Data Stewardship Committee: Minutes of February 23, 2015

In attendance: Craig Abbey (OIA), Gary Pacer (EAS), Brian O'Connor (CAS), Tom Wendt (VPRE), David Love (SEAS), Leah Feroleto (SW), Troy Joseph (GEMS), Chris Connor (GEMS/UG Admissions), Greg Olsen (VPEM), John Gottardy (Financial Aid), Sue Krzystofiak (HR), Beth Corry (Financial Services), Shirley Walker (Student Accounts), Michael Koziej (Campus Living), Kelly Hayes-McAlonie (Capital Planning), Tom Okon (Business Reporting and Services), Mark Molnar (OIA), Laurie Barnum (Resource Planning) Peter Elkin (Biomedical Informatics), Rachel Link (OIA).

Meeting called to order at 4:00 p.m. by Gary Pacer.

Gary thanked all in attendance for making the time to participate and asked attendees to introduce themselves and the areas represented. Gary gave a brief overview of the Data Stewardship Committee, which is a subsidiary of the Data Governance Council. Craig Abbey and Gary Pacer are co-chairing the DSC, and the initial invitation sent to the DSC members included the charge from the Provost for both the DSC and DGC. Thirteen data domains are represented on the committee, providing a university-wide perspective of how data governance is managed across the institution, how data are defined for particular projects, and identifying common vernacular to solve or address problems. The DSC membership is comprised of Data Stewards, Ex Officio members, and Ad Lucem (Latin for "to the light") members, who serve to help lead the data stewardship process.

Gary next reminded those in attendance of the work that lies ahead for the committee. The initial focus is to prepare the report recommending permanent data management organization and structure. Required to complete this report are: researching what models exist at other institutions; formalizing definitions of data steward and data trustee roles; and identifying and inventorying key institutional data elements across domains. 3-5 elements from each domain will be selected to help build an inventory of definitions; define attributes; store materials in an accessible repository; and finally, being work of conforming the UB institutional data warehouse to official definitions.

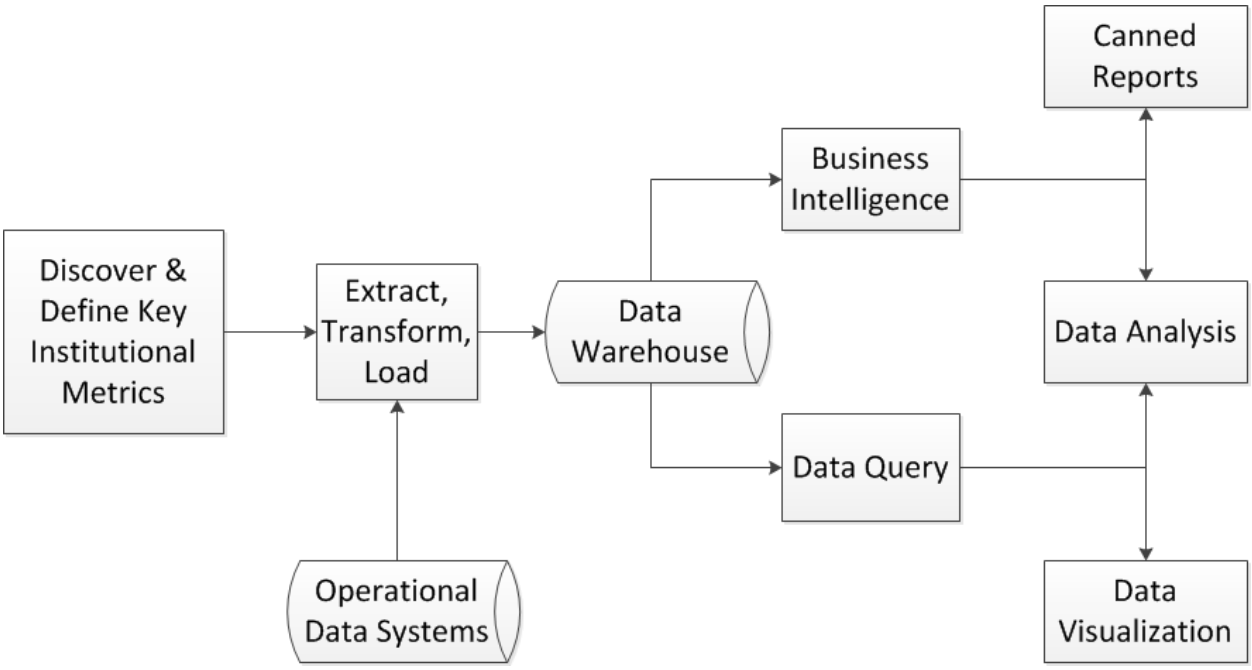
Gary stated that by April 2015, the DSC must prepare a progress report to the Data Governance Council, prepare an inventory of key data elements, and provide preliminary definitions of students, degrees, and faculty. By June 2015, the recommendations for the permanent data management structure and institutional definitions are due, and implementation will begin.

Gary presented the Data Governance website to committee members: the URL was previously emailed by Craig and is http://www.buffalo.edu/provost/policies-and-resources/data-governance.html

This website will serve as the repository for collateral used across the Data Governance efforts at UB. All individuals listed on the committee should have access to upload documents: if additional people need access to do this, or cannot upload documents themselves, they should contact Craig. Each working

team has a subdirectory to store information relevant to their work. Gary encouraged all committee members to use the website for information.

Craig next presented a flowchart of management of data and data-related assets to explain where the data governance process is headed. The process flow is illustrated below:



The Extract, Transform, and Load process (also referred to as ETL) is how information gets into InfoSource tables or SIRI. Data query tools like Access or SIRI are used to create reports that are used to provide information to guide UB operations. Craig stressed that the first stage – where definitions happen – is vital to the integrity of the process. If definitions are not consistent throughout the process, the possibility of inaccuracy arises.

Next, each working team was given the opportunity to present their progress so far.

Working Team 1 (Structure, Roles, and Responsibilities)'s report was presented by Craig Abbey. The team met last week and talked about the team's structure, roles, and responsibilities. The initial priority was to conduct research into data management systems used by other higher education institutions. Craig and Gary held a conference call with the University of Washington discussing that institution's approach, and have obtained documentation from several other institutions. This information has been stored on the Data Governance site. The purpose of this research is to identify a structure that could work for UB and prepare documentation to support it.

Working Team 1 also discussed the current data governance structure at UB and went over the roles and responsibilities of data trustees and data stewards, and reiterated that UB is considered the owner of all university institutional data, although individual units or departments may have stewardship responsibilities for portions of the data.

Working Team 1 has identified some additional roles and terminologies. The first is that of *data manager*, which describes a university official or their staff who have operational level responsibilities for information management activities related to capture, maintenance and dissemination of data. *Data managers* are also responsible for any data administration activities delegated to them by the data stewards. Overall, data administration responsibilities are shared among data stewards, data managers, and the CIO. *Data users* are individuals who need and use university data as part of their assigned duties or in fulfillment of their role in the university community. The *Data Steward Committee* is comprised of the data stewards, as a group, and is responsible for recommending policies, establishing procedures and guidelines for university-wide data administration activities.

Working Team 1 identified some terminology conflicts with current UB policy: the UB policy library uses different definitions for data steward and data trustee. The Policy Library definitions are related to information security. Working Team 1 will be trying to identify a way to merge these definitions.

Issues and next steps for Working Team 1:

- Background research
- Data definitions and implementations
- Make it clear what data definitions are used in reports and analyses
- Storage of data definitions library
- Governance of operational/reporting systems
- Establish processes for updating definition and security

Craig asked for survey suggestions to collect information about issues users or units might have with data conflicts. Mark Molnar suggested surveying different groups, like the Distributed Data Access group (DDA) or the SIRI users group, or perhaps a focus group to collect responses, as the needs may vary by unit or group.

Working Team 2(Enrollment and Degrees)'s report was presented by Mark Molnar in Kara Saunders' absence. The team has met once so far. At the previous meeting, they dove immediately into the definitions aspect, and have started with creating enrollment definitions. The team identified three types of data:

Operational: Volatile data generated by transactional and related systems to perform local tasks specific to campus enrollment, research, philanthropic, and community service missions. An example of the use of operational data is to create a class list or a student mailing.

Official: Transformed static operational data used to provide a consistent, authoritative representation of the campus for use by a variety of external publics. An example of the use of official data would be to report enrollment to SUNY or IPEDS.

Strategic: Static operational and/or official data used to provide a consistent, authoritative representation of the campus from which to conduct local planning, analytical, reporting, or budget operations. An example of strategic data usage would be to attribute tuition revenue.

Working Team 2 tried to keep their definitions simple for ease of understanding. The team would like to submit these definitions to the rest of the committee for their use. They feel that while they started creating these definitions for enrollment, they can be applied to other types of data as well.

Mark suggested thinking of operational data as "changeable data, changing daily, used by people in units to do their jobs". Official data is then "transformed, static data", which is frozen at a certain time to allow for consistent and reproducible reporting. Strategic data can be considered a fusion of operational and official data, and is typically used for planning purposes. The universe of data that covers all institutional data is operational. Official is a subset of this operational data, while strategic data spans both operational and official. Working Team 2 will add a diagram to their documents library on the Data Governance site that outlines this approach. Craig suggested soliciting input from other units on the data definitions created by Working Team 2 to ensure that the definitions and types of data hold across data domains.

Working Team 3(Faculty)'s report was presented by Sue Krzystofiak. The team has met once so far. The initial meeting focused on revisiting the definition of faculty, beginning with UB's definitions, then SUNY definitions. In all, 12 different faculty definitions were identified, which vary depending on the area examined. The team tried to keep the discussion at the top level, focusing on questions like full or part time status, EOC, librarians, and the like. The team is collecting official faculty definitions from the AAU, IPEDS, and other organizations to help guide the discussion. The next steps they have identified are to collate this material and see where it leads them in terms of recommendations for data definitions. Kelly Hayes-McAlonie asked if these recommendations might come out to the units. Sue answered that yes, they will, because the team had discussed how different definitions or titles are used by campus units. Kelly stated that "adjunct" might mean someone not on the tenure track, yet working full-time, or it might mean a person on campus three hours a week.

Gary stated that some definitions are actually collections of sub-definitions, collecting all members in a particular group. IPEDS definitions, for example, might include people not in other groups – or it might include people already categorized in a group. Craig identified the terms "instructional faculty" vs "ladder faculty" as ones that can lead to confusion, and that Delaware Study and Middle States definitions are also confusing to some. Sue pointed out that faculty titles are used for multiple purposes. The term "visiting associate professor" means different things depending on which unit uses it: in Law, it is used for non-tenured faculty, but in other units implies a temporary appointment. Peter Elkin pointed out that in the Research Foundation, titles may be tied to HR pay grades, but may not tell anyone how often someone is on campus or in the department. Sue confirmed that they are looking at the RF titles usage as well. Craig also brought up the issue of postdocs, who may be appointed under different titles, which makes counting them very difficult.

Next steps:

1. First report to data trustees is due on February 26th
2. Sample draft definitions from each working team at next meeting (tentatively scheduled for 3/16)

3. Complete set of data governance materials from other institution (to be stored on the website)

Meeting adjourned at 4:50 p.m.