

Implementing an Enrollment Data Dictionary at George Mason University

A WHITE PAPER

Sharon R. O'Boyle
GEORGE MASON UNIVERSITY
MAY 2019

Table of Contents

- About the Author 2
- Executive Summary..... 3
- Introduction 3
- Statement of Problem..... 3
- Proposed Solution..... 4
 - Step 1. Get leadership buy-in..... 5
 - Step 2. Form a committee of representatives from all business units who deal with enrollment data.. 5
 - Step 3. Create a comprehensive list of data items. 5
 - Step 4. Define all data items 6
 - Step 5. Rename data items to be more descriptive and unique, where necessary. 7
 - Step 6. Create a data dictionary with the names, sources, limitations and detailed descriptions of the data elements. 7
 - Step 7. Get sign-off. 8
 - Step 8. Publish the data dictionary and make it available to all users of the data..... 8
 - Step 9. Educate users about the data dictionary..... 8
 - Step 10. Review and update on an ongoing basis 8
- Benefits 9
- Conclusion..... 9

About the Author

Sharon R. O'Boyle has over 20 years' experience in data analysis, information management and software development. She has been an Enrollment Management Data Analyst at George Mason University since 2012. She holds a B.A. in Mathematics with a Concentration in Computer and Information Systems from Mansfield University in Pennsylvania and an M.S. In Computer Science from S.U.N.Y. at Buffalo. She is currently pursuing a B.A. in English with a Concentration in Writing and Rhetoric from George Mason University.

Executive Summary

Within George Mason University there is inconsistency in enrollment reports released through different offices because the data elements used in enrollment reporting are not clearly and uniquely defined. This leads to frustration and lack of credibility in the data and prevents university leadership from making confident decisions based on that data. This paper proposes a solution that consists of creating a Data Dictionary that will clearly and uniquely define all data elements used in enrollment reporting. The Data Dictionary would then be used to create standardized enrollment reporting. The audience for this White Paper is all university members who deal with enrollment data and includes employees across all business offices, from top-level administrators down to staff involved in day-to-day data operations.

Introduction

During a recent presentation to employees of the Provost Office, Dr. Ángel Cabrera, President of George Mason University, stated that Mason enrolled the most transfer students of any of the Virginia public universities.¹ But exactly how many transfer students were enrolled at Mason at the end of the Fall 2018 term? This is a perfectly reasonable question and it seems like it would be a very easy question to answer by either the Registrar, the Admissions Office, or the Office of Institutional Effectiveness (OIRE). But things are not always as they seem. If this question was presented to all three of these offices, Dr. Cabrera would almost certainly get three different numbers. This is because the data elements used in enrollment reporting are not clearly and uniquely defined, leading to inconsistency in enrollment reports released through different offices. In this example, both “transfer” and “end-of-term” mean different things to different offices. This leads to frustration and uncertainty and has a negative impact on important decisions that university leadership must make.

This paper describes a 10-step process which would result in the creation of a Data Dictionary. That Data Dictionary would clearly and uniquely define all data elements used in enrollment reporting and then be used to create standardized enrollment reporting. The audience is all university members who deal with enrollment data.

Statement of Problem

Because the data elements used in enrollment reporting are not clearly and uniquely defined, there is inconsistency in enrollment reports released through different offices. For example, consider the “transfer” students in the above request. To the Admissions office, any student who has earned credits at another higher education institution before applying to Mason is

¹ Cabrera, Á. (2019, April 16). *President visit with Provost Office*. George Mason University. Fairfax, VA.

defined as transfer. But in order to satisfy State Council of Higher Education for Virginia (SCHEV) requirements, OIRE cannot define students who have ever attended Mason previously as a transfer. So if the Provost asks about transfers, which definition do we use? Similarly, there has been confusion over the expression “end of term.” For Student Accounts, the last day that a student can drop a class might be considered the end of term. For OIRE, it would be the day that they extract their final semester numbers. Other offices may consider it to be the last day of classes, the last day of finals, or graduation day. This lack of consistent definitions leads to confusion, hours of time spent trying to validate the data, and loss of credibility in the data. This problem is amplified because in universities, as in other businesses, data is often needed quickly in order to make a data-driven decision.²

Higher education institutions, including Mason, store data in multiple data management systems. Some such systems at Mason include Banner, Salesforce and the OIRE dataset. To confound the issue even more, some systems (ex. Banner) have real-time data, while other data sources contain data “snapshotted” at a specific time (ex. OIRE). Often each system is a stand-alone “silo” and does not integrate with the other systems. Out of frustration, some employees may resort to compiling their own data set (called a “shadow database”) which results in additional inconsistencies.³

Proposed Solution

The increased use of data processing and electronic data interchange heavily relies on accurate, reliable, controllable, and verifiable data recorded in databases. One of the prerequisites for a correct and proper use and interpretation of data is that both users and owners of data have a common understanding of the meaning and descriptive characteristics (e.g., representation) of that data.⁴

The International Organization for Standardization (ISO) makes the above contention as quoted in Jill Clark’s article “Managing a Data Dictionary.” The problem of inconsistent and confusing data names and definitions is widespread among organizations and is not unique to Mason. Institutions including the University of Notre Dame, Arizona State University and the University of Chicago are working on implementing programs to overcome problems such as these.⁵ One tool used in these efforts is the data dictionary and that is the solution being proposed here.

There are different types of data dictionaries. One type is a very technical document used for database developers and other IT professionals. That is not the type of data dictionary that will

² Powers, K., & Krist, P. (2018). Managing complexity and chaos. In K. Powers & A. Henderson (Ed.), *Cultivating a data culture in higher education*. doi: <https://doi-org.mutex.gmu.edu/10.4324/9781315171326>. p. 99.

³ Powers, p. 100-101 & 105.

⁴ Clark, J., Demster, B., & Solberg, C. (2012). Managing a data dictionary. *Journal of AHIMA*, 83(1), p. 48.

⁵ Chapple, M. (2013). Speaking the Same Language: Building a Data Governance Program for Institutional Impact. *EDUCAUSE Review*, 48(6), pp. 21&26.

be described here. This data dictionary will be a “higher level list of business terms and metrics” that needs to have enough detail to prevent any uncertainty about the meaning and use.⁶ It will be a comprehensive list that everyone involved in enrollment will understand. It will also place the data in the appropriate context and note any limitations. Consistency will then occur “when all data users are using the same set of standards and shared definitions.”⁷ A data dictionary is both a form of documentation and a form of communication that creates common ground among all business users. Everyone will then have the same vocabulary and the same definitions when referring to the data elements. It will solve the problem of inconsistent naming conventions, inconsistent data definitions and inconsistent data element values (ex. Male/Female vs M/F). “The data dictionary is the foundational document for maintaining the integrity of an organization’s data.”⁸ This paper will describe a 10-step process for creating and implementing a data dictionary at Mason.

Step 1. Get leadership buy-in.

Creating a data dictionary is no small task. Experience has shown that the average data element entry requires approximately ten person-hours to develop. And the persons putting in those hours will be coming from throughout the university. So the support of university leadership is crucial to ensure that sufficient resources are available.⁹

Step 2. Form a committee of representatives from all business units who deal with enrollment data.

Within each university office, one or more experts should be identified to work on the data dictionary “based on their knowledge and skills.”¹⁰ These individuals will work together as a team through the development, implementation and maintenance phases of the data dictionary.

Step 3. Create a comprehensive list of data items.

Developing the comprehensive data item list will be a team effort. Each business unit should review all its enrollment related reports and operations and record every descriptor (elements such as name, major, etc.) and metrics (quantitative data such as headcount, number of credit hours, etc.) that they use. At this stage no one is attempting to define any data items; identifying is the goal. The output for each team should be a table such as the following:

⁶ Anderson, C. (2018) Data Dictionary: a how to and best practices. Retrieved from <https://medium.com/@leapingllamas/data-dictionary-a-how-to-and-best-practices-a09a685dcd61>.

⁷ Powers, p. 103 &105.

⁸ Clark, p. 48 & 49.

⁹ Chapple, p. 24.

¹⁰ Atchison, E. (2018). Determining appropriate data sources. In K. Powers & A. Henderson (Ed.), *Cultivating a data culture in higher education*. doi: <https://doi-org.mutex.gmu.edu/10.4324/9781315171326>. p. 121.

Business Unit Name	Data Element Name	Descriptor or Metric?	Example(s)	Data Element Source	Reports Where Used
Enrollment Management	Market Group	Descriptor	VA, Northeast, International	MicroStrategy	Freshman Admits by Market
Enrollment Management	# Completed Apps	Metric	1, 2, 10, 100	MicroStrategy	Term Comparison Report

Table 1. Sample Table for Identifying Data Elements (Adapted from Anderson)¹¹

Step 4. Define all data items.

Once you have the comprehensive list of data terms, next is the key step of defining all the elements clearly, accurately and uniquely.

Start by locating and gathering all existing documentation. Then meet with each office separately and obtain as much information from them as possible, according to their areas of expertise. Be sure to get the exact calculations performed for each metric. Don't rely on the current definition of any data element. Instead, seek ideas about how the element should be defined.¹² Note any limitations or known data quality issues of the data element and include the source that will be the trusted source. See Figure 1 for a depiction of the process.

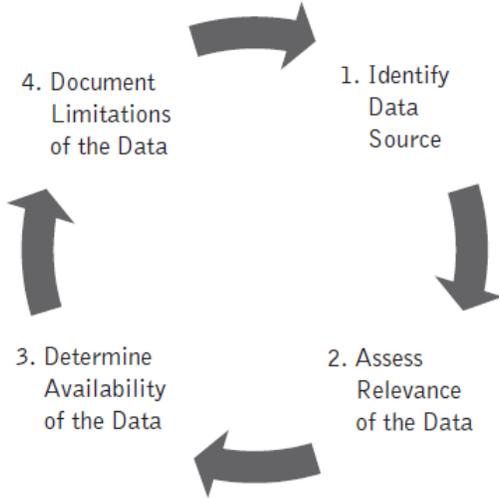


Figure 1. Process of Identifying Data Elements and Sources¹³

Do a review to make sure that the data will be accurate and reliable and then choose the source that will be most accurate, reliable and timely. This will help in the effort to develop confidence in the data. Determine relevance of the data. For example, multiple student addresses are available (ex. permanent address, campus address, mailing address). Which address(es) will be relevant to different reporting scenarios, and how would each instance of

¹¹ Anderson.
¹² Anderson.
¹³ Atchison, p. 116 & 120-121.

address best be labeled? For an element like address, the data dictionary should include how and when it is updated, so an indication of whether an element is dynamic or static should be included.¹⁴ It is critical to keep state and federal reporting requirements in mind during this process to ensure that all reporting needs will be met.¹⁵

Step 5. Rename data items to be more descriptive and unique, where necessary.

Once all the data elements have been defined, there will be elements that present confusion. This can be due to a number of reasons including:

- There are multiple elements that have the same name but mean different things.
- There are multiple elements that have different names but mean the same thing.
- Names seem to mean one thing, but actually mean something else.
- Names are so cryptic that users have no idea what they mean.

This will likely be one of the lengthier and most contentious steps in this process. Members of the team will undoubtedly feel very strongly about their office’s current naming conventions and usage. Debate and compromise will be needed until all team members either agree to use only one term and definition or to use multiple terms with names that will be modified to highlight the differences.¹⁶ For example: TRANSFER_EM (for Enrollment Management’s definition of transfer student) vs TRANSFER_OIRE (for OIRE’s definition).

Step 6. Create a data dictionary with the names, sources, limitations and detailed descriptions of the data elements.

The following pieces of information should be defined for each data element in the list:

Information Needed	Example 1	Example 2
Data Element Name	Market Group	# Completed Apps
Current Description	4-Category grouping based on residency code and address	# applications that have all supporting documents received
Data Source	Academic DataMart	Academic DataMart
Attribute or Metric	Attribute	Metric
If metric, how calculated?	N/A	Simple count
Dynamic or Static	Dynamic	Dynamic
If Dynamic, how often updated?	When residency code changes	As applications become complete
Known limitations or issues	The state will be captured as of date when attribute is created	None

Table 2. Sample Table for Capturing Information About Data Elements

¹⁴ Atchison, p 118-120.

¹⁵ Clark, p. 50.

¹⁶ Anderson.

Step 7. Get sign-off.

After the data dictionary is complete, have all the stakeholders review the document. Make any corrections or updates and then have everyone sign off on the document signaling that all offices are on board.¹⁷

Step 8. Publish the data dictionary and make it available to all users of the data.

The data dictionary will not be useful unless it is accessible. Publish the data dictionary in multiple formats, modes and/or locations to ensure that it will very visible and easy to access.¹⁸ If enrollment data users do not have easy access to the data dictionary, it will not accomplish its purpose.

Step 9. Educate users about the data dictionary.

In order for the dictionary to be useful, it is imperative that all users understand the meaning, source and limitations of any data item that they collect, use, analyze, or interact with at any level. So in addition to having ready access to the data dictionary itself, it is important that ongoing user training is available. This training should include:

- a thorough introductory module for users being newly introduced to the data
- a module that discusses any changes, and the reason and process for making those changes
- a module that stresses the actions that each user should be taking to ensure continued data integrity.¹⁹

Step 10. Review and update on an ongoing basis.

Defining the data elements and publishing the data dictionary is not just a onetime activity. Rather, the data dictionary should be viewed as a dynamic document. It must be maintained as part of a lifecycle process as data needs evolve in response to changes in the university's operations and strategies.²⁰ It is important that updates are made in a systematic and controlled process.

¹⁷ Anderson.

¹⁸ Anderson.

¹⁹ Clark, p. 50-51.

²⁰ Chapple, p. 22 & Clark, p. 49.

Benefits

Having standard, consistent data element names that everyone in the university involved in enrollment reporting agrees on and understands will bring benefits to all. University leadership, from the top levels of President and Provost down to faculty or administrators running an individual program within one of the colleges, will have access to data that they understand and have confidence in. Owners of the data in the various business units such as Admissions or OIRE will be able to explain their numbers and be spared the integrity challenges to their data and also the hours spent trying to answer questions and resolve discrepancies. And, most important of all, the entire university will benefit from the availability of valid, timely and reliable data that will result in improved data-driven business decisions.

Conclusion

Universities, including George Mason University, have a problem caused by a lack of definitions and consistency in the data that they use for enrollment reporting. A Data Dictionary that defines each data element and explains the source and limitations of each element can be very helpful in reducing confusion and frustration. The steps outlined above present a process for developing such a data dictionary, leading to more accurate, consistent and reliable enrollment reporting.