DEBRISAT'S DEBRIS CATEGORIZATION SYSTEM: A DATABASE-BASED SOLUTION TO THE BIG DATA CHALLENGES IMPACTING THE FUTURE REGULATION OF ORBITAL DEBRIS

Joe Kleespies(1), Norman Fitz-Coy(2)

(1) University of Florida, Gainesville, FL, United States, Email: joeykleespies@ufl.edu
(2) University of Florida, Gainesville, FL, United States, Email: nfc@ufl.edu

ABSTRACT

Orbital debris propagation threatens the future of space exploration. To address the orbital debris challenge, future regulations governing space operations need to be discussed, debated, and implemented. These regulations must be based on solid sets of scientific data. DebriSat is an ongoing hypervelocity impact (HVI) experiment that aims to generate new scientific data to empower informed decisions on future orbital debris regulation. The DebriSat HVI experiment produced hundreds of thousands of debris fragments, which are currently being collected, characterized, and recorded. The DebriSat Debris Categorization System (DCS) is a database-based solution designed to address the big data challenges posed by the DebriSat experiment. This paper outlines the standard big data framework and design metrics and describes how they relate to DebriSat and the orbital debris community. Furthermore, this paper describes the methodology and processes employed during the design and implementation of the DCS.

1 INTRODUCTION

The worsening orbital debris climate is becoming more and more threatening to the future of space exploration. Even without the influx of new debris, the Kessler Syndrome postulates that the orbital debris environment will continue to worsen due to continuous collisions between existing debris fragments. [1] To address the orbital debris challenge, future regulations governing space operations will need to be discussed, debated, and implemented. It is essential that these discussions and decisions are based on solid sets of scientific data. The focus now of the orbital debris community is to generate data sets needed to develop models to accurately characterize the orbital debris environment and empower informed decisions on future orbital debris regulation.

In 1992, the Satellite Orbital Debris Characterization Impact Test (SOCIT) was conducted by NASA. [2] The subject of SOCIT was a 1960's-era Navy navigation satellite. The impact test produced debris fragments that were used to create distributions of radar cross section, mass, size, and shape of the fragments. The data generated by SOCIT were used to update the NASA Standard Breakup model in 2001. However, the updated model showed a bias toward small objects in environments with many known small particles. [3]

DebriSat is another ongoing hypervelocity impact (HVI) experiment that aims to generate new, modern data to update and modernize breakup models and to address the biases in the current NASA Standard Breakup Model. [4] The subject of the DebriSat experiment was a representative LEO satellite produced using modern processes and materials. The experiment is currently in the post-impact phase where debris fragments generating from the HVI experiment are collected, characterized, and recorded. Based on existing models, the initial estimate for the total number of debris fragments generated by the test was approximately 85,000; to date, over 130,000 fragments have been recorded and the total number of fragments is currently predicted to exceed 300,000. [5]

The primary data being catalogued for each debris fragment generated by the DebriSat HVI experiment include mass, shape, color, size, and density. Additionally, there are multiple high-resolution images associated with each debris fragment that are used in various space-carving and imaging methodologies to accurately and reliably characterizes the size of each fragment (i.e., characteristic length, average cross-sectional area, and volume). Furthermore, there is a plethora of other metadata collected on each debris fragment used for tracking, cataloguing, and other miscellaneous purposes. Ultimately, the amount of data collected for each individual fragment multiplied by large number of debris fragments poses a classic big data management challenge that is becoming more common as similar large data sets are sought and generated.

2 BIG DATA AND REGULATION

Big data has had and increasing influence on regulatory decisions ranging from commerce, to climate, to now orbital debris. However, with this increasing influence comes the responsibility to ensure big data is being used correctly and responsibly. Companies like IBM and Ernst & Young have developed a high-level description of big data entitled "The Four V's of Big Data" to help identify the four major dimensions needed to utilize big data effectively and responsibly. [6]

Proc. 7th European Conference on Space Debris, Darmstadt, Germany, 18–21 April 2017, published by the ESA Space Debris Office
Ed. T. Flohrer & F. Schmitz, (http://spacedebris2017.sdo.esoc.esa.int, June 2017)

2.1 The Four V's of Big Data

The Four V's of Big Data are *volume*, *variety*, *velocity*, and *veracity*. A visualization of The Four V's of Big Data is shown in Figure 1.

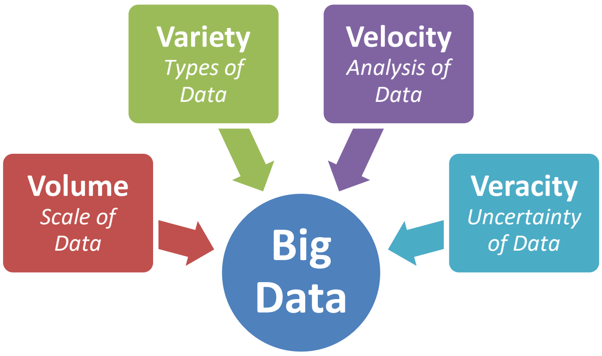


Figure 1. The Four V's of Big Data

Volume describes the scale of the data being collected and analysed. For example, the number of experiments performed or the number of sensors in a network can be used as a metric for the *volume* of data.

Variety describes how diverse the data set is and how many different types of data there are (i.e. textual, imagery, binary, etc.).

Velocity describes the process of analysing the data set and how fast data can be generated and analysed. For example, a network with 1000 sensors has a higher velocity than a network with 100 sensors.

Veracity describes how trustworthy or uncertain a data set is. *Veracity* can describe error in the data due to process or source.

2.2 Challenges of Big Data

While having good *volume*, *variety*, *velocity*, and *veracity* produces a good big data analysis, each of The Four V's of Big Data can introduce pitfalls, shortcomings, and ultimately error in the resulting analysis if not sufficiently satisfied.

2.2.1 Volume

"Lack of data" is a common phrase used to describe not enough volume in a big data analysis. For example, the authors in [7] describe how a lack of suitable data and models hampers the effort to measure food sustainability in the UK and enact suitable regulation and policy to address the challenge. They go on to describe how "there generally exists both a data and policy governance gap at the regional level." This is a classic case of how a lack of data and therefore lack of *volume* inhibits effective use of big data analysis to affect regulatory and policy decisions.

This example draws parallels to the current orbital debris regulatory environment. There are only a few experiments to date that have produced data to support big data analysis and conclusions to support regulatory decisions. There is a low *volume* of data regarding orbital debris, and DebriSat and the Debris Categorization System (DCS) aims to change that.

Further, the authors of [7] go on to describe how "despite the wealth of data available, it can be argued that type of data does not help fully understand food systems at the regional level." This furthers another point that the *volume* of data must be pointed, useful, and targeted. That is, the data set must have a good *variety*.

2.2.2 Variety

While a data set may contain a large *volume* of general or miscellaneous data, a "lack of data" can be extended to describe a "lack of useful data." As the authors in [7] stated, the type of data collected in regard to food sustainability in the UK are irrelevant and not useful for conclusive analysis of the underlying food systems.

Similarly, the authors in [8] describe how a lack of *variety* hinders the efforts to forecast climate change's effects on biodiversity. The authors discuss how "we can now build videogame-like environments with computers where we can create multiple versions of Earth and ask what the implications under different scenarios are, but our ability to learn from these tools is constrained by the kinds of data we have." The authors in [8] go on to discuss how much of the data and models are descriptive in nature and merely capture statistical correlations and observations. There is a need for more mechanistic data and models that describe and predict the underlying processes causing the correlations and observations in the descriptive models.

Once more, this example of a lack of *variety* parallels with the current orbital debris regulatory environment. Experiments and models to date are based on older data sets with limited *varieties*. The lack of *variety* and modern data causes biases in the breakup models and hinders the effort to achieve a comprehensive understanding of the orbital debris climate. DebriSat and the DCS aim to address this lack of *variety* by collecting a more modern data set with a broad suite of data types.

2.2.3 Velocity

The *velocity* of a data set refers to how fast a data set can be generated and analysed. Big data has recently started evolving to include more and more "fast data" where data is generated very quickly, streamed into an analysis tool, then analysed in almost real-time. The author in [9] uses an example of how stock market predictions have progressed with the rise of fast data. Ten years ago, investors typically couldn't react to a critical piece of market information until the day after it happens. With fast data producing analysis and conclusions in almost real time, investors can now start to act of market changes almost immediately.

In the case of orbital debris, this concept of fast data can be used to update or adapt existing models with real-time observations, data, analysis, and conclusions to produce more accurate and targeted conjunction analyses, breakup predictions, etc. Using existing infrastructure like Analytical Graphics, Inc.'s Commercial Space Operations Center (ComSpOC) with a large array of observation stations in combination with model data from experiments like DebriSat, quality big data analysis with fast data streaming can be achieved thus producing a high level of *velocity* for the analysis.

2.2.4 Veracity

In big data, *veracity* is a function of the least-trusted link in the source and analysis chain. Because good big data sets have high *volume*, *variety*, and *velocity*, each error in a data set is compounded by the "butterfly effect." [10] A minor error in such a complex system can compound and affect the analysis and conclusion. Therefore, it is paramount to achieve a high level of *veracity* in a data set to minimize the number of errors. DebriSat and the DCS have implemented various verification and validation processes that can be used and extended by future projects to achieve a comparable level of *veracity* in their data sets and analyses.

2.3 DebriSat and Big Data

DebriSat's data set continues to grow as more processes and methodologies are introduced. Currently, the estimate for the total count of debris fragments is in excess of 300,000 and the number of collected and recorded debris continues to grow as shown in Figure 2.

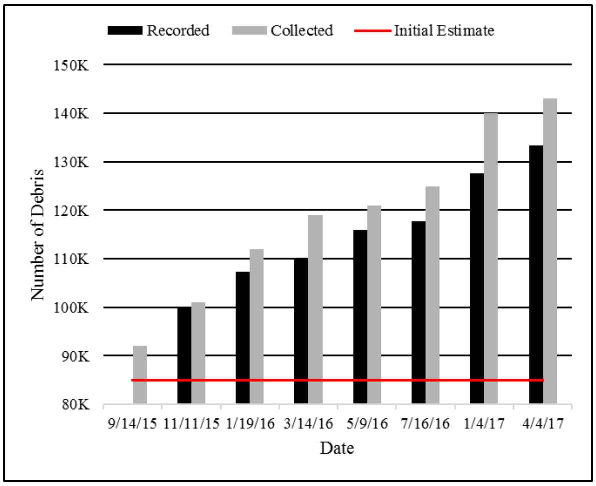


Figure 2. Number of Collected and Recorded Debris

Each of these debris fragments will be individually collected, characterized, and recorded. Ultimately, each debris fragment will have broad suite of associated data types and fields ranging from textual data for measurements such as mass, shape, and color to full imagery data used in various space-carving methodologies to determine the size and dimensions of each fragment. The number of debris fragments multiplied by the size of the associated data sets for each debris fragment pose a classic big data management challenge.

3 DEBRIS CATEGORIZATION SYSTEM

In response to DebriSat's big data management challenges, the DebriSat Debris Categorization System (DCS) was developed and implemented to provide a management structure for DebriSat's data that ensures the high levels of *volume*, *variety*, *velocity*, and *veracity*.

3.1 System Overview

The DCS is comprised of a front-end user interface and a back-end database engine and backup infrastructure. The back-end database engine processes all database requests and stores all of DebriSat's characterization data. The back-end backup infrastructure ensures the data is backed up in multiple locations every day to protect against data corruption or loss. The front-end layer provides an interface that guides users through the characterization process and facilitates verification and validation of the data. Figure 3 provides a visual overview of the DCS.

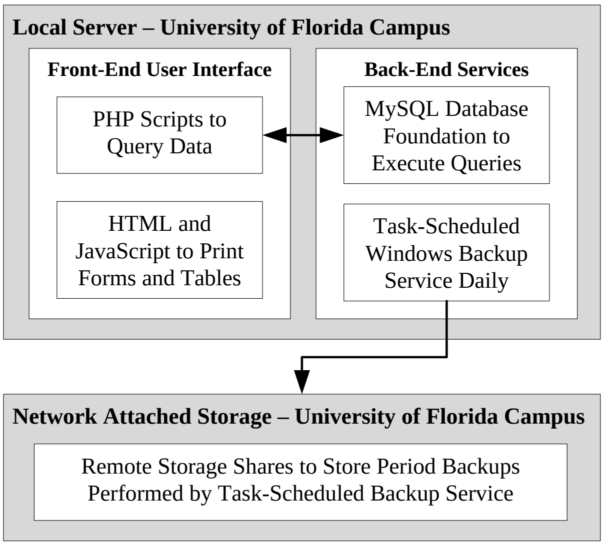


Figure 3. Visual Overview of the DCS

3.2 Back-End Layer Design

The back-end layer of the DCS includes the database engine, web service infrastructure, and the data backup infrastructure. The back-end layer serves the front-end user interface through a Microsoft IIS web service. Access control to both the database engine and the front-end user interface is facilitated using Microsoft Active

Directory (AD) Services. The full DCS database engine is backed up daily as a full VSS copy using Task Scheduler and an externally located network attached storage (NAS) device. Currently, the database engine at the heart of the DCS back-end layer is the MySQL database engine. MySQL provides the relational table structure used to store all data and entries produced by the DCS.

In order to enable secure and distributed access to the DCS, both the front-end and back-end layers of the DCS are hosted on the University of Florida internal network. To access the system, users need to either connect directly to the UF network or remotely connect to the network via virtual private network (VPN). The actual processing of debris fragments takes place at an off-campus facility. Users at the facility connect to the UF network remotely via VPN. Figure 4 provides a visualization of how the back-end layer is hosted and how it interfaces with the front-end layer and the end users.

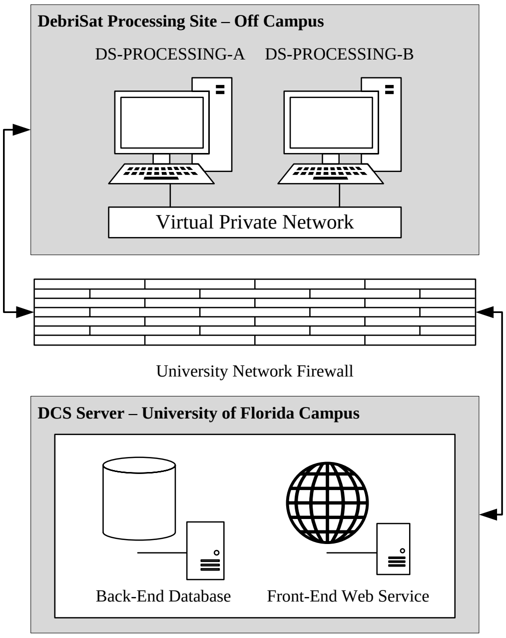


Figure 4. DCS Network Interfaces

The DCS back-end database engine includes several data tables that store activity, debris entries, soft-catch foam panel entries, general announcements and updates, color choices, shape choices, and material choices. Table 1 lists and describes these tables in more detail.

Each of the tables listed in Table 1 include a multitude of columns for each data type and field for each table entry. For example, the *dcs_debrisat_debris* table contains 174 columns for each unique data type or field. The back-end database engine can be queried using server query language (SQL).

Table 1. DCS Back-End Database Engine Tables

Table Name	Description
dcs_activity	Log any and all activity and queries performed on data.
dcs_debrisat_debris	Store all data for debris fragments from the test.
dcs_debrisat_foam	Store all data for soft-catch foam panels from the test.
dcs_debris_colors	Store all possible color choices for fragment characterization.
dcs_debris_shapes	Store all possible shape choices for debris fragment characterization.
dcs_debris_materials	Store all possible material choices for debris fragment characterization.
dcs_updates	Store general announcements and updates posted on the system.

3.3 Front-End Layer Design

The front-end layer of the DCS is written in HTML, JavaScript, PHP, and SQL. HTML and JavaScript are used to format and display web pages and forms that the user interfaces with. PHP and SQL are used to interface with the back-end layer of the DCS. The front-end graphical user interface (GUI) is designed to guide the user through the debris characterization process shown in Figure 5.

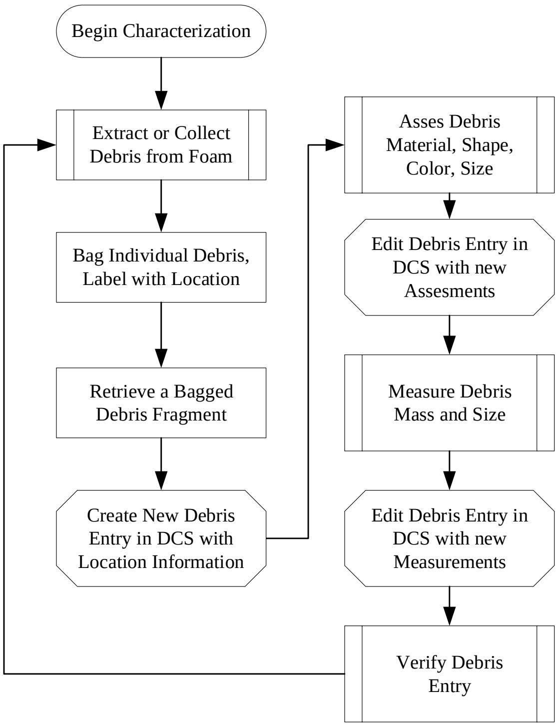
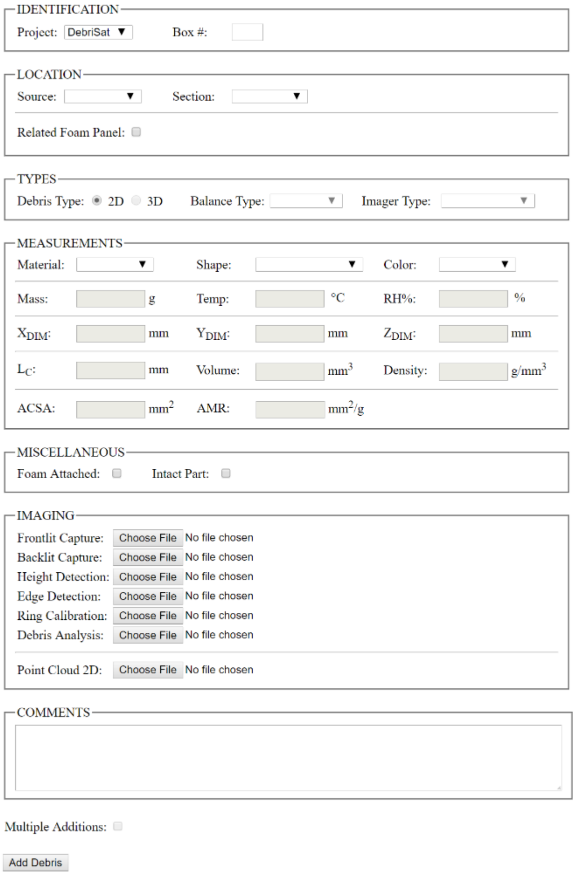


Figure 5. DebriSat Debris Characterization Process

The layout of each web page on the DCS front-end layer is designed to guide the user through a step in the characterization process. For example, the first thing a user does after collecting a debris fragment is to add an entry in the system for the debris fragment and obtain a unique identity generated by the server. This is all done through the “Add Debris” page on the front-end layer of the DCS. Figure 6 shows a screenshot of the current design of the “Add Debris” page.



IDENTIFICATION
Project: DebriSat
Box #:

LOCATION
Source:
Section:
Related Foam Panel: ☐

TYPES
Debris Type: ● 2D ○ 3D
Balance Type:
Imager Type:

MEASUREMENTS
Material:
Shape:
Color:
Mass: g
Temp: °C
RH%: %
X_{DIM}: mm
Y_{DIM}: mm
Z_{DIM}: mm
L_C: mm
Volume: mm^3
Density: g/mm^3
ACSA: mm^2
AMR: mm^2/g

MISCELLANEOUS
Foam Attached: ☐
Intact Part: ☐

IMAGING
Frontlit Capture: Choose File No file chosen
Backlit Capture: Choose File No file chosen
Height Detection: Choose File No file chosen
Edge Detection: Choose File No file chosen
Ring Calibration: Choose File No file chosen
Debris Analysis: Choose File No file chosen
Point Cloud 2D: Choose File No file chosen

COMMENTS

Multiple Additions: ☐

Add Debris

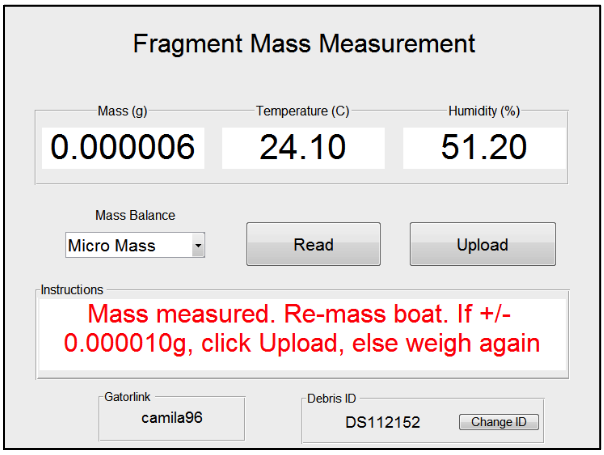
Figure 6. Screenshot of the "Add Debris" Page

The screenshot in Figure 6 shows how the page is broken down into seven distinct sections: identification, location, types, measurements, miscellaneous, imaging, and comments. The order of the sections is also significant as a user will naturally start at the top of the page and work their way through to the bottom. When the user is done entering data, the form is checked and validated and the user is automatically brought to another page that allows them to print out a barcode with the unique identification generated for the debris fragment and attach it to the bag with the debris fragment.

This design principle also applies to the rest of the DCS front-end web pages including the “Edit Debris” and “Verify Debris” pages. Each section of the web pages corresponds to a different step in the characterization process. When a debris entry is first created, only the identification and location sections are filled. As measurements are taken in the characterization process, the debris entry is edited and various fields in the measurements section are populated. Finally, once all data fields have been populated, each debris entry is verified and validated individually by another user.

3.4 Automated Characterization

In addition to the back-end layer and front-end layer, the DCS also interfaces with several external measurement systems used during the characterization process. These systems are designed to automate, as much as possible, the measurement steps of the characterization process. Each measurement system has its own GUI and interfaces directly with the DCS back-end database engine through the use of Open Database Connectivity (ODBC) connectors. Figure 7 shows a screenshot of the GUI for the external mass measurement system.



Fragment Mass Measurement
Mass (g): 0.000006
Temperature (C): 24.10
Humidity (%): 51.20
Mass Balance: Micro Mass
Instructions: Mass measured. Re-mass boat. If +/- 0.000010g, click Upload, else weigh again
Gatorlink: camila96
Debris ID: DS112152

Figure 7. Screenshot of the Mass Measurement GUI

Figure 8 shows a visual overview of the external mass measurement system interfaces with the DCS back-end.

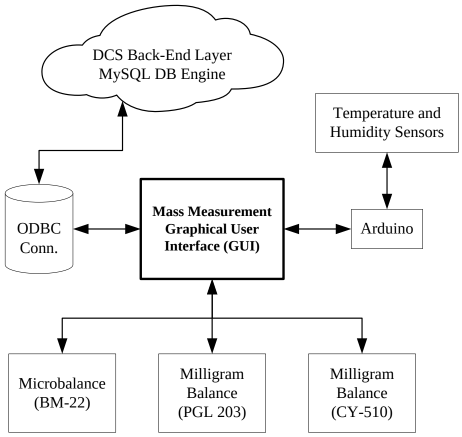


Figure 8. Mass Measurement System Interfaces

Similar GUIs and interfaces are used for the 2D and 3D imaging external measurement systems. Modifications made to debris entries by the external measurement systems are also validated before they are executed and they are formatted constituently with regular edits made on the DCS front-end user interface.

These external measurement systems aim to remove as much human error as possible in the measurement of the debris fragments. User simply place debris fragments in the measurement system, and the rest of the measurement process is automated once the user clicks “measure”.

3.5 Achieving the Four V’s of Big Data

The DCS is designed to ensure the DebriSat data set and analysis achieves suitable levels of *volume, variety, velocity*, and *veracity*.

3.5.1 Volume

DebriSat’s estimated 300,000+ debris fragments provide a significantly large data set to characterize and analyse. This data set includes debris fragments down to two millimeters in size. Furthermore, this data set includes a wide variety of diverse types of debris fragments ranging from flat CFRP slivers to large chunks of PCB or anodized aluminium. This large number of debris fragments and broad range of debris fragment types enables a good data set *volume*. The DCS provides a structure for this *volume* of data to be managed and stored. The estimated size of the final data set stored by the DCS is approximately 14 TB.

3.5.2 Variety

The DCS especially facilitates a high *variety* of data for the DebriSat data set. Each debris entry in the DCS includes a plethora of data types ranging from textual data to imagery data. In addition to measurement data, each debris fragment contains a unique identifier, version history, several timestamps, relational links to other data fields, and a multitude of other metadata fields used to create a mechanistic perspective on each debris fragment. Furthermore, as new measurement processes and procedures are added to the DebriSat characterization process, the *variety* of the data set increases. For example, the addition of a 3D imaging measurement system adds more than 126 new data fields and several new data types to the already diverse data set.

3.5.3 Velocity

The design of the DCS occurred in parallel with the design of many of the stages of the characterization process. As a result, the DCS is intrinsically designed to follow the characterization process; therefore, the main pages of the DCS front-end served to the user are streamlined to guide the user through the characterization process as efficiently as possible. This increases the *velocity* of the DebriSat data set generation/collection and analysis. Furthermore, the use of automated external measurement systems further removes the human delay from the characterization process and also increases the *velocity* of the analysis.

3.5.4 Veracity

Verification, validation, and therefore *veracity* was one of the guiding design principles of the DCS. Each of the processes and procedures used to characterize the DebriSat data set using the DCS are designed to ensure data is verified and validated before it is finalized. At each step of the characterization process, the forms used to enter data on the DCS are validated to ensure the data is formatted correctly and contains no errors. Additionally, once each debris entry has all its fields populated, it is verified and validated by an independent user before it is finalized. Once a debris entry is verified, it is locked and can no longer be edited to ensure the integrity of the finalized data. All of these verification and validation mechanisms add to the *veracity* of the DebriSat data set and analysis.

4 LESSONS LEARNED

Big data analysis is a powerful tool that, if used correctly, can provide insightful and useful conclusions that can inform important regulator decisions that can impact an industry for years to come. In order for big data analysis to be effective, it is essential to have data sets with high *volume*, *variety*, *velocity*, and *veracity*. There are many recent examples of how data sets in different industries lack a sufficient level of one or more of these “Four V’s of Big Data”. Insufficient *volume*, *variety*, *velocity*, or *veracity* can drastically inhibit the ability to make predictions or conclusions with an experiment’s data set.

In the orbital debris context, there is a lot of opportunity to leverage big data analytics to improve the understanding of the current climate; however, it is important that as these opportunities are explored the lessons learned from other experiments with insufficient *volume*, *variety*, *velocity* or *veracity* are heeded and used to avoid some of the pitfalls big data can introduce.

DebriSat and the DCS aim to leverage big data analytics to provide a useful data set to the orbital debris community to use to update and modernize current orbital breakup models. The DCS employs several design techniques such as form validation, a relational structure, and a verification procedure to ensure that the DebriSat data set is treated correctly with respect to the Four V’s of Big Data. The design and development of the DCS utilized this big data design mentality. It is essential to follow this design mentality as other orbital debris experiments are designed and developed. In this way, we can ensure that the data sets produced by these future experiments will be useful and achieve high levels of *volume, variety, velocity* and *veracity*.

5 REFERENCES

1. D.J. Kessler, N.L. Johnson, J.-C. Liou, M. Matney, The Kessler Syndrome: implications to future space operations, AAS 10-016, Advances in the Astronautical Sciences, vol. 137, American Astronautical Society, 2010, pp. 47–62.

2. D.M. Hogg, T.M. Cunningham, and W.M. Isbell, Final Report on the SOCIT Series of Hypervelocity Impact Tests, Report No. WL-TR-93-7025, Wright Laboratory, Armament Directorate, Wright-Patterson Air Force Base, Ohio, July 1993.

3. “Orbital Debris Environment: Detection and Monitoring,” in Limiting Future Collision Risk to Spacecraft: An Assessment of NASA's Meteoroid and Orbital Debris Programs, Washington, D.C.: National Academies Press, 2011, pp. 19–20.

4. J. C. Liou, N. Fitz-Coy, S. Clark, M. Werremeyer, T. Huynh, M. Sorge, M. Voelker, and O. J. Debrisat - a planned laboratory based satellite impact experiment for breakup fragment characterization. In L. Ouwehand, editor, Sixth European Conference on Space Debris. ESA Communications, August 2013a. ISBN 978-92-9221-287-2.

5. Rivero, M., Kleespies, J., Patankar, K., Fitz-Coy, N., Liou, J.-C., Sorge, M., Huynh, T., Opiela, J., Cowardin, H., and Krisko, P., "Characterization of Debris from the DebriSat Hypervelocity Test," Proceedings of the 66th International Astronautical Congress, IAC-15-A6.2.9x30343, Jerusalem, Israel, October 2015.

6. IBM, “The Four V's of Big Data”. Accessible: http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

7. A. Morley, B. Ramsingh, M. Dooris, A. Farrier and A. Hollingsworth, "Regional Responses: The Sustainable Food North West Research Collaboration", Food Research Collaboration, 2017.

8. Urban, M. C. et al. Improving the forecast for biodiversity under climate change. Science 353, pii: aad8466 (2016).

9. M. Asay, "Why your company's big data is not nearly fast enough - TechRepublic", TechRepublic, 2017. [Online]. Available: http://www.techrepublic.com/article/why-your-companys-big-data-is-not-nearly-fast-enough/. [Accessed: 06- Apr- 2017].

10. S. Sarsfield, "How a Small Data Error Becomes a Big Data Quality Problem", Data Informed, 2017. [Online]. Available: http://data-informed.com/how-a-small-data-error-becomes-a-big-data-quality-problem/. [Accessed: 06- Apr- 2017].