

Interview with Professor David Radcliffe of the CATH lab at Virginia Tech

Note on this interview: My recording application for this interview failed, so the information below is a compilation of the notes that I took and is not verbatim. To read more about how Professor Radcliffe became involved with humanities computing, you can see an article online as well: <http://spenserians.cath.vt.edu/Project.php>

Rebecca Jones: How did your interest in digital humanities develop?

David Radcliffe: Quite pragmatically. It was before DH had become a term. I was trying to find tools to perform the research that I needed. I had bibliographies on note cards that I wanted to migrate using a word processor. From there, I started to learn SQL and other programming tools. For me it was a problem solving enterprise. It was really funny, I would have a list of 500 citations and using a word processor, 30 seconds later I would have a process that could pull up the requested items.

Jones: How did you learn your skills or programming languages?

Radcliffe: I was self-taught, we're talking the late 70's, early 80's. You could do a Cobalt class at community college. You had to learn it by yourself. I'm an antiquarian, so I had lots and lots of information that I needed to work with and needed to figure out how to do it.

Jones: We've talked a little about the history of DH. Did you work with TEI? Were you a part of the standardization of TEI?

Radcliffe: I inherited TEI, took a class back in 1992, back before the internet, it was pretty primitive. In 1992, you couldn't do much with TEI, and so I went back to working on databases. Didn't take it up again until 2007 or 2008, when it had come a long way and then it was ready for more common use.

Can you imagine using TEI with nothing to use it with? You just had a useless document. Unless you were a programmer you couldn't use it at all. Webpages started being a possibility after that.

Jones: Were you involved in developing the theory of how research would take place in your lab? How did you and your peers approach developing a theory or a practice for thinking about how the CATH lab should operate?

Radcliffe: It was much more about practice than it was about developing theory. It was about problem solving, working with people and working on documents. Trying to find a way to take 20,000 documents with water marks and then trying to find ways to image it, mark it, combine it and organize the documents. We were practicing it first and then we were reading the theories on DH. In our experience theory followed practice. McGann's *Radiant Textuality*, one of the first really worthwhile books on DH became a source of theory and inspiration for this, but it really was about problem solving first.

Jones: What type of criteria was used to establish best practices?

Radcliffe: Trial and error, did it run or did it fail? There weren't standards back when we started. Standards evolved as we were working. With most of the work that we did, I'm not sure that we had standards at the time. SQL was the best database. TEI became a standard 15 years ago, unmarked texts and XML was huge. I started using as much XML as I could as it became a standard. What are the standards? Does it fail? Does it work?

There were two platforms at the time, Macintosh and Windows, or Microsoft before there was Windows. It was about deciding which platforms would work best for the project. And then the World Wide Web hit and that expanded the standards. You could start viewing projects through browsers, and everyone could see what was going on. HTML was very important as a standard. This may be ancient history, but I can't emphasize how pragmatic and practice-based this work was, and theory came after.

Jones: Were there any philosophical or practical aims when deciding how the work would be done?

Radcliffe: In so far as there were bibliographies and other projects, there were standards for accuracy. This included standardizing titles, dates, and more. You were making a project, and was there peer-review? No. Remember, none of this counted for tenure, so you did it because you wanted to and because it could help your research. This changed about 10 years ago, and then there was chatter and talk about DH.

Jones: When taking on or assisting with a new project, what is your methodology for determining what type of technology to use?

Radcliffe: Is it text or numbers based? It is something that can be done in a relational database or handled in XML? Is the object regular or irregular?

Jones: Then XML could be used for irregular projects?

Radcliffe: Yes, XML is for irregular data. Typical databases are for information and numbers in rows. Books or letters are irregular in form, including paragraphs and structure. Each record could be different, and TEI and XML were good for this. Databases are good for numbers, not documents.

Jones: How do you collaborate in the CATH?

Radcliffe: The CATH lab is really more an association of fellow laborers than a room. We work from home and then collaborate. We don't spend a lot of time in the lab.

Jones: Is technology a tool to read literature or an object to be explored?

Radcliffe: In my line of work it's more as a tool but I believe this may be a generational difference. For those of us who were working before the internet, it was more about building the tools and the practices. Now it is more about consumers than back when we were building the web. There was a different emphasis.

Jones: How important is promoting scholarly work through social networking in your line of work?

Radcliffe: I think for you it is terribly important, but for me, not at all. I think that for those in their 30's or 40's they are working furiously on this, but for me it is still email. The projects that I'm working on take tens of thousands of hours, and it would be difficult to complete those projects if I were spending 3 to 4 hours a day on social networking.

Jones: What does a typical project look like?

Radcliffe: Earlier, they were mostly about building databases. I've worked on a Spenserian database that has compiled bibliographies for the last fifteen years. I've worked on compiling Spenser's work, and then imitations of Spenser, and then imitations of imitations of Spenser, and so on. Technologically the projects are simple, plain text projects. For the last 10 years I've been working

on Lord Byron and his biographies, memoirs, and letters. These are marked up in TEI. My work is about reconstructing relationships, Byron's relationships. I'm trying to find family members, business partners, and more.

Jones: Is the end goal to find this information so that it can be shared and others can use it to find new information?

Radcliffe: Exactly. Really some of the new technologies may allow us to use NoSQL databases to dump the information into the cloud, so that all the information can be queried by clever people to find new factoids. It's to take all the documents and mark those using RDF triplets (little statements) and have them all tossed out to be read alongside other databases. This can allow research to be done across hundreds of thousands of documents, which is then better for graphing and creating visual images.

Jones: Is this a type of distant reading?

Radcliffe: No, this is hand work. It's different because it's not based on probabilities, but based on facts. It may be distant in that not everyone will read everything.

Jones: What visuals are you hoping to find?

Radcliffe: I'm creating this database for future people who are cleverer than I so that they can find Byron's mother, people who grew up in his mother's town, and then find his father, and this can be used for Venn diagrams. Our people could query those people in Worcester who commented on Byron's Childe Harold, and query that information to create graphs, nodes and edges, timelines, maps, or however they want to use it. I'm busy scooping up the data so that it can be used in ways that I can't imagine. But you have to have the data first. And it's hand work because you have to distinguish John Brown from John Brown, and Mary Smith from Mary Smith.

Jones: How do you find the next great tool in digital humanities? Is it through your IT department or do you do your own research? If you do your own research, what webpages/trade magazines or other publications do you use?

Radcliffe: I usually try to do the research, and I keep up with list serves, and try to keep aware of what's new. There is an increasing number of new technology, so that's difficult to do. I've also spent 15 years doing a project in a particular language into particular data fields that I use, so I'm not always eager to switch to a new technology. I try to make use of the time that I've done. XML is really useful because it can be transformed into anything else by someone like a developer who knows a lot about programming.

Jones: So then you try to find ways to make your work easily transferred or translated?

Radcliffe: Right.

Jones: We have been talking about C.P. Snow's "The Two Cultures" and how humanities and the sciences are often pitted against each other, or how humanities have the need to legitimize themselves. Do you deal with these issues of access and fluency? Have you encountered any issues dealing with access and fluency?

Radcliffe: As a philologist, my work is often wrapped up with technology. We're interested in facts and information, but for my colleagues working on the more traditionally English research, yes they

run into this. As digital humanists we're always spanning CP Snow's cultures. We think of ourselves as scientists. English comes out of a hermeneutic tradition and involves more interpretative work. DH always had more of an emphasis on making, or using physical objects to study. The emphasis is different. In English, when you're writing a peer-reviewed article, you're not doing the type-setting or the formatting. When you're doing DH, you do the whole bit. You make the formatting, the design, you self-publish, it's a whole different thing. I should say what I do is more humanities computing. DH is becoming more interpretative and more about hermeneutics.