**Christopher Roberts**

**MSFC EM-04**

**Microgravity Science Database Development**

1. **Introduction**

Throughout NASA’s history, the agency has developed a plethora of complex systems, such as the International Space Station and the space shuttle, and performed research in several fields spanning the gamut from psychology to welding and materials research. Throughout these studies, an extensive amount of data has been generated and unfortunately at times regenerated. As Barend Mons states “Huge sums of taxpayer funds go to waste because such data cannot be reused.”[2] While his comments were directed at the state of data management in the European Union, it is no less valid for data management practices in the United States. The issues surrounding data management, including storage, retrieval, and analysis, will continue to be of utmost importance as the agency aims to responsibly utilize funds and gather the maximum benefit from flight and ground experiments.

While the development of database management strategies, can be applied to most, if not all the research topics applicable to NASA, advanced manufacturing, and in particular additive manufacturing, is a critical technology which enables future NASA missions and the goal of a sustained lunar or Martian presence. Furthermore, Additive Manufacturing (AM) crosscuts two areas of high importance as noted by the decadal survey team: “Materials synthesis and processing to control microstructures and properties” and “Research on processes for in situ resource utilization”.

AM, first explored in its modern form in the late 1980s, is perhaps the first widely adopted manufacturing method that developed solely within the digital era. As such, a large quantity of data, such as powder characteristics, environmental conditions, processing parameters, and test data, is generated during and associated with each production run from both manual entry and automated sensors. While this plethora of data is invaluable during research and analysis of build anomalies, this data is not typically utilized as a predictive tool and is seldom analyzed as a whole in a production setting. Proper data management and analysis techniques directly impact the ability of NASA to develop standards and quality parts and processes.

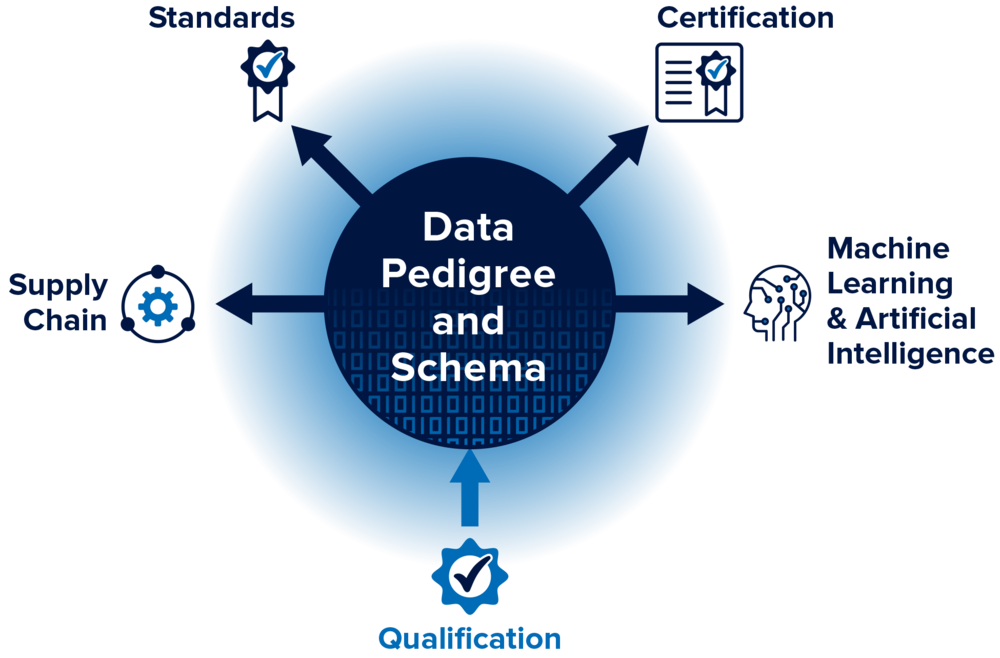


Figure : From ASTM Specialty Workshop Series: Additive Manufacturing Data Management and Schema [1]

The large quantity of data typically generated by AM processes and the criticality of the technology for enabling sustained explorations missions, makes AM exceptional system for demonstrating modern data management and analysis practices. Following demonstration of a wholistic approach to data management, practices can be generalized to other fields such as biological sciences and fundamental physics. By retaining data in a manner easy for researchers to access, NASA can enable future discoveries and more easily identify the effects of microgravity on experimental variables.

1. **Investment Opportunities**

There are multiple areas within data management that are worthy of further investment; three of which are: 1) Database Development and Testing 2) User Interface Development and 3) Application of Modern Analysis Techniques. Databases and the set interface are part of the infrastructure of information. Recently there is a push by industry and government to implements the ideals of FAIR (Findability, Accessibility, Interoperability, and Reusability) data management [3]. Funding should be focused on enabling modern systems which allow users to easily access data with a minimal learning curve or knowledge of proprietary data formats. In each area listed above AM data can be used to demonstrate principles and benefits of data management.

* 1. **Database Development, Testing, and Maintinence**

At the core of data management is the development, testing and maintenance of database and data archives. The structure and schema of a database directly affects both the usefulness and accessibility of a database beyond the principal investigator; however, often the schema of a database is developed by a separate team divorced from the science team who understands the underlaying relationships between data types and variable sets.

An AM database used to demonstrate best practices in data management, would enable insight into relationships between variables and identify the relative importance of parameters on defect development. One of the primary questions of interest to NASA in relation to AM and data management is: “What is the minimum viable data set necessary to be retained when certifying components for critical application?”. To understand the minimum viable data set, researchers must start with a complete dataset and analyze the influence of each variable on defect development. A risk-based approach can then be used to balance the costs of retaining data and the fidelity needed for a particular application. Furthermore, retaining a full dataset not focused on a particular subset of AM, such as powder data or processing parameters alone, would enable the database to be experiment agnostic. The same data could be utilized by one researcher to investigate the influence of powder morphology while another may be interested in the gas flow rate’s influence on mechanical properties. In this way the database may serve as a testbed for a variety of projects within minimal physical specimen development. As testbeds become available on orbit, process specific parameters may be added to the data set and algorithms verified for use in ground-based systems can be applied to flight payloads to elicit differences between ground and flight data.

While the concept of a generalized database for AM or for any other field has several advantages, the primary disadvantage is cost. To be useful for several modern analysis techniques such as artificial intelligence and machine learning, a large amount of data must be ingested into the database. However, as the data may be utilized for a variety of research topics costs can be recouped over time. Furthermore, open source and academic data may be added to the system and ranked accordingly. For instance, data collected using a certified test house containing the full data set would be ranked highly while open-source data contain a partial data set would be lower ranked. This approach allows for a large amount of data to be generated while allowing the researcher to make decisions as to what quality of data is needed for their experiment.

* 1. **User Interface Development**

While collation and archiving of data is the core of data management, the user interface largely determines the usefulness of the data to users beyond the initial research group. Clear and consistent practices can help inform potential users of data relationships and availability of particular data sets. Investment in the development of Application Program Interfaces (APIs) and Graphical User Interfaces (GUIs) allow consistent access to data structures through both manual extraction and programmatic routes. The intended use case and underlaying data should inform the development of these programs and should be outlined as the program develops from its inception.

For an initial demonstration utilizing AM data, the most effective user interface is likely APIs which allow access to the database by a variety of common AI/ML tools. Two of the most common toolsets used in industry and academia are MatLab™ and Python. While MatLab™ is not an open-source program, it is widely utilized in engineering courses and allows both undergraduate and graduate student to explore data in a similar environment increasing engagement of early career professionals and students in microgravity experiments. Python, in contrast, is freely available with a vide variety of statistical and machine learning tool sets available. APIs allowing direct communication between the database and the programing environment would allow the use of large datasets without the need to store a localized copy of the archive.

* 1. **Application of Modern Analysis Techniques**

The final area of data management in which NASA investment would be beneficial is the demonstration of modern analysis techniques and the investment in training and education for the NASA workforce. Over the past several years AI/ML has been heralded as the future of data analysis and big data applications. While AI/ML has constraints and challenges, it provides a new tool that can supplement traditional statistical analysis. Many modern technologies, such as AM, provide a vast quantity of data with subtle relationships between variables. AI/ML has the potential to analyze data and processes in real-time and predict potential outcomes.

For an initial demonstration with AM, AI/ML could be utilized to help understand the relationship between parameter sets, in process monitoring data, and final part performance and materials characterization. As AM technologies become available on orbit, additional insights into how microgravity affects grain development, defect density, and dimensional stability may also be aided by AI/ML. Furthermore, as the process becomes characterized, the system can learn what is a “good” build verses a “bad” build, allowing the operator to cease operation if there is a low chance of success thereby conserving feedstock and power.

1. **Conclusion**

Enhanced data management and training for engineers and scientists to use modern analysis techniques, such as AI/ML, are vital for enabling and enhancing NASA research as the agency focuses on sustained, long duration missions beyond low earth orbit. Not only does data management and analysis help limit the number of tests that must be repeated due to loss of data, these techniques allow technical authorities to make engineering, risk-based judgements such as where additional testing is needed and where we have enough confidence in existing knowledge to forgo specialized testing. Additional benefits may also be realized by the data enabling Integrated Computational Materials Engineering (ICME) which has the potential to speed of discovery of new materials and processes. In the scientific community, FAIR data management practices allow a larger number of researchers to benefit from flight experiment data and secondary discoveries to be made outside of the original experiments design. While multiple areas may benefit from integrating data management techniques early into the experimental design process, AM experiments and research provides a critical opportunity to demonstrate the development of these systems while providing benefit to NASA and the US industrial base. Within this paper, the author has laid out three key areas for further investment which can be demonstrated using AM data while meeting the focus areas of the decadal survey team.

**References:**

[1] ASTM. (n.d.). *Specialty workshop series: Additive Manufacturing Data Management and schema*. Specialty Workshop Series: Additive Manufacturing Data Management and Schema. Retrieved October 26, 2021, from https://amcoe.org/events/specialty-workshop-series-additive-manufacturing-data-management-and-schema?rq=Additive+Manufacturing+Data+Management+and+Schema.

[2] Mons, B. (2020, February 25). *Invest 5% of research funds in ensuring data are reusable*. Nature News. Retrieved October 26, 2021, from https://www.nature.com/articles/d41586-020-00505-7.

[3] Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). *The FAIR Guiding Principles for scientific data management and stewardship. In Scientific Data* (Vol. 3, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/sdata.2016.18