Data Stewardship Committee: Minutes of May 11, 2015

In attendance: Brian O'Connor (CAS), Leah Feroleto (Social Work), Greg Olsen (Data Analytics), Peter Elkin (Biomedical Informatics/SMBS), Kim Yousey-Elsener (Assessment/SL), Corey Hill (Student Accounts), Rachel Link (OIA), Gary Pacer (EAS), Craig Abbey (OIA), Kara Saunders (Registrar), Chris Connor (GEMS), Tom Okon (UBS), Sue Krystofiak (HR), Mark Molnar (OIA), Kelly Hayes McAlonie (CPG), Ashley Kravitz (Resource Planning).

Meeting called to order at 4:02 by Gary Pacer.

Gary provided an overview of the meeting's agenda, starting with an administrative update, the status of the report due on June 11 to the Data Governance Council, and reports from the sub-teams and working groups.

Craig and Gary would like to have completed information from each of the sub-teams and working groups by June 1 to allow some time for formatting and editing of the documents received in order to prepare a comprehensive package for the DGC. The topic of scheduling the next DSC meeting was also raised. Since classes have ended, the question was raised asking how often the DSC should meet over the summer, and whether the current 4 p.m. slot is still the best fit for the committee members. A Doodle poll will be sent out to members of the committee polling for the next meeting time, and discussion will commence later about the frequency and timing of summer meetings.

Craig stated that Jose Aviles, new Director of Admissions, will join the DSC as its newest member. Jose was not present at today's meeting but will join in at later meetings.

Gary spoke about the EAB data governance webinar series. The first in a series of four, this webinar focused on data governance excellence. Topics of interest included data governance for performance management, institutional ownership of data, bicameral data governance committees, fast-cycle decision frameworks, and data dictionaries. Gary presented an overview of the slides from the session, and a pdf of these slides has been uploaded to the Data Stewardship Committee's website for committee members to access. Links to UNLV's data dictionaries will also be shared with the committee: these data dictionaries were shared during EAB's webinar.

Gary's feeling after participating in the webinar was that there are some areas within data governance where we at UB are doing well and other areas where we could improve. The next webinar in the series will be on May 21, and Gary encouraged committee members to participate.

Peter Elkin mentioned ISO TC37 standards, which cover standardization of terminology, as a way to achieve consistency within data definitions. Peter felt that these standards might help the working teams when creating definitions.

Working teams presented their updates next. Craig presented for team 1, Structure/Roles/Responsibilities. The Team has discussed the possibility of a Chief Data Officer role, and found that Stanford is now hiring a Data Stewardship Coordinator role. The position description has been posted to the DSC website, under Working Team 1 documents, and will be sent to the committee as well. The team has been working on the recommendations draft and making the data governance process sustainable, and has been looser on definitions formatting as they get ideas into place. The draft data definitions flowchart was presented, and has been uploaded to the DSC website as well. Craig suggested a vote on adopting this flowchart and process at the next meeting.

In terms of definitions, data trustees and stewards are defined, and data manager and data user have been added. Kara asked if the team has considered whether these definitions might add to workload for staff members that are in those roles. Craig stated that this definition is intended to identify the role, because people are already doing these tasks. However, if the institution sees a need for additional staff, they will then need to identify the resources for those lines. Craig also stated that the team still needs to work with the CIO in order to figure out data classification schemes, the process for creating data dictionaries, and access to this information. Additionally, there is a need for information about a strategic plan for data warehousing and business intelligence, which will be handled by Tom Okon, and data quality questions and reporting.

Craig next passed around some data grouping visualizations prepared by Stanford. Stanford provided a calendar map that lists all academic and unit calendars, along with payroll dates and other information. They also shared faculty groupings, degree information, and student maps. These visualizations have been posted to the DSC website under Working Team 1 documents and will also be shared with committee members via email.

Team 2 presented next, with Kara providing the updated. This has been an iterative process for them. They are using and excel table document prepared by Mark to share definitions. The team has identified their highest priority items to complete by June 1 and feel they can meet this deadline. They have drafts of many items and just need to establish consensus at this point.

Team 3's update was presented by Sue Kryzstofiak. The team has started to look at other definitions now. They started with the SUNY Board of Trustees (BOT) definitions, and have moved on to adding IPDS as well. The matrix is being updated and has been posted to the DSC website under Working Team 2's documents, and a link sent out to committee members. Peter Elkin mentioned the qualified academic rank terminology being confusing to some, and stated the importance of using the term "qualified" only in this definition to avoid confusion in other areas. He has seen the use of the term "unqualified academic rank" internally, and suggested a check on this usage and its inclusion if necessary in the definitions.

The work remaining for Team 3 at this point consists of titles: for example, student assistant versus TA/GA versus Research Foundation titles. They recommend ensuring consistency among these, and on the term "student" to make sure that the definitions hold across domains.

Craig presented for the Financial Aid working team since John Gottardy could not attend. The team is working on definitions for Pell and debt at graduation now, as well as coding for financial aid types to ensure that they are coded appropriately as state, federal, or institutional aid. Some aid may come from one source and pass through the institution, but coding must be checked to ensure it is done correctly. The team is working on an 11 year picture of financial aid to get a better idea of unmet need and where to better target financial aid.

Corey Hill presented for Student Accounts. He and Laura Stevens put together a list of terms, and found many of these from other modules. They will reach out to those in other areas to ensure mutual understanding of definitions. They are using the excel table spreadsheet Mark developed to track their definitions. Mark will add official definitions to some areas, in particular for tuition revenue, as there is overlap here. These tuition revenue definitions will then help inform other definitions.

Greg Olsen presented for Undergraduate Admissions. The group met for the first time last week and came up with terms to define. They will rank and prioritize these in order to complete their tasks by June 1.

Chris Connor presented for Graduate Admissions. They met last week for the first time. While there is consistency in terminology and usage of terms in GrAdmit, this is not always so in external applications received by units. GEMS has a data dictionary and other documentation, which the team will go over later this week.

Space's report was presented by Craig. Many definitions come from SUNY or the federal government, and they will document and organize these definitions as a team. Craig also presented for Bibliometrics. Austin will work on the definitions for data from Web of Science and Scopus, while Craig will provide information from Academic Analytics. Craig will also meet with Beth Corry in Business Services later this week to discuss definitions from that area.

Gary opened up discussion by asking the committee how easy they found it to do definitions. Mark stated the need to keep definitions simple enough for quick understanding by senior leadership, while still containing the "meat" of the information present. This is an iterative process and understanding definitions and different categories provides an explanation of how different numbers may be calculated. Sue mentioned that team 2 found that some terms, like GFT, are not covered in the official SUNY definitions, so additional work needed to be done for that term. She also encouraged anyone with questions about faculty or staff definitions contact her. Greg mentioned that the process has been a good way to gain perspective and learn what others are doing. Kara hoped that information from the final report will help inform areas as well. Tom stated that the process could help identify inconsistencies or areas where more information is needed. To expound on that, Craig said that data integrity comes next. We must know what we are measuring before we can determine if it is inaccurate.

Gary next asked the group about tools used in the process. Excel seems to be useful for working groups to house data definition drafts at this point. Brian asked what other institutions use. Craig mentioned the UNLV data dictionary next, and how the terms link back to the concepts. Gary mentioned Data

Cookbook, which already houses Common Data Set and IPEDS definitions, but for UB to use it would charge a fee, and not many use this. Stanford uses Colibri now.

Craig wrapped up the meeting by stating that the DGC wants to see progress at the June meeting. Brian asked if the 3 to 5 definitions were the limit to send for each team. Craig replied to send as many as you can, recognizing that this is an iterative process, and those definitions will be needed later. The DGC is aware that this is the tip of the iceberg in terms of definitions. The easiest way to provide definitions might be to use SUNY BOT definitions, SUNY_IR definitions, or IPEDS definitions. It is harder to make attributes to units from these, or to show them from a strategic/operational/official standpoint. Gary also mentioned that an evaluation of the process would be helpful to communicate to the DGC.

Meeting adjourned at 4:57. Next meeting TBA, will be scheduled by Doodle Poll.