**[Slide 1: Title]**

In May of 1905, a review of *The Troll Garden*, Willa Cather’s first published collection of short stories, appeared in *The* *New York Times*. The title of the review is “Promising Stories” and the length is three paragraphs, or what seems to be a rather short notice when we compare it to reviews of *The Troll Garden* from *Harper’s Weekly*, *Bookman,* or *Atlantic Monthly*, all of which are included in *Willa Cather*: *The Contemporary Revi*ews.[[1]](#footnote-1)

**[Slide 2: Troll Garden]**

But what if we compare this review of *The Troll Garden* to other reviews in *The New York Times*, or other reviews in *The New York Times* in 1905, or other reviews in *The New York Times* in May of 1905 to characterize its length? *The New York Times* archive API makes these comparisons relatively easy, as one can quite easily construct a Python script to retrieve review metadata, including review word counts, by month and year. And here is what we find:

**[Slide 3: NYT-1]**

These word lengths do not have a “normal distribution” or the iconic bell curve we might remember from a statistics class. Instead, they range from 29 to 4,492 words long. A review cannot be shorter than 0 words, the median review length is 287 words, and the largest review length is much larger than most of the reviews. We therefore see a distribution converging to log-normal. Using a log function, we can convert the values to a normal distribution and analyze each observation’s relationship to the norms of the data.

**[Slide 4: NYT-2]**

The review of Cather’s collection has a z-score of -.3059, which is to say that it is less than one third of one standard deviation away from the norm. A statistically accurate description would characterize the review as being within the typical length range for *The New York Times* in April 1905.

Perhaps someone in the audience here today is already wondering what the point this introduction has been. It’s not an unfair question. In this paper, I focus on book reviews of the turn-of-the century United States as crucial, currently undervalued book historical objects. I frame my analysis with several important stages of the data lifecycle—planning, creation/collection, processing, analysis, preservation/distribution/discovery, and re-use. This model, I will argue, has enormous potential to reveal longstanding compatibilities between book history, and large scale humanities computing.

Before I go too far, I must acknowledge that the Data Lifecycle has become something of a librarian’s slideshow cliché.

**[Slide 5: lifecycle-all]**

Lifecycle diagrams are liable to show up as flow charts; brightly colored circles with arrows; A mostly blue geometric diamond shape with one green triangle that says “Dependent Processes”; swirly thingies making up an outer circle that seems to be swirling around an inner circle that’s swirling in the opposite direction; an infographic with the universal icons for a server, a computer, downloading from the cloud and … I think that’s a Word War II fighter plane? And many more. Just past this preponderance of chatter and hype, however, is potentially crucial perspective on how research data move through an ecosystem of interlinked practitioners. For this reason, I want to emphasize the model as described by the Data Documentation Initiative.

**[Slide 6: DDI Lifecycle] -** Ball, A., 2012. Review of Data Management Lifecycle Models. University of Bath.

The DDI Lifecycle version 3.0 emphasizes study concept, data collection, data processing, data archiving and distribution, data discovery, and data analysis. As with most models, repurposing is considered the end of one iteration of the lifecycle and the beginning of new one. I will discuss these elements in non-chronological order as I relate them to historical book reviews.

* Phase I: Study Concept/Planning; pair with Analyze
  + Return to Cather
  + Other review types, especially Bookman, a newspaper, as well as Goodreads and Amazon to show later developments.
  + Describe some of the other investigative work we did
* Phase II: Collection/Acquisition
* Phase III: Processing
* Phases IV, V, VI: Preservation, Distribution, Discovery
  + Publication is sometimes its own phase

Conclusion: Why phase VIII is reuse

Even the most theory-heavy humanistic scholarship is peppered with what we might call pseudo-quantitative claims or characterizations. A quick measurement like the one I reenacted reminds us that human intuition often mischaracterizes subtle things like size, frequency, density, or duration, especially when we are focusing on how one case relates to a norm because anchoring bias, attentional bias, confirmation bias, and overconfidence are all potentially at play.[[2]](#footnote-2) Positioning oneself against “positivistic, strictly quantitative, mechanistic, reductive and literal” measures and methods has become increasingly fashionable in the humanities, but the humanities have never abandoned our need to speak of external, observable phenomena and divide those phenomena into legible units.[[3]](#footnote-3) Bibliographers and book historians, all too often, have quietly reminded literary studies scholars that careful observation and tabulation lays the foundation for close reading and even literary theory.

Second, I’m hoping a measure as simple yet still deceptive as review length will underline the importance of seeing computational analysis and data stewardship practices as interdependent roles.

* Phase I: Study Concept/Planning; pair with Analyze
  + Return to Cather
  + Other review types, especially Bookman, a newspaper, as well as Goodreads and Amazon to show later developments.
  + Describe some of the other investigative work we did
* Phase II: Collection/Acquisition
* Phase III: Processing
* Phases IV, V, VI: Preservation, Distribution, Discovery
  + Publication is sometimes its own phase
* Conclusion: Why phase VIII is reuse

The cultural cachet conveyed by this list of periodical titles would perhaps suggest that *The Troll Garden was* widely reviewed, or at least widely reviewed for a debut short story collection. Based on several memorable lines, we might also consider this a mostly positive review with minor qualifications:

“In this collection of seven stories the author has shown a great deal of deep feeling and real ability” (303).

“One wishes … that Miss Cather would eschew the complex psychological attitude which she so frequently displays” (303).

“‘Flavia and Her Artists,’ holding up to ridicule that senseless toadyism to genius which obliterates entirely any appreciation for real worth which has not already been accepted by the world, is perhaps the best story in the collection” (303).

“There is something about ‘The Garden Lodge’ which rings shallow in spite of the pathos in which it is enveloped, and the same might be said of ‘A Death in the Desert.”

“Nevertheless, the little book contains some excellent work between its covers, and is well worth while if only as an olive branch of promise” (303).

Certainly the *New York Times* review was kinder than Bessie du Bois who, writing for the Bookman, called it “a collection of freak stories that are either lurid, hysterical or unwholesome” reminiscent of a Sunday newspaper’s “coloured supplement,” which was turn-of-the-century name for what we would now call the funny paper or the comic section (612).

Every day document so ubiquitous as to seem self-evident,

It is seldom if ever treated as a highly rehearsed and performative genre with deep rhetorical and material codes of information transfer.

* The data lifecycle
  + Study Concept/planning: design, assess, modeling, looking ahead etc.
  + Collection/acquisition: make, purchase, collect, datify
  + Processing: anonymize, cleanup, merge, convert, tag, ingest, normalize, munge, memory saving tasks
  + Archiving/preserve or destroy: curation, stewardship, context for outside audience, document practices, articulate absence and uncertainties
  + Distribution/share dataset: ideally in a discoverable way
  + Discovery: readers, emphasizes other using finding and access your data
  + Analyze: computational and/or non-computational
  + Repurposing/reuse
  + Publish: scholarship based upon the data, contains a more rigorous interpretation of results
* How I'm interpreting the data lifecycle
  + Spiral aspect
  + Recognition of multiple stakeholders
* Why talk about book reviews this way?
  + Deeply tied to a direction I would like to see my discipline go in
  + Be very clear about what doesn't exist that should. I especially need to emphasize searchability, accessibility, and computability as different things.
    - “big data” line from Kirschenbaum and Werner
    - look at book reviews proposal and FER
    - look at recent blog post
    - Piper, “There will be numbers”

It would be wrong, and intellectually limiting, to see this undertaking solely as computer science applied to culture. Cultural analytics requires a wholesale rethinking of *both* of these categories. Computation forces us to rethink our current disciplinary practices in the humanities from the ground up. What counts as evidence? What is the relationship between theory and practice? How do we account for the technological mediations of our critique?

* + - Underwood on literary history
    - Laura Mandell
    - look at Bode, Piper’s response
  + I want to argue that the roots of bibliography and book history are fundamentally compatible with large-scale analytics, but also having certain incompatibilities that need to be addressed
* Broad goals:
  + Bring the reader close to physical proxies of primary source book reviews
    - Cather review as example of information contained there in
    - Connect to Cather: the contemporary reviews
    - Poe, DuBois, Fanny Butcher, William Stanley Braithwaite, Gilder and Howells, Lovecraft, Goodreads, Amazon, Kirkus, Audible, Rotten Tomatoes
  + Emphasize how book historically minded scholars have been using an interpreting book reviews
    - Bode
    - Hegel
    - Ashley Champagne
    - Nina Baym
    - Davidson, Revolution and the Word
    - Reread Suarez for mentions of book reviews
    - Book History – reception and readership
    - Bourdieu
    - Janice Radway
    - Joan Shelley Rubin
    - Genette

1. *Willa Cather*: *The Contemporary Reviews,* edited by Margaret Anne O’Connor, which organizes Cather’s reviews by work reviewed rather than grouping them by periodical of origin or placing them in chronological order. [↑](#footnote-ref-1)
2. Anchoring is the tendency to overvalue a particular trait. Attention bias is the tendency to be affected by recurring thoughts. Confirmation bias describes a tendency where the “selection and evaluation of information that would confirm a focal hypothesis is given priority—or even exclusivity” (Rajsic, Wilson, and Pratt, “Confirmation Bias in Visual Search,” 1353). Overconfidence is a tendency to overvalue one’s own answers to questions, qwhihc might in turn prevent someone from taking a measurement or consulting an outside source. [↑](#footnote-ref-2)
3. Drucker, “Humanistic Theory and Digital Scholarship,” 86. [↑](#footnote-ref-3)