Music Research Data Management

Francesca Giannetti

22 February 2019

Title Slide

Teaching (Music) Data Literacy

My remarks will focus on teaching data literacy as a component of digital humanities and data services outreach. I will illustrate my points using analogies to more established frameworks for understanding research, and by examining a digital project in music as an example of data as (proto-) argument.

Slide 1 (purposefully blank)

My particular work context focuses to a large extent on digital humanities pedagogy, more so than digital project development and digital preservation, so I will let Anna and Jon fill us in on those topics. In my role as digital humanities librarian, I work with students and faculty at my institution who are interested in digital methods but who may not have prior experience with machine-readable data formats.

Slide 2: Trivium

I like to make analogies to older frameworks when describing digital work in order to see that work as a logical continuation of past practice. The medieval trivium has origins in the rhetorical tradition of ancient Greece; it along with the four mathematical arts laid the groundwork for the medieval university. The trivium was comprised of the three verbal arts: grammar, dialectic or logic, and rhetoric. Grammar, or the building blocks of language, is the foundation upon which all other liberal arts are built. Dialectic is the study of reasoning or logic. And rhetoric combines the two preceding arts to create flowing, persuasive arguments.

Slide 3: I/O

I then compare this with a simple input/output chart in order to invite students to see an abstract pipeline that could represent any number of things.

Slide 4: Traditional Research Process

Next, I show an admittedly simplified chart of the research process, in which the inputs become all the literature we read when we do research, our own thinking about that literature, and the eventual product, usually a written paper, that we produce.

Slide 5: Digital Research Process

And last, this is my estimation of a digital research process, which is a little messier, but it still nonetheless shows inputs and outputs, and it has a place for the mediation of our thoughts as well as any computer process we may apply to our inputs or data.

In my teaching, I talk about data as both a source and a product of the research pipeline. In practice, this can mean multiple things. In a literal sense, researchers may base their analysis on original data that they collect or create. And if they use computational techniques, they may produce derived data as a result of that computational process. More abstractly though, the data that researchers create for the purposes of research will already contain embedded values and assumptions. Since researchers typically create data that is specifically oriented towards their research needs, it becomes difficult to speak of the data as demonstrating any kind of truth value. It's more useful to evaluate the data for their appropriateness and persuasiveness relative to a stated research goal. In this way, the original data already contain the seeds of the argument, and for this reason, they are at least as significant as the findings. Researchers' inputs constitute an important part of the research project as a whole. For methodological openness and transparency, it is important to document and share original data, together with the methods, and the findings, so that other researchers may find them and potentially reuse and adapt them. In this way, data are input and output both in the sense that they already contain their arguments, and they may get reused to develop new arguments.

Becoming proficient in digital methods takes a great deal of time. In the past, I have introduced students to digital humanities by inviting them to evaluate one or more scholarly digital projects. Digital projects in many ways mimic the structure of print scholarship. Each stage of the pipeline is perhaps more diverse and multifaceted, but there's still a rhetorical arc to the whole.

Slide 6: Evaluating Data

While I still do this, increasingly I find it more productive to focus on one part of the pipeline: data. The interest of asking students to evaluate research data is twofold. One, the ability to assess data is analogous to the grammar of data literacy; it's a critical first step towards developing the ability to manipulate, summarize, and present data. Two, datasets are generally created to pursue specific research goals. This means that some questions may be productively explored with those data just as others may be entirely out of scope. Since most of us have the weakness of imagining data to represent some universal, unassailable truth, it is helpful to examine actual research data to see the ways in which this is false. Highlighting the constructed nature of data overlaps with the goals of critical information literacy, and in this way I can situate my practice as close to library information literacy instruction, while exploring some of the unique aspects of data-driven research processes.

But I should take a step back and first define data. Somewhat like Suzanne Briet and her definition of a document, I take a liberal view of "data" to mean any component of a whole that we choose to make an object of study. However, most of the time I presuppose that we are talking about digital data, which involves a transformation to make our data understandable to computers. And sometimes I am even more specific than that, which is to say that I am referring to machine-readable data.

Slide 7: Evaluating Music Data

Machine-readable data have the added benefit of being manipulable and tractable for our computers. It's the difference between an image of a score, and various forms of encoded music where instead of just pixels we have information about pitch, meter, and duration, among other things.

Here, we are looking at three different ways of encoding that C in the middle of the bass clef. I like to ask students which of these three encodings is the most concise, which the most verbose, and what they imagine to be possible to accomplish in each. The interest of this question is not to complain about the proliferation of music encoding syntaxes (although maybe that too), but to highlight the affordances and disadvantages of each, relative to a specific purpose.

Slide 8: Meteomozart

I'm going to take a specific example of a project by Raffaele Viglianti who works on dynamic digital scores. About this project, he says:

"Meteomozart is a dynamic score of Mozart's Piano Sonata No.13 in B flat major, K.333/315c... Meteomozart adjusts the score based on the weather at your location (or at a location you set). Different slurs and dynamics are shown, taken from four sources: Mozart's 1783 manuscript, the first printed edition (1784) and two performing editions by Bartók (1911) and Saint-Saëns (1915)."

Slide 9: Meteomozart Data

A couple of observations about this project. Viglianti pursues a very specific question about variations in the use of slurs and dynamics across the autograph manuscript and various print editions. His most robust descriptive markup is directly connected to this question. He also provides some fairly uncontroversial data pertaining to note values, meter, and key signature. It is possible to imagine reusing his MEI file to pursue a different question about either the autograph score or of the printed editions.

Slide 10: Data Literacy Practices

To conclude, I would like to cite these five data literacy practices from a study by

Calzada and Marzal. Notice that, among these practices, I only discuss understanding and evaluating data, which I see as the grammar or building blocks of the more complex practices. I do this for a couple of reasons. It doesn't require as much quantitative or statistical expertise as you might think to study the research data of scholars. For example, humdrum and MEI are meant to be human-readable in addition to machine-readable syntaxes. There's often no need to be a proficient user of the syntax to reuse the music data of scholars in one's teaching. Lastly, it is advantageous to students to have some exposure to these methods, as your introduction may become a launching point for other forms of data-centric work in their research and in their professional life.

Slide 11: Recommended Readings