Public Domain Editions

Chris Forster

This is an extended version of the (less than) two minute "dork short" or "lightning talk" I gave at THATCamp Virginia a while ago (this post has been sitting in the hopper for a while). I offer an observation, an anecdote, and a suggestion.

tl;dr: I'm trying to put together an edition of Claude McKay's *Harlem Shadows.* Would you like to help?

An Observation:

An enormous wealth of public domain material is available on the web, from sources like Project Gutenberg or The Oxford Text Archive or The Internet Archive or Google Books or smaller projects like The Modernist Journals Project.

Yet, in my experience, these texts seem underused. (Am I wrong?)

An Anecdote:

When I was a teaching assistant for UVA's twentieth-century literature survey a few years ago, the professors taught Claude McKay's *Harlem Shadows.* Published in 1922, *Harlem Shadows* is just inside the public domain.

The text they used was a cheap (though still in the neighborhood of

$15; here it is at Amazon) paperback facsimile of the 1922 edition. When I opened this slight paperback, it looked eerily familiar.

Compare:

The top image is from the Google Books edition; the bottom is a scan I just made of the Kessinger edition. Kessinger's "edition" of *Harlem Shadows* is printed from page images available at Google Books (scanned, in turn, from a copy at Indiana University library). They've cleaned up the title page a bit, but look at the distinct pencil marks. That's Kessinger's business: get new ISBNs for Google Books scans and then sell them. (When folks first noticed Kessinger doing this a while ago it caused some consternation.)

(Worth noting: there is a another copy of *Harlem Shadows* (scanned from a copy held at Princeton) in GBooks, which misidentifies Max Eastman in the author metadata; in addition to the two Google Books copies, archive.org has two copies; one from the Library of Congress and one from the University of Toronto, all the same edition. Thoughts on easily breaking up those four PDFs and digitally collating them?)

It seems unfortunate that right now a professor who wants to teach *Harlem Shadows*, ends up assigning Kessinger's rather ugly printout of a Google Books PDF.

A Suggestion:

Can we do something to make public domain texts more useful? Is there a place for (some) scholars to take the lead here? Rather than paying Kessinger to print out Google Books page-scans, could we not use the (in this case, multiple sets of) page-scans available from a variety of sources to put together a lightly marked up version of the text? Couldn't we draw on existing bibliographies to make clear what the book object represented by those scans actually is. And then, from our single encoding, could we not export to multiple formats: PDF (by way of LaTeX, for folks who want to print this

thing out); HTML; and ePub (etc) for eReaders?

Such an idea is not novel; it is merely an expression of the dream of a markup language like TEI. Not so long ago, a proposal for a "A Git Powered Project Gutenberg" lead to a discussion on Hacker News which in turn lead to a hastily arranged group (which just as quickly disarranged itself)—all focused around the idea of making public domain texts better. There is interest in improving the accessibility and usability of public domain texts and it isn't confined to academic literature departments.

Scholars could play a key role here by helping to establish a good text and providing annotations and glosses or other contextual material. In my wilder moments I imagine scholars providing a base text which than then becomes the staple, raw ingredient in a variety of remix editions, produced for audiences varying from high school to the college classroom, and beyond. These texts in turn could be cut and remixed to produce a roll-your-own anthology.

An Acknowledgment and a Goal:

There are some excellent reasons why **I** shouldn't be doing this. First, in the specific case of *Harlem Shadows*, I am not a specialist in American, African American, or Caribbean literature in general, nor in Claude McKay's work in particular. Nor am I an expert in text markup. Nor am I sufficiently well versed in the dark bibliographical arts to really be handling the complexities of putting together a proper critical edition.

With those reservations stated, I'm trying to carve some time out to work on this nonetheless. One's reach should exceed one's grasp, else what's a public domain for? But boy would I love some help.

I've converted the plaintext, OCR'd version of *Harlem Shadows* available through archive.org to a lightly marked up TEI version of that text. This markup itself is worthy of scrutiny; but I wanted to have *something* to start with on the way to producing a proofread,

bibliographically sound, TEI-version of the text; to that I'd like to add annotations and textual notes, as well as supplementary material—early reviews, maybe McKay's prose from this period, as relevant. Think Norton Critical Edition (minus the criticism which is likely too thorny a permissions matter; though I'd love to proved wrong on this front).

To begin:

- here is a github repository with my initial stab at marking up the text.
- here is a wiki to organize future work. (Let me know if you want to be added to the wiki).

(A minor technical note: For a while I was imagining that it would be possible to use stand-off markup to keep text and annotation completely separate. This would be great for many reasons; in theory, one could have different sets of notes for different audiences (the high school versus the college class room; a reading versus a scholarly edition); from the little reading I've done, that seems not easily feasible at the moment. For software developers, however, the problem of how to combine constantly evolving sets of dependent texts is simply a fact of life; version control systems, like git, provide some help in managing this problem.)

As a preliminary schedule: begin finalizing markup of the edition by the end of the summer. Continue collecting and adding supplementary material and annotations in the Fall. Then start working on processing the text out to desired formats (the TEI Stylesheets provide a great place to start); so that this time next summer, an edition of sorts (available in multiple formats) is done.

For now I'd be interested in other folks sharing their thoughts, criticism, or enthusiasm. Or, better yet, take some of this material and fix it or fork it.