

Overview of the Discussion Between Computer Science, Sociology and the GSLIS in Planning a New Master's Program in Data (Science/Management/Librarianship) During Fall 2012

Background to the discussion: Across multiple disciplines and professions, including LIS, the composite area of Data Science/Data Management/Data Librarianship is getting a lot of attention. It is sexy. At Queens College, there is a strong desire within the social science division to get a piece of this action and we have been invited to join in. Within the LIS profession, there is much attention to data management/data librarianship/data curation/data science at the present time and the GSLIS might want to consider how this new focus supports our goals and objectives. (For three exemplary LIS efforts in this area, see:)

<http://sils.unc.edu/news/2012/pmc-kick-off>;

<http://cirss.lis.illinois.edu/CollMeta/dcep.html>;

<http://ischool.syr.edu/future/cas/datascience.aspx> .

On September 25, 2012 Dean Weinberg called a meeting with the representatives from natural sciences and social sciences to discuss an interdisciplinary collaboration to develop a data science/management Master's program. (The exact name has not been finalized.) GSLIS was represented by Marcum, Ng and myself. Lunch was served. Follow up meetings were held on October 15, and November 28, with representation from Zhigang Xiang – chair of computer science, Andrew Beveridge, Emilio Zagheni and Holly Reed from sociology, Dean Weinberg, and Marcum, Ng and Cool from GSLIS.

At the time of the last meeting I attended, it was determined that the long term goal is to develop a new master's degree, with a more immediate plan to launch (fall 2014, I believe) a concentration or program, with three tracks – one in computer science; one in sociology, anchored in data analytics; and one in library science. Although the original impetus for this program came from the social science division, computer science has joined forces with much valuable contribution.

In addition to the in-person meetings, many email messages have been exchanged over the past several months about possible directions for this program, in particular, courses that might be offered and required. All of these of these emails were forwarded to Bobby Brody when she took over as chair.

The emerging plan, as I understand it, would involve all students to take the following required courses, OR THEIR EQUIVALENTS, AS DETERMINED BY DEPARTMENT:

From **computer science**, the following undergraduate courses would be offered as 600 level graduate courses for non computer science majors: (course descriptions

and syllabi are available)

CSCI 111 – Introduction to Algorithmic Problem Solving

CSCI 212 – Object Oriented java

CSCI 313 – Data Structures

CSCI 120 - Discrete Structures

After taking these courses, or their equivalents, students would be eligible to take a variety of special topics courses as electives, including:

CSCI 780 – Data Mining and Warehousing

From **sociology**, the following two **4 credit** courses, OR THEIR EQUIVALENTS are required:

SOC 710. Basic Analytics. 6 hr. plus conf. and lab; 4 cr. 6 contact hours. Prereq: Introductory undergraduate course in statistics. This course focuses on basic statistical concepts and analytic techniques with an emphasis on application to real-world problems and issues. It is the first in a two-semester sequence.

Emphasizes describing and summarizing data, statistical inference procedures and reasoning, and use of statistical software in analysis.

SOC 712. Advanced Analytics. 6 hr. plus conf. and lab; 4 cr. 6 contact hours.

Prereq.: SOC 710 or equivalent. An examination of advanced statistical methods, inference, and multivariate techniques, such as ANOVA, linear regression, and logistic regression.

From the **GSLIS**. We have not put forth required courses from the GSLIS, and it is not clear that either computer science or sociology is willing to accept required courses from us. However, since the GSLIS has not stated its requirements for participation in this program, the door is not closed on this discussion. **CLEARLY**, this needs to be addressed if we go forward with this joint initiative. As of our last meeting, the following courses were suggested by KB as *possibly* relevant to this initiative, with some modifications, but of course KB will speak further on his own behalf about these courses and the data management program.

LIBSC 729: Metadata and XML..

LIBSC 746: Database construction.

LIBSC 748: Web Programming. (CSS, JavaScript, JavaScript libraries like jQuery, for web environment).

Conclusion at this point – the above is my understanding of the data management program in the social science division. My personal opinion is that this is an exciting area for us to possibly expand into. The division has already laid out an array of courses that we may wish to explore. I think that the courses put forth above from computer science and sociology are more relevant to an “information science”

rather than to a “data curation” track. This seems to be a next topic of conversation for the GSLIS curriculum Committee