**Large Scale Computational Methods for the Printed Book Review:**

**A “Data Lifecycle” Approach**

Book history and sociology of literature, including but not limited to scholars of readership and reception, have made numerous attempts to account for the role of book reviews in print culture. In an American literature context, studies by Nina Baym, William Charvat, Cathy Davidson, and Janice Radway point to the complexities that book reviews raise, as well as their potential to represent important aspects of historical readership. More recently, digital humanities emerging scholars such as Ashley Champagne and Allison Hegel have turned to book review websites like Goodreads to bring distant reading strategies to bear on subject such as literary canon construction and how readers define genre expectations.[[1]](#footnote-1) Yet we might easily read Nina Baym’s *Novels, Readers, and Reviewers* (1984) as a precursor to Champagne and Hegel’s work. Baym’s ambitious study drew conclusions about a growing “vocabulary available for writing about novels” by considering a corpus of more than 2,000 antebellum American book reviews at a time when the average computer had 64 kilobytes of memory and no hard disk.[[2]](#footnote-2)

In my paper, I will focus on book reviews the turn-of-the century United States, in order to underline fundamental compatibilities between large-scale, computational methods and book historical approaches. For two years, I have worked a team of digital humanities colleagues at the University of Pittsburgh on a project titled “Computational Approaches to Textual Networks.”[[3]](#footnote-3) Our primary goal has been to create workflow models to address challenges of producing machine-readable textual corpora with strong relational attributes.[[4]](#footnote-4) In the case of book reviews, this means having high quality metadata to describe each individual review from a periodical, as well as highly faceted information about the book or books being reviewed in each piece, such that a network of actors and texts can be reconstructed. Our efforts have included small-scale digitization, adapting or “up-coding” existing digital assets.[[5]](#footnote-5) My talk will focus on viewing historical book reviews at each stage of the data lifecycle: creation/collection, processing, analysis, publication, archiving/preservation, and re-use. This model, I will argue, has enormous potential to reveal longstanding compatibilities between bibliography/book history and humanities computing, and to create collaborations that push two often-partitioned scholarly communities in new directions.

1. Champagne is a doctoral candidate at University of California, Santa Barbara and Allison Hegel is finishing her dissertation at University of California, Los Angeles. [↑](#footnote-ref-1)
2. Baym, Nina. *Novels, Readers, and Reviewers: Responses to Fiction in Antebellum America*. Ithaca: Cornell UP, 1984. [↑](#footnote-ref-2)
3. The faculty members comprising this team are Benjamin Miller, Alison Langmead, Annette Vee, and me. Graduate students Daniel Libertz and Lucia LoTempio served as summer 2016 graduate student assistants on the project, and undergraduate Zach Luettgen did an independent study on book reviews as networks in Spring 2017. [↑](#footnote-ref-3)
4. In addition to documenting the idiosyncratic considerations of digitizing bibliographical objects like book reviews, our project has also focused on best practices for data modeling and data curation when working with these materials. [↑](#footnote-ref-4)
5. For example, Proquest’s *American Periodicals Series* and the *HathiTrust* collection. [↑](#footnote-ref-5)