information. OAIS makes the content information…visible and available to its designated communities" (Kress & Danko, 2022, p. 443-444). The model is graphically represented as a sort of box with complex inner workings represented by lines showing how data flows through the archive from producer to consumer. The OAIS model is divided into six "functional entities": ingest, archival storage, data management, administration, preservation planning, and access. The ingest function accepts submission information packages (SIPs) from external creators or internal administrators and gets data ready for storage in an archive. This function also includes quality assurance of the SIP, creating an archival information package (AIP) based on the particular organization's standards, using information from the AIP to create a Descriptive Information Package (DIP) for the archive database, and coordinating updates. The function of archival storage is where storage, maintenance, and retrieval of AIPs takes place. Data management takes care of descriptive information of archive holdings as well as administrative information needed to manage the archive. The administration function provides for overall archive operation, including submission agreements, submission audits, management of hardware and software, operations monitoring, establishing and maintaining standards and policies, and customer service. The preservation planning function monitors the archive environment, makes recommendations and plans to keep data accessible and comprehensible over the term. The final function, access, is a customer-facing function that allows users to request and receive information (4.1 functional model, 2015). A preliminary function known as pre-ingest has been added to this model by organizations, such as the United Kingdom Data Archive. This step helps to ensure data quality and accessibility while also ensuring that submissions meet a certain standard, which reduces workload and increases efficiency of the ingestion function (Preservation policy, 2025, p. 14).

USGS Science Data Lifecycle Model (SDLM)

The United States Geological Survey (USGS) Community for Data Integration (CDI) was established in 2009 to address data and information management issues affecting the Bureau's scientific research. An internal working group was established in 2010 to recommend best practices for the bureau, including a data lifecycle model (Faundeen et al., 2014, p. 1). The final version of the USGS Science Data Lifecycle Model (SDLM), which was approved in November 2012, incorporates six primary stages or "elements": plan, acquire, process, analyze, preserve, and publish/share. The plan element is designed to ensure a project's entire data lifecycle is considered at each stage of a project to better determine the appropriate approach, personnel, material resources, financing, and outputs at each stage. The second element, acquire, involves the collection and evaluation of data, whether it is created or acquired from existing sources, and considers "policies and best practices that maintain the provenance and integrity of the data" (p. 2). The process element involves processing the data acquired in the previous step, including defining elements, transformation, and calibration to create a collection of datasets ready for the next stage. The analyze element involves exploration and interpretation of those datasets, generating new datasets or interpretations of the previous data, which are often "published in written reports or machine-readable formats such as map layers or numerical modeling results" (p. 3). The next element, preserve, addresses long-term storage and access of the project's data, metadata, reports, and other documentation. The final primary element, publish/share, involves publication and dissemination of the resulting information "through Web sites, data catalogs, social media, and other venues" (p. 3).

The SDLM is graphically much simpler than OAIS, with the primary elements represented as a simple chain of boxes with arrows showing the flow of information. Beneath that are three long arrows, each spanning the length of the model. These arrows represent the "cross-cutting elements" describe, manage, and backup and secure, which are "critical activities [that] must be performed continually across all stages of the lifecycle to help support effective data management" (Faundeen et al., 2014, p. 3). The first of these elements, describe, addresses the need for ongoing metadata and documentation throughout each primary stage to help ensure accuracy and comprehension later. Manage quality, the second element, incorporates data quality control via monitoring and adjustment across each stage to help maintain accuracy. The final cross-cutting element is backup and secure. This element serves as a reminder for routine backups at every step to minimize or prevent "the physical loss of data because of hardware or software failure, natural disasters, or human error before the final Preservation [sic] of the data" (p. 3).

Comparing the Two

Even though they may look quite different, OAIS and SDLM share some key, basic similarities. They are, after all, both lifecycle models with the same basic goal. The OAIS ingest function is analogous to the SDLM acquire element, where collection of some form of information or data takes place. The ingest function also shares commonalities with the cross-cut element of describe, where data descriptions and metadata are created. Likewise, UK Data Archive's pre-ingest function shares some similarities with the plan element found in the USGS model as pre-ingest involves work at the outset of the lifecycle that is designed to facilitate the subsequent stages. The OAIS functions of archival storage and data management find counterparts in the SLDM elements of preserve, manage quality, and backup and secure. In these components, both models focus on data quality, preservation, and long-term access and understandability. And finally, both models move toward long-term preservation and making data available to potential consumers, whether via queries (OAIS access function) or outright publication (SDLM publish element).

That being said, however, there are some stark contrasts between OAIS and SDLM, largely arising from their foundational differences. OAIS, as the name implies, focuses on archival data. Information does not just pass through and out; rather, the model is recursive. Data comes into the archive and is shared with users, but it remains in the organization – it "lives" there, as it were. On the other hand, SDLM is a linear model. While the USGS does keep an eye toward long-term preservation and usability, its lifecycle model does not concern itself with the actual archival process. The model is about collecting data, processing it, and reporting on the results; it focuses on what can be learned from the data. While the USGS may house some of its own information, the publications that result from its processing are just as likely to “live” elsewhere in the end. This difference in focus can also be seen in how the models portray what happens with the data. While both strive to make sure the data remains accurate and stress monitoring and quality control, OAIS does not address analyzing or processing the information in any mathematical or scientific sense. It simply collects and makes available. SDLM, on the other hand, is primarily concerned with accuracy so that the data can be correctly analyzed, processed, and reported on. It is, first and foremost, a scientific data model.

Conclusion

Data can be complex to sort through and save for future use. The Open Archival Information System and U.S. Geological Survey Science Data Lifecycle Model both exist to help make sense of the processes involved in ensuring that data is preserved and useful in the long-term. However, not all data is the same, and neither are the reasons for collecting it. OAIS exists to make sense of how archives take in information, store it, and make it available to users when they need it. The SDLM, on the other hand, primarily exists to make sure data is accurate and usable for a specific project leading to a particular report or other dissemination – although it does keep an eye on long-term accessibility and usability, as well. Interestingly, despite (or perhaps because of) their differences, these two models are not mutually exclusive. Data collected by the USGS using the SDLM could easily be used in a report that eventually finds its way to a digital archive that makes use of the OAIS model. It all depends on the purpose of the organization using the data.

References

4.1 functional model. (2015, October 5). Digital Preservation Coalition. https://wiki.dpconline.org/index.php?title=4.1_FUNCTIONAL_MODEL

Faundeen, J. L., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., Hutchison, V. B., Martín, E., Montgomery, E. T., Ladino, C. C., Tessler, S., & Zolly, L. S. (2014). *The United States Geological Survey Science Data Lifecycle Model* (Open-File Report Nos. 2013–1265). United States Geological Survey. https://doi.org/10.3133/ofr20131265

Preservation policy. (2025, February 17). UK Data Archive. https://dam.data-archive.ac.uk/controlled/cd062-preservationpolicy.pdf

Wolfgang Kress, & Danko, D. (Eds.). (2022). 15.3.16 Preservation. In *Springer Handbook of Geographic Information* (2nd ed., pp. 442–444). Springer Nature. https://app.knovel.com/hotlink/pdf/id:kt013N1PK1/springer-handbook-geographic/open-archival-information