What has DY done for me lately? A lot

Thank you all for your attendance and also your endurance. Before I begin, just a little house keeping. This talk is available as an interactive presentation and can be accessed through the web address below, or if you have good aim with your QR scanner you can also pull it up on your phone directly. There are some electronic doodads and widgets embedded in the presentation that are interesting to play around with, and, if you’re that kind of person, you can also check my work. I don’t want to spend too much time talking about individual charts, so you’ll have to take it on faith that the charts mean what I say they mean, and then zip through them rather expeditiously.

This talk will touch upon some of the research that has been possible with the data inside the DY database. Its larger aim though is to make a case for our data model, and how our methodology can be fruitfully be applied to other modernist and, gasp, non-modernist authors. While a discourse on method, usually is not a show-stopper at a conference, it is an important discussion to start if we want computational methods to have any type of sustained and lasting impact on how we study modernism. To my mind, the DH ecosystem currently suffers from two major issues: longevity and interoperability. Certainly, I am not alone here and these two issues have plagued DH since its inception.

DH projects have fractional life-cycles compared to print scholarship. Projects that are launched with great fanfare often become defunct only a few years later. Solving the longevity issues is not merely an academic question it is also an ethical one. Let’s face it, DH programs suck up an incredible amount of funding. For 2023, the NEH announced it will disburse 2.2 million dollars in grants for DH projects. For that amount of money, you could hire 30 “traditional” English professors who could churn out articles and books in print that won’t simply evaporate with the latest browser update.

Additionally, DH projects are rarely able to speak to each other without significant data wrangling. This is more than merely a technical issue. Practically, it means that DH projects therefore do not benefit from the network effects visible in other disciplines such as political science, economics, or the hard sciences. There data from individual projects can be recycled for other scholarship. For example, voting data from the US can be compared to voting data from Canada, even if there are important nuances in how that data is collected. Meanwhile, in the humanities, the data from a project about author X can rarely be used effectively as a point of comparison with author Y. Obviously, there are important exceptions to this such as TEI, The Seshat Global History Databank, HathiTrust, and others, but, by and large, this is the landscape.

In this sense, longevity and interoperability are intimately related. By making the data interoperable we are ensuring its reincarnation and upcycling in other projects.

I would argue that the data model created, refined, and tested for nearly a decade by the Digital Yoknapatawpha team represents a good starting point for a more universal framework.

Without going into too much technical detail, we can say that DY collects three main data entities: characters, locations, events.

Each of these entities, in turn, have any number of attributes from characters’ gender, race, and class to which locations in Yoknapatawpha are built and destroyed.

The relationship between these three entities represents one of the fundamental structures of fiction: people (characters) doing something (events) somewhere (location). This flexible structure could arguably be applied and modified for most any author. Doing so would allow us to ask and answer questions that exceed a single author, and start to map the larger literary ecosystem. This in turn could help us reimagine the perennial question of this conference: What do we mean when we say Modernism?

Lighting Example: Faulkner and Plot Structure

One of the distinguish features of Faulkner, and, perhaps Modernism, is narrative experimentation. In our data, we have information that will let use visualize Faulkner’s use of time and narrative voice.