Data Stewardship Committee: Minutes of September 21, 2015

In attendance: Craig Abbey (OIA), Gary Pacer (EAS), David Love (SEAS), Mark Molnar (OIA), John Gottardy (Financial Aid), Kara Saunders (Registrar), Brian O’Connor, (CAS), Ashley Kravitz (Resource Planning), Rachel Link (OIA), Greg Olsen (Data Analytics), Beth Corry (Business Services), Nigel Marriner (Registrar), Tom Okon (Business Services), Jose Aviles (Admissions), Alex Nelson (GEMS).

Meeting called to order at 4:03 p.m. by Gary Pacer.

Gary discussed the meeting he and Craig had with the Data Governance Council (DGC) on September 10. Part of the DGC’s ask was for the creation of the process to create, return, and approve definitions. The DGC approved the process as presented by the DSC with one caveat: if definitions are approved at the DSC level, the DGC does not need to see them for review. If definition disputes occur at the DSC level, the definitions should be escalated to the DGC for clarification. The result of this decision means that the schematic that was designed by Craig and Gary and previously presented to the DSC at the August meeting will need to be updated. Craig stated that most of the groups have a good handle on definitions and most of the definitional conflicts are being hashed out at the group level.

John Gottardy provided some examples of item type scrubbing in Financial Aid. As they have worked through the various item types used in Financial Aid, they have identified discrepancies and coding issues in previous time periods. John asked if the preference is to alter old data towards a new understanding or to provide data mapping to explain how to convert old values to new ones. Craig indicated that this example is similar to what has happened with enrollment in the past: the School of Informatics closed, the College of Arts and Sciences was recast – and led to discussions about how to accurately provide trend analysis. In terms of Financial Aid information, however, many of the historical numbers have been reported to outside agencies like NYSED and IPEDS, although we should correct things that are in error. John mentioned that an example of inconsistency was how “disbursed aid” was calculated – whether it meant all money passing through versus other ways of calculating disbursements, and wants to try to come to an understanding of how to calculate disbursement. Grants and scholarships for athletics should all be “scholarships”, even if they are called “grants in aid” within the system. John indicated that the data could be scrubbed, but that he wants to map the old information to the correct definitions to allow for historical trend analysis. Craig suggested the John work with Mike Randall in OIA to work on this project further. Gary asked the question: once definitions are approved, when do we cleanse the data? We have not yet discussed this topic, but should consider it. John said that Financial Aid would not correct prior submissions to NYSED or IPEDS, but wants to include explanations for potential discrepancies. Mark suggested that the Hub implementation process could provide some ideas: footnote that business processes have changed and move forward.

Craig next spoke about Data Cookbook. This is a hosted solution for metadata repository. It is used by numerous institutions and is designed for higher education, and is the most common tool used for this
purpose by AAU institutions. We can show our definitions to other institutions and borrow from other institutions’ work: for example, if we find that another institution has defined a term in a way that works for us, it could serve as a starting point for us to develop a definition. Data Cookbook also includes all IPEDS and Common Data Set definitions as well. Numerous OIA staff as well as David and Gary attended a webinar demonstrating the product earlier this summer, and the decision was made to move forward with the recommendation that this product be purchased. Other competing external products were too large for UB’s needs. We have received approval to purchase Data Cookbook for data governance. Currently, trying to decide on the level of support needed: there are multiple tiers, and include workflow processes and training modules. More information on Data Cookbook will be presented at the next DSC meeting.

Gary asked for suggestions for the definition approval process. Should each group pick a definition for review? Beth asked if other groups used the spreadsheet template and if they found it easy to use. The Financial workgroup has uploaded theirs to the shared folder and includes action items for each meeting. Perhaps a format like this could be used for review: the document with definitions as approved by the groups could be uploaded to a common shared area for feedback. Mark asked about a voting process for definitions, and Craig replied that although we have attempted to make the process transparent, has anyone been missed or not included? Maybe the definitions could be sent to a wider audience to allow further representation and feedback. Beth suggested that financial definitions could be sent to the CFO group for review. Kara indicated that this is supposed to be an iterative process, and we expect some changes along the way. Craig’s concern is that we could potentially force definitions, and that groups may not have thought of potential issues. Integration with the Data Cookbook process could be considered for definition approval as well, since there are places for various levels of approval built in to the product. Mark asked how other institutions handle the approval process, and Craig replied that the process seems to vary by institution, although the Data Cookbook process is used by some. Craig suggested that some definitions could be brought forward for discussion at the next meeting, and documentation collected into one place. Gary stated that the DSC website would be the place to store documentation for this effort. However, the Human Resources files will need to be updated with Mark’s succession into his role. Rachel will reach out to the group coordinators individually in the next week and obtain current definition lists to post to the DSC site.

David asked about the initial load of definitions into Data Cookbook: would it be a batch process and then individual loads after the initial phase? The HUB process started with a small group pushing out data and then shrank to one person updating and maintaining the process after the bulk of the work was done. Should individuals be added to working groups now for definition review? Brian stated that his group has been so deeply immersed in the work for some time that it would be difficult to quickly bring a new person up to speed. Mark felt that the groups should stay as they are for now, citing similar experiences to Brian’s, and stated that anyone who would join now might have a lot of catching up to do with interpretation or terminology. Mark suggested that in the future, subject matter experts (SME’s) could be added to the groups, and could be asked for input and ways to improve the process. The cross-modular HUB group meets regularly and could serve as a model for the definition approval process. This issue is tied in with a larger data governance picture. Involving more experts provides
intelligent feedback and makes the process richer for all. Gary’s concern is that this could seem like a daunting process for the groups that have hundreds of definitions, and suggested that perhaps we could just start with key definitions or ones with areas of contention. Craig stated that we could gather the definitions first. Depending on how long it takes to get Data Cookbook up and running and use its built-in workflows, we can either start there or reconsider if it will take some time to get Data Cookbook ready for use.

Gary asked where the groups are now, and where they are headed. Beth spoke for Financial Services. The group has broken down terms into areas, like financial, procurement, etc. for subgroups to review. The procurement subgroup is going through these and then sending those up to the financial group for review. Overall, there are about 300 terms. They are covering 20 terms per meeting. It is going well, although the process is slow. One member is comparing terms to SIRI, and another to forms on the website for compliance, and noting what will need to be updated. As the Resource Planning dashboards are launched, they will also be checking and correcting terms.

Gary spoke for Human Resources. Mark Coldren is catching up now and has met with Sue Kryzstofiak, but the group has not yet gotten together. They will need to reconvene soon. 50-60 key definitions have been finished and are ready for the review process, but the group is not done yet.

Mark spoke for Enrollment. They have identified more than 70 dimensions for 15 critical report definitions. Although there are many reports out there, only a handful of reports are used extensively—these contain terms like headcount, credit hours by department, and the like. They hope to make progress in the next two weeks to bring these definitions to the larger group for review.

John spoke for Financial Aid. The biggest issues they have identified were the identification of item type discrepancies, which he spoke about earlier in today’s meeting.

Jose spoke for Undergraduate Admissions. Due to staff changes, they have not yet met, although they plan to soon. The responsibility shifts in the office means that he will be involving Danielle Ianni in the group as well. Greg stated that they have key items in order and had found common ground as well as crossover on many definitions.

Craig spoke for Space. They had more than 400 definitions well-defined due to legal requirements on terminology, and should be complete.

Mark spoke for Research. They have a base compendium of definitions to start from. Beth indicated there is likely to be crossover with financial definitions, and will send those to Mark for review. Mark also spoke for Student Accounts. They have not met in a month, but the group has some definitions done. Entity-Hierarchy is basically complete at this point.

Gary concluded the meeting by stating that meeting invitations have been sent out through June 2016. The Monday at 4 p.m. time seemed to work for the majority of the group. The next meeting will be October 19 at 4 p.m. in Capen 567.
Future meeting dates are below – all will be held in Capen 567 from 4p.m.to 5 p.m.

- 10/19/15
- 11/23/15
- 12/21/15
- 1/25/16
- 2/22/16
- 3/21/16
- 4/18/16
- 5/23/16
- 6/20/16

Meeting adjourned at 4:52.