Data Stewardship Committee: Minutes of April 18, 2016

In attendance: Brian O'Connor (CAS), Mark Coldren (HR), David Love (SEAS), Greg Olsen (DAT), Mychal Ostuni (GEMS), Rachel Link (OIA), Beth Corry (Financial Services), Ashley Kravitz (RP), Nigel Marriner (Registrar), Joe Mantione (OIA), Leah Feroleto (SW), Mark Molnar (OIA), Michele Sedor (OIA), Gary Pacer (EAS), Craig Abbey (OIA).

Meeting called to order at 4:01 by Gary. Gary reviewed the agenda for today, which was:

- Data Cookbook Implementation
- Business Intelligence Task Force/ Oracle/Ciber Insight
- Higher Education Data Warehouse (HEDW) Conference
- Review and approve financial terms
- Ethnicity discussion

Michele provided an update on the Data Cookbook implementation. Initital content training has been completed, and workflow training will be completed this week. Trainings are being recorded, and will be available online soon. Approximately 20 people have completed the training program so far. The implementation team is now working on integrating Data Cookbook with OBIEE and Tableau. Michele has contacted other institutions to see how they have done this and what we could learn from their experiences. A Data Cookbook web link is available off of the Data Governance website. Access to the documentation repository here is currently limited to those who have taken training as well as the implementation team. Once the Shibboleth authentication is available for Data Cookbook, individuals will be able to log in using their UBIT and associated password, rather than the Data Cookbook-provided ones.

Gary spoke about updated to the Oracle/Ciber visit in February. The team returned on April 7 to meet with Gary, and will be on campus next week to meet with the Data Governance Council. Gary felt that the Oracle team was well-prepared for their visit and had collected a lot of information. They did recommend a data warehouse, which UB is not quite ready for yet. The Oracle team also met with Peter Elkin, Kara Saunders, Lee Melvin, and Tom Okon. Their report is targeted for review on April 29. The hope is that we obtain something useful to use as a blueprint for future activities.

Gary next spoke about the HEDW conference that was recently held in Rochester. Attendees at this annual even included Stony Brook, Washington, Purdue, Cornell, U of Rochester, Arizona, and many other institutions. The major themes of the conference were: Provide better data and provide data better, which leads to data quality and data accessibility; and Spread the word – get data out there. Gary shared UMBC's Rex site, which is similar to SIRI here at UB – it is a reporting portal site available to campus members, featuring canned and ad-hoc reporting. This was presented as an example of self-service business intelligence, and a potential idea for what BI could look like at UB.

The top twelve issues for higher education as identified by HEDW were:

1. Data governance
2. Data quality
3. Student success
4. Data visualization (Gary noted many institutions use Tableau)
5. Meta data and data definitions (Data Cookbook is gaining users)
6. Self service
7. Predictive analytics (Important, but what to do with the results?)
8. BI strategy
9. Data informed culture
10. Standard reports and dashboards
11. Integration
12. BI program governance

University of Maryland Baltimore County presented a session discussing an Educause tool to allow campuses to measure BI progress relative to peers, and a matrix that displays maturity and deployment. Additionally, they can show an analytics maturity dimension score. Gary stated that we may wish to look into this and see if there would be benefits to participating in this project.

Cornell presented a session on the role of Chief Data Officer. At Cornell, this role leads the planning, development, and delivery of innovative strategy, effective solutions, and effective operational process that set a high standard of excellence for broadly conceived data architecture, BI, and data analytics solutions. Cornell's current data landscape has a structured and administrative focus. They are noteworthy in higher education for their work in data warehousing, but Gary felt UB is close to them.

Notre Dame presented a session on data access. The key point made in the session was to presume good intent; the onus was on the data steward to prove why people cannot have access to data. More schools are starting to talk about this type of access. In this model, access is restricted only when there is clear risk of significant harm to the university. Craig pointed out that this policy is a bit unusual in light of many private AAU institutions not opting to share much data with AAUDE (the AAU Data Exchange). Gary stated that institutions with this model do provide appropriate training and certification, and that the access applies to both data and reports. Craig pointed out the contrast between this approach and UB: at UB, access is provided but training is not. People can get confused about what tables or variables to use.

Craig spoke about a session at the recent AAUDE conference on the topic of data governance. UB has a good handle on entity hierarchy, and other institutions have nothing like this. They spend huge amounts of time combining and connecting data. While there are ways to improve, we are doing a good job on that front. Many others at the AAUDE conference were interested in the topic of data governance.

Gary opened the floor to questions about the financial definitions. Beth reviewed the process. The team began by reviewing financial reports for terms, and used the financial glossary as well to build a list of starting terms. Research Foundation, Resource Planning, Financial Services, and academic support units were all represented on the committee. Definitions were complete as of one months ago, but additional requests may have added information. The comments field on the spreadsheet previously distributed links to that additional information. Beth pointed out that the procurement and travel systems will be updating to new systems, so their terms may change.

Gary asked about NACUBO terminology. These terms tie to account numbers and are used for specific utilization, and are all expense related. Beth stated that staff are being trained on the proper use of accounting codes. Mark asked if there are efforts to consolidate accounts or sunset them, and Beth stated that as people wish to do more finite reporting, they ask for more account numbers. The importance of linking account numbers to entities to ensure appropriate reporting was stressed. The financial terms were accepted and approved.

A brief discussion on Ethnicity coding was next. There is inconsistent reporting of ethnicities across multiple InfoSource views. Ethnicity is derived in two ways, depending on if a primary ethnicity code was flagged or not. SUNY_IR views use the IPEDS definitions of ethnicity. Mark asked the group how to resolve this, and how we can ensure consistency across faculty and staff data too. HUB support will encourage the use of the primary indicator.

Nigel asked if the system wants only a single row, and stated that it seems dangerous to suggest a single race to a multi race student. Michele stated we can use the two or more indicator in these fields. Mark mentioned issues with HR data, which is self-reported, as it needs to be: it should not be reported by looking at someone and determining their ethnicity. There are also issues when a person reports a race or ethnicity on their UB employment application, but fails to report it once hired. Should that application material be imported? Craig stated the old system would not allow for 2+ races, and would pick the minority race if present. In EDI, which is different than IPEDS, there are race and ethnicity, and internationals may have races as well. Gary stated that HR data on ethnicity comes from multiple sources, and UBF is not reporting race/ethnicity consistently. Mark thought that self-reported data should be collected annually, and an educational campaign indicating the usefulness of these data might help with reporting rates.

Mark asked if a group to explore using the EDI data in SUNY_IR could be created. Greg stated that it is important to note that the ethnicity calculation algorithms in the application process differ from the later parts of the student process. Different people impact pulls by collecting information from various places. Brian mentioned that a person could be a student and later and employee, leading to potentially different information in multiple places instead of a centrally located database. Identifying people with flags as “student” or “employee” might help. A recommendation was made to talk to the HUB support groups to see what they recommend. They might wish to audit all tables with ethnicity and figure out where it comes from and how it is derived.

Meeting adjourned at 4:50 p.m. Next meeting scheduled for **May 23, 2016** in **Capen 567.**