Data Stewardship Committee: Minutes of January 25, 2016

In attendance: Kelly Hayes McAlonie (Capital Planning), Mychal Ostoni (GEMS), Leah Feroleto (Social Work), David Love (SEAS), Brian O'Connor (CAS), Rachel Link (OIA), Gary Pacer (EAS), Craig Abbey (OIA), Mark Molnar (OIA), Michele Sedor (OIA), Nigel Marriner (Registrar) Kim Yousey-Elsener (Student Life), Beth Corry (Financial Services), Mark Coldren (HR), Jose Aviles (UG Admissions), Shirley Walker (Student Accounts), Tom Okon (Business Services), Joe Mantione (OIA), John Gottardy (Financial Aid), Ashley Kravitz (Resource Planning), Charles Zukoski (Provost).

Meeting called to order at 10:29 by Craig Abbey. Craig introduced Dr. Zukoski, UB's Provost, to the group, and reviewed the meeting's agenda:

- Introductions
- Data Stewardship Committee and Data Governance Background
- Reports from Working Teams
- Next Steps

Craig asked each attendee to introduce themselves and reviewed the year's agenda, which was as follows:

- Data Governance Council has approved the data definition process
- Next step: establish a metadata repository – Michele Sedor will speak more on the Data Cookbook selection and implementation plans
- Finish drafting and approving key metrics
- How do we get to a single source of the truth?
- How do we roll out data definitions?

Craig next covered a review of the data governance principles:

- Recognition of primary data sources
- Unambiguous transformation of data
- Transparency
- Data quality, integrity, and security
- Systemic and continuous
- Feasible
- Alignment of data management approaches
- Evaluative

Craig provided a brief review of the general roles and responsibilities graphic that has been shared with the committee previously, and explained the purpose of the DSC, DGC, working teams, and subject matter experts. The role of Data Stewardship Coordinator was again reviewed, with the intent of further discussion with the Data Governance Council on what this role might look like in coming months.

Michele Sedor spoke next to discuss the Data Cookbook purchase and implementation plans. Data Cookbook is a hosted metadata repository with a higher education focus. Three other SUNY institutions are using this product: Empire State, SUNY Downstate, and Jefferson Community College, and seven AAUDE peer institutions are also clients. Data Cookbook has a search database of terms and workflow model to help with the data definitions process. Links can be embedded in the product dictionary to connect to OBIEE or Tableau as well.

Michele shared the proposed timeline for the Data Cookbook implementation:

- January: Administrative kickoff and setup
- February: Content Creation Training (starting with subject matter experts)
- March: Workflow training
- October: Project completion goal

Michele and Jon Havey from Institutional Analysis will be the administrators for Data Cookbook.

Provost Zukoski asked about the embedded links: do these work from the data dictionary into SIRI? Craig and Gary confirmed that this can happen, and that a library of official reports can be created within Data Cookbook. Examples of some of these reports could be official headcounts, faculty counts, and enrollments, and the reports could link back to the definitions.

Beth asked what definitions would be captured within Data Cookbook. Michele replied that there is some configurability within the product and that she and Jon will work with the working teams to get information into the definitions. Beth also asked if spreadsheets with definitions could be uploaded, and Gary indicated that they are working to determine if it easier to make changes to files, then upload, or edit already-uploaded files. The administrators will be working with the Data Cookbook consultant on this process. Craig commented that the dictionary included in Data Cookbook will help us get to a common vernacular as well. More information on Data Cookbook will be distributed soon.

Provost Zukoski thanked the group for allowing him to be a part of the updates at today’s meeting. He often does “data battles” with deans and it is helpful for him to be here to participate. The work that committee does is important in order for the university to run efficiently. UB will be more reliant on correct data and consistent definitions, and having those will make everyone’s jobs easier.

Craig next asked each working team to present their updates. Working Team 1(Enrollment and Degrees) spoke first, thanking all of the members of their team for their assistance. The team’s main charge was to identify and document common structure and processes. Mark gave examples of how discrepancies can arise. The team had many discussions over time, and the varied perspectives of team members provided valuable insight into the process. A significant challenge the team identified was reconciling numbers from transactional systems and official definitions. The team provided definitions for three types of data: operational, official, and strategic. The team, as part of its work, came up with a list and matrix of definitions relevant in these areas. The data definitions process forced focus and creation of definitions, and use of prioritization measures to satisfy immediate needs versus those less urgent made the process much smoother.

Tableau and OBIEE have led to an increase in reporting needs, and reports used for administrative decisions were the starting point in definitions. Mark mentioned that the 20,000 foot view approach was used in creating data definitions from these reports. Nigel mentioned that the team began with 88 terms, and identified 30 as immediate priorities. These 30 terms are all defined. 36 more terms are identified as top priorities and are next for the team to define. 22 terms remain that will need to be prioritized. The data definitions prepared by the team show where definitions and reports originated in order to provide a view of the similarities and contrasts between terms. Report frequencies and data tables are also included. Brian O'Connor spoke to one particular term the team created: fractionalized: which can be used to count "parts" of students who might be enrolled in two or more units, or in two or more plans. Brian was tasked with disseminating this term among other campus offices and reports that it is starting to work and others are using this term.

Beth Corry spoke about Team 2 (Finance)'s progress to date. More than 200 financial terms have been identified, and the team is down to the last 40 that need to have definitions approved by the team. The terms are sent out for comments prior to meetings. Central offices, academic support, resource planning, and the research foundation are all represented on this committee. Discussion over terms has been useful. The team is including all financial terms that appear in Info Source, SIRI, and financial forms. Procurement terms are on hold at this point. The deadline for this team to finish their work is early March. Although many terms were externally defined, the team has "UB-ized" some terms for every day definition purposes.

Mark Coldren provided Working Team 3 (Human Resources)'s update. The team is ready to have the 42 definitions they have created presented to the DSC. Many of these terms came through the Board of Trustees documentation. They have identified 15 types of faculty. The team is also looking at snapshot dates and timing of these. The next group of definitions will include volunteers, positions, and consistency on workloads. The team plans to present next month.

Team 4 (Undergraduate Admissions)'s update was presented by Jose. The progress has lagged a bit as work has shifted to admissions priorities rather than definitions, but 33 terms have been identified. Work on the terms and definitions has allowed education on how work is assigned in the office. An example was the term "completed application". As a result, undergraduate applications are now read twice and ISCOMPLETE is not marked to 1 (meaning "complete") until the second reading is complete.

New projects include adding many terms related to the redesigned SAT, which will include more fields than the old version. The team recommended adding all of these fields to Info Source even though they may not be impactful for 3 to 5 years. This way, the information will be available when it is needed. A table for these new elements will be created, and can be imported into Data Cookbook as well. Craig mentioned that PS_RPT_APPLICATIONS_V, the primary undergraduate applications table in Info Source, is due for an update, as it has not been looked at since the HUB go-live date, and could use a refresh to add more relevant fields. Jose agreed with this assessment, and mentioned that that the table could be organized to have fields in alphabetical order and remove some rarely used elements.

Working Team 5 (Graduate Admissions) – recent changes in GEMS staffing meant that the previous lead, Chris Connor, is now with Engineering and Mychal Ostuni will now attend these meetings. Mark and Joe will work with Mychal and Chris' successor, once identified, to make some progress. More information will be forthcoming.

Working Team 6 (Student Financials): Shirley presented the update for this team. They initially thought they would have few terms to define, then realized that some terms are unique to the billing process and would need definitions specific to student financials. Examples included billing career, full or part time, and in or out of state. A total of 56 items were identified and this number may increase as they do collections for the campus and work with tuition receivable data. The team continues to meet to see what else needs to be done. Mark mentioned that tuition revenue process for SUNY reporting is done in conjunction with student financials, and that this is strategic data. The process takes official data (enrollment), merges it with financial data, and then the unit are tasked with meeting targets set as part of the process.

Working Team 7 (Financial Aid): John Gottardy spoke about the team's progress to date. They began with pulling part surveys to identify inconsistencies in data reported out. For example, grant in aid scholarships: what are these – grants or aid? Item types are being scrubbed next. A total of 1,289 item types exist. Other projects include updating endowed scholarships and external awards as the proper data types with correct labels and attributions. The term "aid received" can be problematic as well: does the aid need to funnel through the system? Provost Zukoski stated a problem with TAP is that the money can only be attributed to a course associated with the students major. John stated that this leads to requiring a definition of a student's major or a primary major. These are examples of what can happen with a lack of definitions.

Working Team 8 (Entity Hierarchy) – Mark provided a brief overview of the entity hierarchy process and structure, explaining the entity is the "glue" that allows information to link to structures across the enterprise. Most of our systems have entity hierarchy in their DNA so we can use this structure to tie things together. We launched a new system last summer. 60 definitions were created from standard displays and views of organizational structure.

Working Team 9 (Space): Kelly mentioned that this team had the easiest task, as their definitions are already done as they comply with the SUNY space standards. Kelly worked with Sean Brodfuehrer and Dave Barnas and determined that they are completely compliant with SUNY. SUNY is transitioning to the Facilities Inventory Classification Standards, which Dave is coordinating. They have not customized any definitions from SUNY and use them as-is.

Working Team 10 (Research): Tom Wendt was not able to attend, so Mark provided an update. 42 definitions are complete, and many were already defined by external agencies.

Provost Zukoski stated that this looks like all teams have these definitions and are doing data cleanup, but asked how these definitions will be rolled out to the broader campus community. Brian spoke again to his work to get the term "fractionalized" in use by transmitting this definition in meetings he attends to reports produced. Provost Zukoski stated his desire for one source of the truth, or what is defined as the truth. He asked if it was difficult to update reports if an administrator wanted new reports or additional information, and was assured that this would be part of the ongoing data governance effort.

Craig next spoke about the next steps for BIS and strategic planning. We have SIRI, but students are not covered well in that system. There are scattered resources and units doing their own reporting. There should be a way to standardize these reports or to get them more aligned. There may be a need to be more rigid in reporting. This could be done by including standards in the report or linking to definitions.

The data definition approval schedule is as follows:

- February
    - HR (including faculty)
    - Entity Hierarchy
    - Space
- March
    - Finance
    - Student Financials (including tuition revenue)

Definitions can be turned into Info Source models. For example, a field called "voting faculty' could be added to the employee tables and individuals could be coded as appropriate. This could then be used with SIRI or other reporting tools to yield a count of voting faculty by querying appropriately.

Craig spoke briefly about the BI Strategy Task Force: they had hoped Oracle would have come for a group visit by this point but have been unable to schedule it yet.

The next Data Stewardship Committee meeting is set for **February 22, 2016 at 4 p.m. in Capen 567.**