

# Transcript

[00:00] [background music]

**Derek Bruff:** [00:01] Welcome to season two of “Leading Lines,” a podcast from Vanderbilt University. I’m your host, Derek Bruff, Director of the Vanderbilt Center for Teaching. We hope you enjoyed our first season, all 10 episodes of which are available on our website [leadinglinespod.com](https://leadinglinespod.com). We’re looking forward to more explorations of creative, intentional and effective uses of technology to enhance student learning here in season two.

[00:28] In this episode, we feature an interview with Kathryn Tomasek, associate professor of history at Wheaton College. Kathryn is interviewed by my colleague, Cliff Anderson, Associate University Librarian for Research and Learning here at Vanderbilt.

[00:40] Cliff met Kathryn at the 2016 Alliance of Digital and Humanities Organizations Conference in Krakow, Poland where Kathryn shared some of her work engaging undergraduate students in the digital humanities.

[00:50] Kathryn’s work focuses on transcription and markup of historical of texts. She and her students are active in TEI, the Text Encoding Initiative.

[00:59] In the interview, Kathryn discusses her experience of getting started with text encoding, the value of teaching all students how machines talk to each other and the role that text encoding can play in helping students engage in the kind of close reading that’s critical for historical analysis.

[01:13] [background music]

**Cliff Anderson:** [01:15] I am here with Kathryn Tomasek who is a Professor at Wheaton College, professor of history. We’re going to talk today about some of the really innovative

work that she's done using digital humanities, in particular the TEI, with undergraduates.

Welcome, Kathryn.

**Kathryn Tomasek:** [01:31] Hi, Cliff.

**Cliff:** [01:36] Why don't we start talking a little bit about your own background, how you got to where you are professionally and how you developed your interest in digital humanities?

**Kathryn:** [01:48] I started off studying something that didn't seem to have anything to do with digital humanities or digital history. I would say I was trained as an analog historian. I went to grad school at Wisconsin and I did a dissertation on women in Fourierist communities. Fourierism was a so-called the Utopian Movement of the 1840s. It has a lot to do with Transcendentalism. I was interested in women's work.

[02:18] Partly as a result of that and partly just because there was an opportunity, the interest in women's work -- I guess I would say -- I attended some workshops that were organized through NITLE (<http://www.nitle.org/>) and the Mellon Foundation in 2004, a million years ago. They were about using text encoding in the classroom.

[02:41] We learned some text encoding and we did a pilot experiment in a women's history class. My colleague in the archives, Zeph Stickney, had purchased the journal of the daughter of a Baptist minister from the period following the US Civil War. We used it in a US women's history class to give students an opportunity to encounter a primary source.

[03:21] We had them do some transcription and then we had them do some markup. The students loved it. They felt like they got to know Mariah Wood, and to really care about her life. They really wanted to know what happened in her life after the period of the journal, which I consider to be a real success because people from the past always seem to be fairly distant and not real people, or they can seem that way. This gave students an opportunity to identify with someone and get to know her and care about her.

**Cliff:** [04:01] Obviously, that forces a close reading of the text that students who probably don't have coming into college, don't necessarily have a background in that close reading.

[04:18] Maybe you can explain just a little bit for those who maybe listen to this and come into educational technologies from different perspectives, a little bit about what the TEI is

and how that plays a role in textual editing.

**Kathryn:** [04:34] The Text Encoding Initiative started out as a way to think about how to turn humanities' materials into data that computers could analyze. This was before there was good OCR and even people who might think that OCR is the way to go for turning text into something that's machine readable.

[05:12] For scholars in the humanities, particularly people coming from literary studies where there was a strong tradition of scholarly editing, OCR's didn't seem like it was ever going to be the thing, the right way to go. These folks developed a method...This was before HTML, can that possibly be the case?

**Cliff:** [05:42] That rings a bell for me. We'll have to check on this, but I believe SGML was what they were using at first, right?

**Kathryn:** [05:52] That's exactly right. Standard General Markup Language. What these folks found was that SGML was not adequate for expressing the kinds of characteristics that they wanted to express about texts.

[06:11] In fact, one of the people who was, at the time, a young person in the lead group of the TEI and who is now an elder statesman, Michael Sperberg-McQueen, was one of the people who was part of the group that developed XML, which is Extensible Markup Language. That's a vocabulary for data that's behind almost every application that ordinary people use in our daily lives, things like Excel and Word.

**Cliff:** [06:50] It brings to the point that these are technologies that are ubiquitous, but we don't necessarily see them. Part of the challenge is bringing them to the forum and using them in a critical way and a conscious way, right?

**Kathryn:** [07:05] Yes.

**Cliff:** [07:05] They structure the data in a way that isn't dictated by a tool, but is actually dictated by intellectual standards.

**Kathryn:** [07:13] Exactly, and it's not so much about the marketplace but about, as you say, intellectual standards.

[07:20] One of the things that I like about teaching students the TEI is that they get to see not just sort of all the things we can consume on the web, but they get to see a little bit of what's under the hood. That was more true before Steve Jobs and our friends at Apple came up with the iPad and we went into the world of apps.

[07:53] It's still really important because there is code underneath all of that stuff. The more that ordinary non-engineer folks can understand that stuff or at least have a little bit of a notion of what code looks like and how people and machines talk to each other, the better I think that is for what we call the Digital Age.

**Cliff:** [08:25] We've met previously, but we met in Krakow for the Digital Humanities 2016 Conference and you had brought some students with you and you held a seminar in advance for the conference.

[08:45] One of the things that I thought was fascinating in the way that you led that seminar was to walk through a sample TEI document and then just have the students say, "How would we mark this up?" and just iteratively add markup to it as students came up with the ideas.

[09:02] I have to say, as I've told you before, just so incredibly impressed by the students' ability to think about the multiple ways in you might enrich a text. So obviously, you've been teaching them a lot. It was a great display of their knowledge.

**Kathryn:** [09:18] Thanks. They're brilliant students. There were four of them and they were the students who worked with me on a summer research project for six weeks before we went to Krakow.

[09:31] It was a really great culmination of their summer research to have an opportunity to attend a workshop that was led by my colleague, Georg Vogeler, and then also to attend sessions of DH, so that they could get a sense of where some of what they were learning fit into the larger, international, scholarly universe.

**Cliff:** [09:58] One of the things about digital humanities, and I know that there have been a lot of efforts to make this less true and I'm sure it is less true than it used to be, but in the beginning, I think a lot of digital humanities research was clustered around large research universities, and mainly focused on graduate students.

[10:24] You're working at a liberal arts college and you're working with undergraduates. How have you been able to adapt these trends in DH for undergraduate teaching?

**Kathryn:** [10:36] First, I have to really credit Julia Flanders and the Women Writers Project, which is now at Northeastern University, because Julia is one of the people who, with her colleague, Syd Bauman who's a Senior Programmer at the WWP. One of the main jobs that Julia has taken on, and had been working on really powerfully in the first decade of this century, has been just teaching people the TEI.

[11:17] The workshops that I attended were ones that Julia taught and that were focused particularly on how we might use the TEI for teaching undergraduates. I never would've done any of this if, on the one hand, there hadn't been funding and on the other hand, Julia hadn't been there to do this teaching.

[11:46] Truthfully, Cliff, I jumped in because someone in our academic computing area said, "Oh, this would be kind of keen. Let's do this." I wasn't really someone who had done digital humanities before or any of those kinds of things, but what I discovered, and what was really appealing to me, was the way that asking students to do transcription and markup could teach something that's really a challenge to teach as an undergraduate instructor of history, that is how important it is to spend time with your sources.

[12:35] We teach students in the history department, or in the history major, we try to bring them along to a place where by their senior year, at some point, they can do original research and come up with a product, which we've traditionally thought of as a research paper, that is genuine history. That fits into the discipline.

[13:10] One of the things that I think is always challenging to help students understand is that necessity of spending time with your sources and becoming immersed in them. One of the things that transcription and markup, I hesitate to use this word, but forces is that close reading that's really necessary in order to be able to write a historical narrative.

**Cliff:** [13:48] I really appreciate that, especially as a former special collections librarian. That is a more difficult skill to teach than it seems. It's just because we are so used, in other contexts, to reading quickly and trying to skim through as much as we can, as quickly as we can, that to really slow down and to read carefully, it's a hard challenge for people.

[14:16] It's a hard challenge sometimes because of the digital technologies that we use, right? I don't think our ability to concentrate is necessarily being improved by a lot of the technologies. I've learned this, and we should also mention that you've been leading a research project in the last couple of years called MEDEA. Remind me what MEDEA stands for? I always stumble a little on this.

**Kathryn:** [14:44] It stands for Modeling semantically Enhanced Digital Edition of Accounts. The idea is that there are an embarrassing number of account books lying around in archives that we haven't done a lot with. If I say it that way, then I run into the danger that social and economic historians will jump up and down and say, "But we've been sampling that stuff for a long time."

[15:17] That's true, but, for instance, documentary projects like the papers of people like Thomas Jefferson and George Washington and John Adams have avoided work with the account books. Documentary projects are about transcription of the vast papers of people like the Founding Fathers.

[15:49] They've avoided them because they seem too complex, partly because there's a lot of formatting that goes into a ledger, for example. It's actually very challenging both in print and digitally to represent the table format.

[16:11] I don't know if you've ever tried to make a website that had a table on it. It's not an easy thing. I was going to say it used to not be, but I don't want to sound like the old lady in the room. They seem too complex to represent fully, like in a full transcription.

[16:39] The fact is that they're incredibly rich. There is all kinds of information in account books. Not just about money and prices, which has been the focus of economic historians, but about people and how they interact and their daily lives.

[16:59] Some colleagues and I are interested in coming up with some best practices for transcription and markup of account books so that it would be possible to do some comparisons across time and space. That's really what MEDEA is about.

**Cliff:** [17:18] Having participated in that project myself and coming to terms with the diversity of ways in which this financial information was recorded, the real challenge is when you're crossing such large historical periods and geographic locations, what you realize is

that there isn't going to be a single standard. There may be multiple overlapping standards for different purposes.

[17:50] Again, it's kind of what you want to achieve and what you want to get out of working through these sources. Although, I think at least one of the takeaways I took from this is that establishing a good base edition with something like the TEI gives you the flexibility then to extract data from that in different formats. The TEI works really well in that context as a sort of, would you say, diplomatic transcription, a way to do diplomatic transcription of these sources?

**Kathryn:** [18:27] Yeah, the diplomatic edition is one product that then if you have a good edition, it's possible to extract different bits of data based on what your research question is.

[18:45] One of the things that my colleague, Georg Vogeler does at the digital repository that he hosts at Karl-Franzens University in Graz in Austria, in this repository, what they host are editions that have been made in TEI and then there are various kinds of exports that you can do. You can export the XML if that's what you want. You can also export into Excel if what you want is the numerical data.

[19:27] That's actually a very important innovation that comes with thinking about account books from the transcription perspective, the perspective of producing an edition as opposed to the sampling perspective that is the focus of social science history.

[19:52] With the edition, you present the entire account and then allow folks who come after you to make the decision about what to select out of the account as opposed to making the selections for your series -- for instance, a price series or something like that -- and then leaving behind data that has already been affected by the choices of a particular set of research questions.

**Cliff:** [20:22] That was actually maybe even a binding for a lot of the researchers who participated when they were in the room with researchers who had different interests and said, "Well, why did you make this decision in advance because this data could have been very useful to me."

[20:40] It's partly that we all have our disciplinary blinders on and that means some of the choices we make might not be the right choices for other people who are doing different

types of research.

**Kathryn:** [20:53] That's one of the really important takeaways from the folks who use text encoding, is that the goal is to produce an edition that one can use for one's own research and then leave behind something that another researcher can use for their own purposes that you or I can't even imagine at the moment.

[21:22] There are a lot of people in the TEI who think that's really the most exciting thing, is that you can make something, make your own conclusions out of it, make your own arguments, ask your own research questions, and then leave something that is complete in a certain kind of way that someone else can come and put together with something that you would never think of putting it together with and make something new a long way down the line.

[21:53] That's one of the really exciting things about digital humanities, I think, is that the possibilities of the reuse of the information in interesting ways that you would never think of.

**Cliff:** [22:07] As we wind up our conversation — this is such a rich, fascinating conversation — I want to bring it back to the students again and just ask you, maybe, when you're teaching students, where do they have the hardest time with the TEI? What are the stumbling blocks that you see?

**Kathryn:** [22:31] I try really hard to remember that the students that I teach are not necessarily students who would encounter any kind of computer code anyplace else, because they might, for instance, have what's referred to as math anxiety, and I'm not particularly keen about that, that the notion of something like math anxiety, I don't necessarily think that's the real thing.

[23:05] I do think that sometimes students are a little put off, a little scared of the angle brackets just because they look weird, they're unfamiliar, even though one of the things about the TEI is that it's meant to be human readable as well as machine readable.

[23:28] There are students who are just the tiniest bit freaked out by something that doesn't look like words that you see printed on a page or a screen. One of the cool things, one of the ways to deal with that is that the editor that we use, which is Oxygen, actually has two modes. There's one, a mode called author mode and one called editor mode.

[23:55] If students are freaked out by the angle brackets, they can do the initial transcription in this author mode, which is a cleaner mode that allows them just to see the letters and numbers on the screen.

[24:15] That's the place where students face the biggest challenge and where I'm really pleased to say that the tool allows us to deal with the challenge in a useful way.

**Cliff:** [24:34] I can also say we've met similar challenges even working with librarians who are unfamiliar with XML and, with a little help from CSS and the author mode, you can go a long way of easing people into things. Then they learn if, oh, they need to go in and make some changes on the markup, they can switch and they get their toes in the water and it becomes more of a familiar environment to them.

**Kathryn:** [25:02] Exactly.

**Cliff:** [25:06] Again, this has been a really wonderful conversation. We always end by asking our guest, what is your favorite analog technology for educational technology?

**Kathryn:** [25:16] Discussion.

[25:16] [laughter]

**Kathryn:** [25:25] Am I allowed to just use one word?

**Cliff:** [25:28] That's a great word. Are you thinking seminar style conversing?

**Kathryn:** [25:31] Yeah, I've never been a person who was particularly comfortable with lecture. I've always felt like the classroom works best when it's interactive. Surely you know this, Cliff, you have to have a plan for what's going to come out in a class meeting, no matter what. You can achieve that by lecturing for exactly 80 minutes and feeling really great when you finish whatever it is that you were going to say at exactly the 80-minute mark. That's one option, right?

[26:20] But I really believe in active learning and I really think that students learn the most when they are actively engaged in thinking things through. I really like even when it may look like I'm lecturing, I really like bringing the students in and using discussion.

[26:48] I don't just mean the Socratic Method, because I hate the "What is teacher thinking" game, but getting the students talking and giving them an opportunity to come up with ideas on their own is my favorite thing to do.

[27:07] The fact is that transcription and markup and all these other things that I do are just a different way to do that.

**Cliff:** [27:15] That's a perfect place to end this conversation. Thank you again. This has been so lovely to talk with you. I've learned a lot. There will be show notes about many things that we talked about, too.

[27:29] Thanks again, Kathryn, really appreciate it.

**Kathryn:** [27:29] Thanks for having me, Cliff.

[27:31] [background music]

**Derek:** [27:31] That was Kathryn Tomasek, associate professor of history at Wheaton College. In the show notes, you'll find links to Kathryn's home page and some of the projects she mentioned in her interview.

[27:40] You can find those show notes on our website, [leadinglinespod.com](https://leadinglinespod.com). We welcome your comments and questions there and on Twitter, where our handle is @leadinglinespod. You can subscribe to our podcast on iTunes or your other favorite podcast app.

[27:53] I've heard several interviews we have on deck for season two and I do think you'll want to subscribe.

[27:59] Leading Lines is produced by the Center for Teaching and the Vanderbilt Institution for Digital Learning, the Office of Scholarly Communications and the Associate Provost for Digital Learning, all at Vanderbilt University.

[28:09] Look for new episodes the first and third Monday of each month. I'm your host, Derek Bruff. Thanks for listening.

