

Data Stewardship Committee: Minutes of March 16, 2015

In attendance: Brian O'Conner (CAS), Greg Olsen (Enrollment Management), Leah Feroletto (SW), Troy Joseph (GEMS), David Love (SEAS), Kara Saunders (Registrar), Gary Pacer (EAS), Craig Abbey (OIA), Tom Okon (UBS), Sue Krzystofiak (HR), Michael Koziej (Campus Living), John Gottardy (FA), Rachel Link (OIA)

Meeting called to order at 4:01 p.m. by Craig Abbey.

Gary went over the agenda, starting with a demonstration of the committee's website. Gary showed the group where various information and files are stored, and where working group materials are posted. Kara asked what sort of materials should be posted to working group sub-areas. Gary replied that while minutes are not necessary, materials worthy of being shared should be posted, especially as the committee moves toward standardizing definitions. This should include any documentation or materials that could be useful to the DSC as a whole. Craig suggested the addition of an email alert feature to the website that would alert committee members whenever a file was uploaded or updated.

Gary mentioned that he and Craig were to meet with the Data Governance Council on February 26, but this meeting was pushed back to April 2. Gary feels that the working groups will have the ability to share material at that point, and that the DGC will provide feedback to the DSC as well.

Working teams presented their updates next. Kara Saunders presented for Working Team 2 - Enrollment and Degrees. The team has had a few meetings so far. The diagram of primary data relationships (available on the DSC site under Working Team 2's section) was again reviewed, showing how operational, strategic, and official data definitions are related. The team wants to hone in on the types of definitions used, realizing that definitions vary by situation. Craig noted that the data definitions of operation, strategic, and official make sense from a student standpoint, and asked the group if they made sense from an HR perspective. Craig also noted that counts for numbers of students vary depending on funding types, enrollment status, and other values. When counting staff members and using payroll data, we should remember that payroll data is operational data, and for IPEDS counting, SUNY may transform some HR data. Tom Okon stated that the three data definition types appear to match with SIRI data views and that HR data would seem to be captured by this framework.

Kara stated the next goal for the working team is to work on defining data elements, like term or session, semester, student (with subset categories like new, returning, continuing, matriculated/non-matriculated, and level), academic major, plan, program, GPA/QPA, headcount, and full or part time status. They have started with defining student and major for now, and will be discussing them in the working team meeting scheduled for later this week. Gary asked how the team will begin with a student definition, and Kara replied that one source is the official SUNY definition. At Purdue, the method used was to have all individuals contribute a definition, and then look for commonalities and propose a

solution. Official definitions should come from the outside, while strategic and operational definitions should come from internal sources.

Rachel Link presented Working Team 1's research on data governance at other AAU institutions. The team looked at a total of 15 AAUDE institutions to identify which have established a data governance structure, what that structure looks like, and what responsibilities lie with the committee or office responsible.

13 of the 15 institutions researched have some sort of data governance structure, and the remaining two are in the process of implementing it. The most common reporting structure, found at 10 institutions, is to the Chief Information Officer or equivalent. Other reporting areas include Finance, Institutional Analysis, Provost, or Enrollment Management.

Data management policies are generally the responsibility of Data Governance Committees, and are usually publically available. For 10 of the 11 institutions that had data definitions available, data stewards, trustees, custodians and/or users were defined, and these definitions were consistent with the definitions that the DSC has used so far. Data access policies are also commonly specified, and generally center on the classification level of the data, and how access is granted along those lines.

Data governance websites were also examined, and range from minimal to extensive. Typically, at least some information is behind a firewall or otherwise restricted from public view. Websites tend to be housed on the CIO's site. Almost all institutions maintain extensive data dictionaries and explanations of terminology. Definitions vary only slightly between institutions. Several institutions provide guidance on writing data definitions.

Two model institutions were also briefly presented: the University of Washington (with whom Craig and Gary have spoken about data governance) and Stanford. Washington has heavily invested in data governance and data quality and provides an extensive site documenting the process. Their Data Management Committee was established by the provost. The charge laid out the need for a consistent approach for data governance; defining clear and commonly used structures, models, definitions and process, and to provide recommendations and guidance concerning institutional data, especially regarding access, strategic planning, quality, and security. The goals were to come up with university-wide data management policies, standards, guidelines, and procedures; provide input on strategic planning as it related to data management, recommend controls for assessing data management and value, and to coordinate compliance efforts. The Office of Information management sets forth these standards. UW also defines operational, tactical, and strategic data – the only other institutions researched that did this type of categorization. Data Management is seen as an ongoing concern and the Data Management Committee (comparable to our Data Governance Committee) is an ongoing one. The Standards Implementation Committee is comparable to our DSC, and members are accountable to the DMC. This group is tasked with reviewing and recording standards and procedures, and tracking and recording revisions. They also can assist in implementation and enforcement of procedures, standards, and guidelines.

Stanford began data governance in 2011. Data governance supports the Business Intelligence competency center and creation of BI dashboards. Most of the data governance activities fall under Stanford's Data Stewardship group, and there are four active subgroups (finance, HR, research and sponsored projects, and student) within that. Stanford has extensive documentation on the makeup of their groups, meeting schedules, definitions, best-practices, and policy initiatives. Themes include supporting consistency across efforts, encouraging volunteer participation, maintaining motivation, and increasing awareness and engagement at all levels of the university community. Stanford's dedicated DG role is based in and reports to the Office of Institutional Research & Decision Support. Other IRDS staff members are involved in many of Stanford's subject area-specific data stewardship efforts

Stanford has established a data governance maturity model which attempts to guide the process of data governance. Each project is broken down into people (who is responsible, what roles do they have), policies (definitions, implementation, and enforcement), and capabilities (structure and support). These are then broken out into foundational stages, beginning with awareness, moving towards formalization, and resulting in metadata creation. Maturity levels help identify where in the process each project is at, and provides guidance for the suggested next steps. Data definitions are also included, and Stanford stresses the need for high quality and reusable definitions. Interestingly, Stanford also hosts a Confluence site for users to participate in to add information or documentation, along with a WordPress blog site about data governance issues.

Chris Connor mentioned that staff are needed to ramp up the data governance effort, probably something like the Student Services effort, in order to speed up the process. Data needs to be moved into a warehouse, and auditing needs to occur to ensure data validity. Ongoing work will need to be assigned to a dedicated staff member and/or office to ensure timely completion.

Craig provided a brief rundown of what the next steps for the working team will be. The first is to address the lack of training/onboarding. Currently, no data dictionary exists and there is no documentation for report construction. Confluence could be used to build documentation and dictionaries. Next, we have no policy yet on how individuals are appointed to the DGC and what the terms should be, and this should be a priority. Finally, the team will work on identifying data warehousing and BI policy.

Tom and Kara next mentioned a need for a student view in SIRI. Jon Havey used to provide support for this, but since he took a new position, this has not yet been reassigned. Kara mentioned the need for one office to "own" SIRI at all levels. HUB governance used to provide this, but the reorganization after the implementation meant that no one office fills that role currently. The business side of SIRI has groups that manage reporting and run their own areas, but there is a lack of this on the student side. A part of what we as an institution face that is a challenge for data governance is that there is a lack of coordination, research constrains, and lack of focus, which impeded efforts. Sue stated that hearing about data issues is helpful to understand the challenges that lie ahead.

Troy pointed out that student data is different than what is in SIRI. The reports in SIRI are canned, but GEMS is able to provide more specific reporting. Tom said that the capabilities of SIRI are greater than

many know, and the operational needs have been the primary focus of SIRI efforts. Troy stated that for many who use student data, it is easier to go to Info Source and run a query than to use SIRI. Tom mentioned that back-end data modeling is happening on other views that would need to happen on a student view to make some requests possible.

Craig said that a goal should be to have “top level” numbers in one place, like a master table of facts that could be used as a starting point for people to agree on. Brian pointed out that issues with text fields like name where additional spaces might occur make lookups very difficult and time-consuming. Tom gave examples of how this could occur: a student applies for two jobs but uses a different name on each. It is not uncommon for international students to use an Americanized name here and not their official name, but it makes matching records difficult. Creating a solid definition of what can be used in the various name fields could help with this.

Sue presented for Working Team 3, and distributed draft definitions from their meetings. They have started with academic rank and qualified academic ranks. Academic Rank uses the titles of Instructor, Assistant Professor/Librarian, Associate Professor/Librarian, Professor, or Librarian. These are tenure track. Qualified Academic Ranks are conferred with the title of Lecturer or the Academic Rank titles prefaced by clinical, research, visiting, or adjunct, and these are non-tenure track lines. Sue stated that clock-stoppage is something the team is examining now as when the tenure clock stops, titles may change from Academic to Qualified Academic Rank.

A definition for GFT (Geographic Full Time) was also provided: these faculty are persons serving on the faculty of a medical center who are not employed on a FT basis for the purposes of fixing compensation payable by the State, but a;; of whose professional services and activities are conducted at the medical center or its affiliated hospitals and are available to the State on a full-time basis for clinical and instructional purposes.

Sue also displayed a matrix created by the team and still in progress that will place titles into the categories of academic rank, qualified academic rank, tenured/tenure track, non-tenured, ladder faculty, non ladder faculty, and voting faculty. The hope is that providing this matrix will allow for easier calculation of what faculty are counted in these groupings, and potentially create an Info Source lookup table for all to use. Emeriti have thus far not appeared in their definitions but will be explored in the future, as will TA and GAs. EOC definitions will also be created. The team has not yet addressed the issue of full or part time status. The status may vary by title as well: there is a full time status for benefits purposes that may not match full time status for voting purposes. David pointed out that PI eligibility for ABET reporting requires certain titles and that individuals appointed as lecturers cannot be PIs for this reporting.

Craig stated that Kelly Hayes-McAlonie will lead up the effort on providing space definitions for the data dictionary/ These should be close to what will be needed as there are federal guidelines that must be followed. Craig described bibliometric definitions and sourcing, and why the counts may vary depending on the timing and sources, like Google Scholar, Web of Science, Scopus, or Academic Analytics. Each counts slightly differently, and librarians could provide some guidelines on this effort.

Financial aid definitions will be needed, and assistance is requested from the Office of Financial Aid for this, especially when it comes to calculating student debt.

Gary stated that we are making progress to the June deadline, and that he and Craig will be having a discussion with the DGC on their progress. The recursive nature of definitions and a lack of existing documentation mean that extensive research and work are involved. Sue stated that the need for documentation now can help guide future efforts as well.

The next meeting will be March 30 at 4 p.m. in 107 Capen. Student definitions will be available by that point.

Meeting adjourned at 5:03 p.m.