

Qualitative Data Reuse: Records of Practice in Educational Research and Teacher Development

A nation of learners requires a nation of teachers adept in different instructional methods who are sensitive to diverse learning styles and able to adapt instruction to address the needs of all students. It also requires researchers and other education professionals who are able to identify exemplary practice, develop and test innovative instructional methods, and provide supports and resources to help teachers improve student learning. To achieve these outcomes, the University of Michigan Schools of Information (UMSI) and Education (SOE) are collaborating on a research project to examine the reuse of qualitative data for teacher education and research. Our work centers on the Teaching & Learning Exploratory (TLE), a qualitative data archive of classroom “records of practice” that include videos of classroom teaching, recordings of content area interviews with students, and work samples from students and teachers. One important focus of the TLE’s collections is on math education. These collections are used by educational researchers, instructors in teacher education programs, as well as by others involved in the study and improvement of teaching. We view the TLE as an ideal setting to examine the qualitative data reuse and to better understand the curation practices required to address these challenges and facilitate reuse. Our three year project covers December 1, 2014 to November 30, 2017.

1. Statement of Need

Our project focuses on qualitative data curation and reuse. Qualitative data presents challenges to both data reusers and data curators. For data reusers, common challenges often arise around issues of transparency in the research design, having enough contextual information to make sense of the data (Atici et al., 2012), the co-creation of the data by researcher and participant, and the perspective that there is no such thing as ‘raw’ qualitative data because they are always filtered through the researcher and embody a degree of interpretation (Carlson & Anderson, 2007). Data curators face particular challenges in dealing with qualitative data, such as the inability to anonymize the data without losing a significant amount of meaning and compromising the data’s value (e.g., Corti, 2007; Cliggett, 2013), uncertainty concerning how much contextual information is needed to make sense of the data (e.g., Faniel & Yakel, 2011), varied and multiple metadata schemas or ontologies covering different disciplines (e.g., Faniel et al., 2013), fewer shared professional norms or history for handling and working with qualitative data (e.g., Corti, 2013), and the sheer variety of file format types (e.g., Carlson & Anderson, 2007; Cliggett, 2013). Qualitative data often includes proprietary file formats and very large multimedia files. In summary, both qualitative data reusers and curators face challenges when dealing with qualitative data and we have identified this as a gap in the research that needs to be addressed.

In this proposal, we focus on qualitative data reuse by examining collections of educational “records of practice”¹ and how they are best curated and utilized to improve educational outcomes. These forms of documentation – especially video records of practice – have been used heavily to improve teaching (Bacevich, 2010; Brophy, 2003; Brouwer, 2011; Villegas-Reimers, 2003; Wang & Hartley, 2003). In our project, we study qualitative data reuse in the context of the Teaching & Learning Exploratory (TLE), and the curatorial practices necessary to properly manage, preserve, and facilitate use of the data. We address the needs and challenges of three distinct audiences in this project: 1) educational researchers, 2) teacher educators², and 3) data curators. The first two audiences are data reusers interested in teaching and learning. Their goals in using the collections are to improve instruction and enhance learning outcomes. In addition, for the teacher educators, this includes

¹ Records of practice are “a detailed documentation of teaching and learning... taken directly from teaching and learning” (National Research Council, 2002, p. 79). They include videos of teaching and learning, documents and objects created during and for lessons, and work samples from teachers and students.

² Throughout this proposal we are defining “teacher education” broadly to include pre-service teacher education, early support of new teachers (induction), and ongoing in-service professional development for experienced teachers. Likewise, we consider “teacher educators” to include those who provide educational experiences for teachers at all levels of the professional continuum.

not only improving their own instruction but also the teaching of the pre-service and in-service teachers with whom they work. Thus, we are interested in understanding the questions teacher educators ask, their patterns of data use, and the ways in which the TLE resources can be improved to increase access to and understanding of the collections. The third audience, data curators, is interested in the best ways to curate and disseminate qualitative data. There are few examples of this activity in the literature and it is often unclear what lessons from social science and scientific data curation can be applied to qualitative data. Our work will provide an in-depth look at the reuse of qualitative data and identify the curatorial activities that most affect reuse.

Data Reusers. The first two audiences – educational researchers and teacher educators – require access to the records of practice to use as evidence and examples. Educational researchers have diverse information needs including developing reliable evaluation systems, studying teachers and teaching to understand the challenges and possibilities of different facets of practice, and analyzing the design, enactment, and impact of educational materials and activities. Teacher educators use these materials in courses and workshops to exemplify or examine practice, help novices learn to critique teaching and evaluate learning, and increase the pedagogical tools and “moves” teachers have to enhance learning. These uses may be as straightforward as selecting and playing a short video segment of a teaching technique or they may comprise an activity involving multiple records of practice that deeply engage novices with the materials. Although this interest in video records of practice is not new, it is playing a more prominent role in teacher professional education, evaluation, and feedback. This is partly due to a greater focus on *practice* and *skill* in teacher preparation (U.S Department of Education, 2011) where video records of practice are often used in training and assessment (e.g. Boest, Sleep, Ball & Bass, 2011; Pecheone & Wei, 2011). There is growing use of classroom videos in the evaluation and measurement of teaching quality (e.g. Bill and Melinda Gates Foundation, 2013; Ho & Kane, 2013), teacher professional development (e.g. Borko, Jacobs, Eiteljorg & Pitman, 2008; Santagata, 2009; Seago, Jacobs & Driscoll, 2010; van Es & Sherin, 2010), and professional growth (e.g. Helms, 2001; Lustick & Sykes, 2006). While we know these audiences are interested in reusing the records of practice, we do not know what their needs are in such areas as discovery and retrieval, what types of contextual information they need, or what tools would help them move from the data in the archives to an end product – whether that is a scholarly article or a course assignment.

Data Curators. Our third audience is data curators who have few examples of qualitative data curation upon which to draw – even fewer within the specific context of the United States. We see this audience as one that will be deeply interested in the policy framework of the TLE that facilitates data reuse, the collection policies around which records of practice are provided, and findings about what constitutes a useful dissemination information package. Our results will highlight similarities and differences in curating qualitative data (from our study) with what is known from the literature about curating data in the social sciences and the sciences.

Researchers have paid little attention to the curation of qualitative data and even less to its reuse. Corti (2006, 2007, 2011, 2013), at the United Kingdom Data Archives, has established curatorial norms and elucidated many of the special considerations of dealing with qualitative data, particularly in the UK context. However, she has not studied the actual experience of data reusers. Parry and Mauthner (2004), Lin (2009), and Cliggett (2013) all discuss the data management issues associated with qualitative data, such as confidentiality, ownership, and anonymity. Cliggett also mentions the importance of repository processes, such as metadata and preservation management. All of these authors admit that qualitative researchers are wary of data sharing but none demonstrate a good understanding of the dynamics of qualitative data reuse. Carlson and Anderson (2007) have come the closest to understanding qualitative data reuse in their comparative case study of four projects across the qualitative-quantitative spectrum. Regarding qualitative data, they conclude that “the one who collected the data and the one who interpreted them were the same person, and this had implications for the potential to meet data reuse requirements, because many assumptions, procedures, processes, and decisions often remained

undocumented tacit knowledge” (Carlson & Anderson, 2007, p. 646). We need to know more both to advance knowledge in different disciplines and to properly curate these data.

While there is limited research on qualitative data curation and particularly reuse, there has been considerable work on data reuse in other disciplines from which we can draw. Research on quantitative social science (Niu & Hedstrom, 2009; Faniel et al., 2012; Kriesberg et al., 2013) has shown the importance of research design, codebooks, and other associated materials, such as publications, in helping data reusers make sense of survey data. Zimmerman’s (2003, 2008) studies of ecologists demonstrated the unifying function of standards and the importance of trust in the data producer. Faniel and Jacobsen’s (2010) research on earthquake engineers and Vertesi and Dourish’s (2011) study of robotics engineers identified both technical and social issues that help and hinder reuse. These studies have given us a wealth of knowledge about how to uncover disciplinary practice, the importance of contextual information (albeit varied) required for data reuse, and the key roles of data producer reputation, publications utilizing the data, and repository trust. The research also shows that certain factors are important in *certain* disciplines while others are important to varying degrees in all, but we have little insight into how these apply to disciplines using qualitative data. We seek to fill this gap.

All three audiences will benefit from this project in different ways. Our findings will help facilitate use of the collections by educational researchers and teacher educators, as we understand more about the qualitative data reuse process. We will also make appropriate changes to the TLE curatorial processes and website that specifically address issues uncovered in our research. Furthermore, by collecting records of practice of the teacher educators using the TLE, we will be able to provide these as part of the TLE collections. Finally, our research will benefit curators of qualitative data by providing more insights into the use of qualitative data – particularly video data – and identify those curatorial actions that most facilitate data reuse.

2. Impact

Our research will have two major impacts: one for the library, archival and museum communities and the other for teacher educators. We will contribute to the library, archival and museum communities by filling a gap in knowledge about qualitative data reuse and how data reusers can best utilize qualitative data. There is little previous research in this area and what exists focuses on the issues from a curatorial perspective. We will focus on the perspective of the data reuser and inform curatorial practice with the data reuser in mind.

We will contribute to teacher education efforts by examining downstream impact. This unique aspect of our research will study pre-service and practicing teachers’ (i.e., the students’) learning opportunities and their uptake from secondary exposure to and reuse of TLE materials. This study of secondary reuse is important for assessing the impact that reuse may have on K-12 classroom teachers and its potential for influencing teaching.

Evaluation. We will use several strategies for evaluation: 1) Creswell’s (2009) “sequential explanatory strategy,” for data collection (discussed in the next section), 2) benchmarking against our project timeline (in section 4 below) and the Schedule of Completion, and 3) input from an external advisory board. Using each of these approaches, we will measure and analyze our research project’s outputs, outcomes, findings, and products.

Research Design. Using Creswell’s (2009) “sequential explanatory strategy,” will allow us to continually evaluate the effectiveness of our research methods in two ways. First, we will be able to ensure that the research design builds toward achieving the benefits (below) for our three audiences. For data curators, our outputs will be publications and presentations that contain recommendations concerning the curation of qualitative data. Given the paucity of research in this area, we can make a unique contribution to professional practice in this area. For educational researchers and teacher educators, our output will be implementing changes to the TLE to

facilitate reuse of the educational records of practice. Our impact goal is to facilitate the work of curators and researchers/teacher educators. We will assess this by the responses of curators (invitations to speak, citations) to our work. For researchers and educators we will be able to monitor changes in research patterns through transaction logs, use of resource materials and supports we post on the TLE site, and the number of users on the site. Another benefit of the “sequential explanatory strategy” is that we will regularly be forced to articulate how each step relates to the previous and subsequent ones in the research design, thus ensuring continuity and increasing our ability to build an appropriate base of evidence to answer our research questions.

Project Timeline. We have developed a realistic project timeline/research plan that we will use as a measurement tool to keep us on schedule. Using this timeline, we will be able to set achievable goals for each term, compare our progress against it, and make appropriate adjustments with the larger project in mind.

Advisory Board. We will use our external advisory board as an evaluation and review team for all project instrumentation and evaluation processes. Our advisory board contains one member with expertise in evaluation and others adept in repository management, teacher education, and educational research. We will use their expertise to ensure that our research design and evaluation plans are appropriate given our goals.

3. Project Design

The goals of this project are to understand the dynamics of data reuse of digital qualitative data, contribute to the discourse and understanding of the best means for curating qualitative data, and to use our findings to improve the curation and services available to reusers at the Teaching & Learning Exploratory (TLE) and beyond. As such, our research questions encompass two areas:

1. Data Reuse: What are the dynamics of the data reuse lifecycle (from selection of data through the reuse of data) in a qualitative digital educational archive? We will formulate a series of sub-questions to address the different stages of the data reuse lifecycle for educational researchers and teacher educators.
2. Data curation: What special issues are involved in curating digital qualitative data for reuse at the TLE?
 - How can qualitative data archives best support TLE data reusers throughout the data reuse lifecycle?
 - What aspects of the TLE’s experience are informative for other types of qualitative data archives?

Site selection. The Teaching & Learning Exploratory (TLE) is a qualitative data archives of classroom “records of practice,” including videos of classroom teaching, interviews with students and teachers, examples of student work, and teacher-generated lesson materials. Records of practice are an important resource for understanding teaching, presenting “a detailed documentation of teaching and learning... taken directly from teaching and learning” (National Research Council, 2002, p. 79). We have selected the TLE for several reasons. First, the TLE is a unique qualitative data resource repository. Unlike many other online content providers that deliver videos of K-12 classrooms for use in educational settings (e.g., Annenberg Learner, the National Board for Professional Teaching Standards (ATLAS), TIMSS Videos), the TLE holds a variety of collections created by many researchers and organizations rather than a single set of materials produced and delivered by one group. In addition, TLE collections are intended for reuse by both educational researchers and practitioners, rather than focused exclusively on researchers (like the Inter-university Consortium for Political and Social Research) or on teachers (like the Teaching Channel, Inside Mathematics, and the content providers mentioned above). One thing that makes this diverse use possible is the priority the TLE places on creating collections of *minimally* edited records of practice that leaves editing decisions with reusers whenever possible. Also, TLE collections have a strong focus on mathematics education, a critical component of STEM education. Second, there are few qualitative data archives in the US. Third, the TLE holds a nice variety of digital materials in its collections of records of practice – including videos of professional practice, a range of classroom artifacts (e.g. homework

papers, lesson plans, photos of lesson preparations and products), and various supports for using the classroom documentation (e.g. seating charts, video transcripts). Fourth, the data can be reused in diverse ways, by researchers and practitioners – not only including study and analysis but also course activities, seminar presentations, and integration into educational materials. TLE reusers cross the research-teaching divide.

Collections currently being made available through the TLE have been available for years on-site at the University of Michigan School of Education or through special arrangements with the data producers. The TLE is now making these materials available through a website to promote broader use. Currently, the TLE is piloting the archives with several school districts, teacher educators, and researchers. The formal launch will be in the spring. The initial featured collection is the Measures of Effective Teaching Extension (METX) collection. It will be joined over the summer with the Grand Rapids Elementary Mathematics Laboratory (GREML) collection. These two collections represent great variation. The METX collection consists of classroom video and scanned lesson artifacts (teacher lesson plans, worksheets and handouts, homework assignments, and images of board work) captured during the 2011-2012 and 2012-2013 academic years. The METX project enlisted more than 350 teacher volunteers from six school districts, collecting video and other artifacts of everyday classroom activities that document thousands of lessons. The recorded lessons focus on English Language Arts and Mathematics classroom topics but other topics also appear. In addition to object-level metadata describing the lesson and classroom context for the video, time-based metadata was also applied to the videos using a scheme based on the Common Core State Standards (Common Core State Standards Initiative, 2010) and the TeachingWorks High Leverage Practices (TeachingWorks, n.d.). GREML is a more in-depth dataset featuring a comprehensive and diverse set of artifacts from a week-long laboratory including all student work and assessments, lesson plans, handouts, transcripts, seating charts, plus 40 hours of video. By December 2014, these collections will be joined by others documenting additional aspects of teaching.

Data collection. We have developed a mixed methods design for data collection in order to study qualitative data reuse because it allows “for the broad purposes of breadth and depth of understanding and corroboration” (Johnson, Onwuegbuzie, & Turner, 2007, p. 123). We will utilize surveys, interviews, observations and web analytics to understand data reuse through a number of different lenses. Furthermore, Morse and Niehaus (2009) note that mixed methods are most appropriate in complex situations with diverse stakeholders. By using this mixed methods approach, we will also be able to employ a “sequential explanatory strategy” (Creswell 2009, p. 211) where each data collection method builds on its predecessors.

Survey of TLE data reusers (baseline). We will conduct a baseline survey of all researchers and teacher educators who have used the TLE’s collections. Our survey goal is to uncover research questions or learning objectives, basic patterns of reuse, the types of materials reused, and what supplementary information (either provided by the TLE or not) was used to understand the collections. The survey will also inform our interview protocols. All TLE registered users will receive the survey as well as those who have previously published or created other products from the collections. We will use Qualtrics, an online survey application freely available at the University of Michigan, to collect this data.

Interviews. After the high-level survey, we will complete 25 in-depth interviews with researchers and 25 with those using the materials in teacher education. In the latter group, our goal is to interview 10 teacher educators that select and show videos to pre-service or practicing teachers and 15 that set up more intensive learning activities using the TLE materials. The interview protocol will draw from the existing literature on data reuse as well as findings from the survey of TLE users and the web analytics.

Web analytics. Throughout the project we will collect web analytics and analyze the transaction logs looking for specific behaviors (paths through the site), discovery processes (search terms/use of the

tags), and downloads of materials. We will also revisit the logs and our analytic efforts periodically to ensure we are collecting the data we need for our analysis.

Using the survey, interviews, and web analytics, we will be able to examine some phenomena, such as discovery and data selection from three perspectives across the two main user groups. This will provide invaluable evidence to inform changes to the website and the creation of tools to facilitate use.

In addition, we are particularly interested in data reuse by teacher educators as well as the impact these instances of reuse may have on pre-service and practicing teachers. Therefore, we will engage in two additional methods of data collection with this group.

Observations. We will conduct 10 video observations of teacher educators using the TLE’s records of practice in an educational context (both pre-service and professional development). We will supplement the video observations with live visits when possible. As with the interviews, we will strive for a mix in the observations so we see teacher educators who just play videos *for* their groups and those who engage their learners more directly in working with the records of practice. Along with the observations, we will collect lesson artifacts developed during and for these activities (e.g. lesson plans, class products) and analyze these as part of the observation. This will assist us in determining what associated documentation should be digitized with or produced for the video. For example, early reusers of video records of practice, regularly made seating charts and transcripts to support their work so, for the GREML collection, the TLE created these and provided them as a resource for data reusers. At the end of the project we will include the observational data and artifacts as a collection within the TLE so teacher educators can see some of the diverse ways the materials have been used.

Survey of pre-service and practicing teachers. Pre-service and practicing teachers will be the *students* in the teacher education settings utilizing the TLE resources. In order to better understand the “secondary” impact of reuse we will survey these individuals. We are particularly interested in this group because teacher learning and uptake is one thing that might directly influence – and make a difference in – what takes place in K-12 classrooms. A modest incentive will be provided for pre-service and professional teachers who fill out the survey. We will again use Qualtrics to collect this data. We see this element in the research design as innovative: It pushes us toward identifying the types of outcomes we would like to see associated with data reuse.

Participant recruitment. We will use a variety of methods to recruit participants, including solicitations of TLE registered users, appeals at professional meetings, snowball sampling of interviewees, and those who have used and cited the collections in the past (not the data producers).

Data analysis. Data analysis will take several forms. For the surveys, we will export our data from Qualtrics in CSV format, clean and recode the data and ingest it into SPSS or R for statistical analysis. In the case for the interviews and observations, we will use NVivo a qualitative data analysis tools to code the transcripts and then analyze them. NVivo allows for the incorporation of other types of artifacts into the dataset and this will be used in the case of the observations. Our goal is to maintain continuity across all these data points by using many of the same coding concepts across all the different data in order to amass the evidence we need to move into the final phase of this project: making changes to the curatorial processes and website of the TLE.

TLE incorporation of findings. In year 3, we will analyze our data and create a plan for revisions to the TLE curatorial processes and website. We cannot anticipate the exact changes that will be needed, but our goal is to let the data guide us to make changes to improve curatorial processes that facilitate reuse and website changes

that ease the discovery, selection, analysis, and creation processes for data reusers. For example, we may decide that the existing metadata does not meet the needs of users and add a schema, such as the *IEEE Learning Object Metadata Standard* (Campbell, 2007) or the Learning Resources Metadata Initiative (LRMI).

3. Diversity

There are two ways in which we have incorporated diversity goals into this project: 1) through our stated audiences: educational researchers and teacher educators and data curators and 2) through the pre-service and professional teachers exposed to materials in the TLE. Our major audiences span two disciplinary communities: education and data curation. We will address each community when disseminating our results and our work will have an impact in each of these communities. Also, the records of practice housed in the TLE represent diverse communities in terms of socio-economic, racial, and geographic orientations. For example, the 350 classrooms documented in the METX collection come from six states in different regions of the US, providing a cross section of the US population. Participating districts represented a wide range of socioeconomic status (SES) levels as indicated by (a) percentages of students receiving free or reduced school lunch and (b) whether or not schools were receiving Title 1 funding. The racial composition of the student bodies also varied greatly as represented in percentages of White, Black, Hispanic, Asian, and “other” minority groups reported for each school by the National Center for Education Statistics. These statistical factors are listed with each video in the METX collection, although these variables are not currently searchable in the system. Our research will determine when and how these (and other) factors allow educational researchers to consider and explore different facets of diversity in the selection and interpretation of materials. Likewise, for teacher educators, we hope to better understand how the diversity of available classroom teaching samples enables them to make a range of classrooms accessible to pre-service and practicing teachers. The TLE collections have the potential to expose teachers to diverse communities and practices to which they might not otherwise have access. In our survey of secondary reusers of the TLE resources, we will inquire about their reactions to this diversity.

4. Project Resources: Personnel, Time, Budget

We have assembled a team of researchers and an advisory board with varied expertise to carry out this project.

Elizabeth Yakel, Ph.D. (Library and Information Science, University of Michigan) is a professor at the University of Michigan School of Information. She will be the Principal Investigator for the project. Previously, her research focused on use and users of traditional archives, and that research increasingly moved into the digital realm, focusing on archives 2.0 and interaction with users. Most recently, Dr. Yakel has focused on data reuse in different disciplinary communities and served as co-PI of the IMLS-funded Dissemination Information Packages for Information Reuse Project (<http://dipir.org>). She has led grants from IMLS, the Mellon Foundation, and the National Historical Publications and Records Commission and has demonstrated ability to guide grant-funded research projects. Dr. Yakel’s regular job responsibilities include research and some of this time (12%) has been designated as cost share in years 2 and 3.

Kara Suzuka, Ph.D. (Education, Michigan State University) is a research scientist at the University of Michigan School of Education, Teaching & Learning Exploratory (TLE) where she leads the research and development for the digital repository. She will be the co-Principal Investigator. Her primary work focuses on developing practice-based educational materials and approaches for use in teacher education and professional development. Her research and materials development work involve studies of classroom practice, utilize multimedia records of classrooms, and center on efforts to engage teachers in teaching practice. On the Qualitative Data Reuse Project, Dr. Suzuka will assist in data collection and analysis and lead the application of the findings to the TLE in year 3. She will also help to supervise the doctoral and master’s students. This project

is close to her work in the TLE and its research findings will contribute to her existing efforts to further develop the TLE and its collections. This project can easily integrate with her current duties and responsibilities that include conducting research and some of this time (10%) has been designated as cost share in each year.

The research will also incorporate doctoral and master’s students as integral members of the research team. Drs. Yakel and Suzuka are deeply are committed to making this a rich educational experience for the students and will work closely with them to help develop and hone their research skills in the course of this project. Both students will be involved in data collection and analysis, although the doctoral student will have a greater role. The master’s student will be responsible for project website maintenance and bibliographic monitoring. Part of the tuition for the doctoral student will be paid by UMSI and is designated as cost-share.

Advisory board. We have assembled an advisory board of four individuals with varied specializations with whom we can consult on various aspects of the project. These individuals are:

Tim Boerst, Ph.D. Dr. Boerst is a professor of practice at the University of Michigan School of Education and Co-Chair of the Elementary Teacher Education Program. His scholarship focuses on mathematics teaching and practice-focused approaches to professional learning – including work with classroom records of practice. Boerst’s expertise will be important in recruiting and studying teacher educators and educational researchers.

Eric Kansa, Ph.D. Dr. Kansa is the director of Open Context, a digital archaeological data repository. As such he deals with qualitative and quantitative data in a variety of formats. He has particular expertise in ontologies and metadata management and will be essential in year 3 as we apply our findings to the TLE.

Nanette Seago. Ms. Seago is a project director at WestEd, a non-profit public research agency in California focused on improving education. Her work centers on teacher professional development in mathematics. She was involved in the earliest efforts to reuse classroom videos and other records of practice (i.e. the TIMSS-R Video Study) and continues new work in this area. We will draw heavily on her expertise as we study practicing teachers.

Ann Zimmerman, Ph.D. Dr. Zimmerman is an expert in scientific collaboration, including interdisciplinary research, and the influence of new technologies on the practice and data reuse. She has examined data reuse in many disciplines including ecology and materials science. Dr. Zimmerman is also an expert in evaluation and a member of the American Evaluation Association and the Michigan Association for Evaluation.

Project Timeline.

Timeframe	Project Activities
December 2014	<ul style="list-style-type: none"> ● <u>Project start-up:</u> Human subjects application, hiring students, website design
January – April 2015	<ul style="list-style-type: none"> ● <u>Survey of TLE data reusers:</u> Develop, pilot, deploy survey; begin data cleaning; analyze ● <u>Interviews:</u> Develop interview protocol based on comprehensive literature review ● <u>Web analytics:</u> Set up and begin collecting web analytics data from the TLE ● <u>Website/Social media:</u> Launch; continue population and dissemination

<p>May – August 2015</p>	<ul style="list-style-type: none"> • <u>Survey of TLE data reusers</u>: Feed survey and web analytics results into the interview protocol; analyze survey; draft findings for dissemination • <u>Interviews</u>: Finalize interview protocol after consideration of survey and web analytics data; pilot interview protocol with both educational researchers and teacher educators; begin recruiting; commence interviews; transcribe, check transcriptions • <u>Web analytics</u>: Continue collection of web analytics data; assess collection strategy and revise as needed • <u>Website/Social Media</u>: Continue population and dissemination
<p>September – December 2015</p>	<ul style="list-style-type: none"> • <u>Survey of TLE data reusers</u>: Finalize data analysis and papers for dissemination • <u>Interviews</u>: Continue interviews; continue transcription and checking; begin drafting codeset for interview analysis • <u>Observations</u>: Develop and pilot observational protocol based on interview/survey data • <u>Web analytics</u>: Continue collection of web analytics data • <u>Website/Social Media</u>: Continue population and dissemination
<p>January – April 2016</p>	<ul style="list-style-type: none"> • <u>Survey of TLE data reusers</u>: Finalize papers and disseminate findings • <u>Interviews</u>: Finish interviews; continue transcription and checking; begin coding • <u>Observations</u>: Recruit teachers for observations; conduct pilot then actual observations • <u>Survey of secondary users of the TLE collections</u> (student or in-service teachers): Develop, pilot, and begin to deploy survey with observations • <u>Web analytics</u>: Continue data collection; assess collection strategy, revise as needed • <u>Website/Social Media</u>: Continue population and dissemination
<p>May – August 2016</p>	<ul style="list-style-type: none"> • <u>Interviews</u>: Continue coding interviews • <u>Observations</u>: Continue observations • <u>Survey of secondary users of the TLE collections</u> (student or in-service teachers): Continue administering survey with observations • <u>Web analytics</u>: Continue collection of web analytics data • <u>Website/Social Media</u>: Continue population and dissemination
<p>September – December 2016</p>	<ul style="list-style-type: none"> • <u>Interviews</u>: Finish coding and begin dissemination of results • <u>Observations</u>: Continue observations; develop coding scheme and begin analysis • <u>Survey of secondary users of the TLE collections</u> (student or in-service teachers); begin survey analysis • <u>Web analytics</u>: Continue collection of web analytics data • <u>Website/Social Media</u>: Continue population and dissemination
<p>January – April 2017</p>	<ul style="list-style-type: none"> • <u>Integrate data from the surveys of data reusers and secondary users, interviews, observations, and web analytics to inform TLE development and disseminate to the educational and curation communities</u> • <u>Observations</u>: Finish observations; finish coding observations; begin preparing observations and records of practice for incorporation into the TLE • <u>TLE</u>: Identify important findings; plan revisions for TLE curation processes and website • <u>Web analytics</u>: Continue collection of web analytics data; reassess collection plan • <u>Website/Social Media</u>: Continue population and dissemination

<p>May – August 2017</p>	<ul style="list-style-type: none"> • <u>Integrate data from the surveys of data reusers and secondary users, interviews, observations, and web analytics to inform TLE development and disseminate to the educational and curation communities</u> • <u>TLE</u>: Implement plan changes into curation process and website • <u>Observations</u>: Continue preparing observations and records of practice for incorporation into the TLE • <u>Web analytics</u>: Continue collection of web analytics data • <u>Website/Social Media</u>: Continue population and dissemination
<p>September – November 2017</p>	<ul style="list-style-type: none"> • <u>Integrate data from the surveys of data reusers and secondary users, interviews, observations, and web analytics to inform TLE development and disseminate to the educational and curation communities</u> • <u>TLE</u>: Finalize changes • <u>Observations</u>: Finalize incorporating observations and records of practice in the TLE • <u>Write final report</u> • <u>Finalize all presentation and publications</u> • <u>Website/Social Media</u>: Make last changes to the website, documenting the project

5. Communications Plan

The communication plan for our research will be multifaceted. It will include a project website, presentations and publications, social media to advertise our activities, and a collection within the TLE illustrating reuse. We will establish a website to post project reports, publications and presentations, and use as a social media platform. The website will also host our instrumentation. We will assess the reach of our website through the use of web analytics to monitor use and downloads of information.

We will use conferences to recruit subjects for our study (e.g., Association of Mathematics Teacher Education (AMTE), American Association of Colleges for Teacher Education (AACTE), American Education Research Association (AERA)) and to inform these communities about our work. Furthermore, we will target conferences for data curators, such as the International Digital Curation Conference, to disseminate our findings. We will also develop a communications plan that encompasses both of these communities. Along with this, we will announce our activities (conference attendance) and publications through social media, such as Twitter. Using Twitter, we will engage with members of the digital curation and the educational research/practice fields by both pushing out content as well as creating a network of those we believe would be interested in our work.

Finally, we will create a small collection within the TLE that includes records of practice produced through this project that illustrates some of the diverse ways educational researchers and practitioners reuse TLE resources.

DIGITAL CONTENT SUPPLEMENTARY INFORMATION FORM

Instructions: This form is required as part of grant applications to the Institute of Museum and Library Services that include activities that create certain types of digital content, such as online collections or databases, metadata, new software tools or electronic systems, or digital research datasets. Your responses to the questions on this form are used by IMLS staff and by expert peer reviewers to better understand technical aspects of your proposed work. Please consult the relevant program guidelines for further instructions on when this form should be included as part of your application.

If you need more space for your response, you may append additional pages as part of the single PDF that you upload with your grant proposal through Grants.gov.

Please indicate which of the following digital products you will create or collect during your project.

(Check all that apply):

	If your project will create or collect ...	Then you should complete ...
<input type="checkbox"/>	Born-digital, existing digital, or to-be-digitized content	Part I
<input type="checkbox"/>	New software tools or electronic systems such as databases	Part II
<input checked="" type="checkbox"/>	A digital research dataset	Part III

PART I. Projects Creating Digital Content

A. Selection Methodology

A.1 Describe how you will select non-digital materials for digitization.

A.2 Describe how you will select born-digital or existing digital content for your project collection.

B. Converting Non-Digital Materials to Digital Format

B.1 List the types and formats of materials to be digitized and the quantity of each type.

B.2 List the equipment and software that you will use to digitize each of these formats or the name of the digitization services provider who will perform the work.

B.3 List the digital file formats (e.g., TIFF, JPEG, MPEG) that you will produce during the digitization work and the anticipated quality standards for each file format (e.g., resolution, bit-depth, color/grayscale, pixel dimensions, sampling rate).

B.4 If different digital versions of content will be created during the digitization process (e.g., preservation master, access copy, thumbnail) list the type, format, and number of each version.

C. Repurposing Existing Digital Content or Creating New Digital Content

C.1 List the types and formats of born-digital or existing digital content that you will create or repurpose and the quantity of each.

C.2 If you will be creating new born-digital content or converting existing digital content to new formats, list the equipment and software that you will use to create each of these formats or the name of the services provider who will perform the work.

C.3 If you will be converting existing digital content to new formats, list the new digital file formats and relevant information on the anticipated quality standards (e.g., sampling rate, pixel dimensions).

C.4 If different versions of digital content will be created during the conversion or re-purposing process (e.g., preservation master, access copy, thumbnail), list the type, format, and number of each different version.

D. Digital Workflow and Asset Maintenance/Preservation

D.1 Describe your quality control plan.

D.2 Describe your plan for preserving and maintaining digital assets during and after the grant period (e.g., storage systems, data standards, technical documentation, migration planning, commitment of organizational funding for these purposes).

E. Metadata

E.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

E.2 Describe how you will use metadata to enhance the management, discovery, and use of your digital content.

E.3 Explain your strategy for preserving and maintaining metadata created and/or collected during your project, during and after the grant period.

E.4 Explain what metadata-sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content created or repurposed during your project (e.g., an Advanced Programming Interface or other support to allow batch queries and retrieval of metadata).

F. Copyright and Intellectual Property Rights

F.1 Explain the current copyright or intellectual property status of the content you intend to digitize, create, or repurpose. Describe the quantity or percentage of materials that are in the public domain and/or have restrictions that will require you to obtain permissions. If you have already obtained permission to use and provide public access to materials under copyright or other restrictions, provide the quantity of such materials, and the documentation you possess granting such permissions.

F.2 If you will need to obtain permissions during your project, describe the process you will use to request and obtain them.

F.3 Are there any materials you will be digitizing, creating, or repurposing that may raise privacy concerns? If so, what is your plan for addressing them?

F.4 If your project will include online users or others outside your organization contributing metadata, social media comments, or other content to your digital resources, describe your plan to obtain releases or permissions from these content contributors. What rights and permissions will you require such contributors to transfer to your organization?

G. Access And Use

G.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

G.2 We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in openly accessible journals, depositing works in openly accessible repositories, and using non-restrictive licenses such as the “CC Zero – No Rights Reserved” that dedicate digital content to the public domain. What ownership rights will your organization assert over the new digital content, and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users of the digital resources.

G.3 Provide URL(s) for any examples of previous digital collections or content your organization has created.

Part II. Projects Creating Software Tools and Electronic Systems

A. General Information

A.1 Describe the software tool or electronic system you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) the system or tool will serve.

A.2 List other existing digital tools that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your new digital content.

B.2 Describe how the intended software or system will extend or interoperate with other existing software applications or systems.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software or system you will create.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software or system.

B.5 Provide URL(s) for examples of any previous software tools or systems your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software or system development to develop and release at least a beta version of these products as open-source software. What ownership rights will your organization assert over the new software or system, and what conditions will you impose on the access and use of this product? Explain any terms of access and conditions of use, why these terms or conditions are justifiable, and how you will notify potential users of the software or system.

C.2 Describe how you will make the software or system available to the public and/or its intended users.

Part III. Projects Creating Digital Research Data (Data Management Planning)

We expect exemplary management and sharing of research data. The purpose of this part of the form is to help us understand your research practices and plans for management of data that will be generated through your project. Please address each question that applies to your proposed project.

1. Summarize the intended purpose of the research, the type of data to be collected or generated, the approximate dates when the data will be generated or collected, and the anticipated volume of data.

The purpose of the research is to better understand the needs of data reusers in a qualitative data archives (specifically those using the Teaching & Learning Exploratory). Diverse data (web analytics, surveys of data reusers of the TLE data (100), surveys of pre-service and practicing teachers exposed to instruction using TLE materials (150), interviews (50), videotaped observations of teacher educators using TLE materials (10) accompanied by the records of practice), will be collected in digital form from January 2015 - December 2016. Based on other research, this is approximately 10 gigabytes of data (primarily the audio from the interviews and video of the observations).

2. Does the proposed research activity generating the dataset(s) require approval by any internal or institutional review panel? If so, has the proposed research activity already been approved? If not, what is your plan for securing approval?

Our project will require approval by the University of Michigan Institutional Review Board (IRB). Immediately after we are approved for funding we will begin the IRB approval process. Since our start date is December 1, 2014 and we do not plan on collecting data until January 2015 and given that both investigators are experienced in conducting human subjects research and have submitted many IRB applications, we do not anticipate any problems in the IRB process.

3. Will you collect any confidential or private information about individuals (e.g., names, contact information, health status) or proprietary information about organizations? If so, detail the specific steps you will take to protect such information while you prepare the research data files for public release.

We will collect names as a by-product of the recruiting, consent, and scheduling processes; however, all data will be coded with a unique participant number so no names will appear on the actual data. Practicing and pre-service teachers who experience data from the TLE in professional development workshops or in the course of their education will be offered an incentive (\$5.00 gift certificate). We will need to collect names for this process but will also separate those forms these from the data.

We will protect privacy in several ways. No names will appear on the data. We will edit the interview transcripts to eliminate identifying information (e.g., employers) or other variables that, if aggregated, would make the participant more easily identifiable. However, Given the nature of our data, we will use our informed consent process to inform people about the difficulty of maintaining anonymity. Our informed consent form will allow the participants to indicate whether their interviews or videotaped observations can be transferred to a repository and opened to other researchers in the future. If they indicate yes, they will be informed that we will strive to protect their privacy but we

4. If you will collect additional documentation such as consent agreements or signed certifications along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

We will consent all participants who are interviewed or observed. These consent forms will be stored in hard copy in a locked cabinet or electronically on a secure server. The consent forms will be linked to the audio through the use of a unique participant number. This number will appear on our tracking spreadsheet, which will also be on a secure server. These will be kept in perpetuity as part of the data documentation. We will also transfer a copy of these to the repository with the data upon completion of our data analysis.

5. How will you manage intellectual property interests related to the dataset(s)? Who will claim ownership of the intellectual property rights related to the dataset(s)? How will those claims of ownership be communicated to others?

The data is co-owned by the investigators and the participants. As noted above, in the informed consent, participants will be informed of their rights as to how the data are used in the future and have the opportunity to consent to deposit in a repository or not. The investigators will deposit their data in a data repository on the UM campus. The investigators will retain ownership of the data but put a creative commons license on it for reuse.

6. Which technologies, instruments, or tools will you use to collect or generate the data? Provide details about hardware or software; electronic formats for data capture or storage; standards or local practices for data content and encoding; controlled vocabularies or other mechanisms for data normalization and consistency; and any other relevant technical requirements or dependencies for understanding, retrieving, displaying, or processing the dataset(s). If the data will be encrypted at any point in its active or inactive life, explain the reasons for choosing to encrypt the data and how the decryption key will be stored, protected, and made available if necessary.

Data Collection Method	Technology	File Format	Transfer to Repository	Encryption
Survey	Qualtrics	xlsx, csv	Yes	No
Survey	SPSS	spss	Yes	No
Interview - Capture	Audacity	mp3, wav	Yes	No
Interview - Transcription	Microsoft Word	docx	Yes	No
Interview - Data analysis	NVivo	nvp	No	No
Observation-Capture	Videotape	mpeg, avi	Yes	No
Observation-Transcription	Microsoft Word	docx	Yes	No
Observation-Records of Practice	Microsoft Office	docx, pdf, pptx, tiff	Yes	No

We will use basic Dublin Core metadata to identify all files in the header/properties section. In all cases we will strive to use the most open formats. The only case where there is proprietary data with no open format is the NVivo analysis of interview data, but we will not turn this over to a repository, we assume that other researchers would want to do their own analysis.

7. What metadata will you capture or create along with the dataset(s)? What standards or schema will you use to express the metadata? Where will the metadata be stored, and in what format(s)? How will you permanently associate and manage the metadata with the dataset(s) it describes?

All interviews and observations will be given a unique identifier and standardized naming conventions for research data. The identifier and naming conventions will also be used to tie in any associated documentation collected during the interviews or observations. For example, we anticipate that the observations of teaching use of the TLE materials will also generate lesson plans, worksheets, and perhaps even student work. These would be linked through metadata.

8. During the research project, where will the data and metadata be stored and on what type of media? Who will have access to the data and/or copies of the data during the project? How many backup copies will you maintain during the project, and how frequently will you refresh the backup copies? Who will be responsible for data backup? Where will you store the backup copies of the data and metadata during the project?

The data will be stored on secure servers at the University of Michigan, access to these servers is only available through a VPN (very private network), specifically granting permission for access to a server, and then only with appropriate authentication with a UM password. Access to the data will be restricted to project personnel (PI, co-PI, and students working on the project). Data backup is done on a weekly basis automatically by UMSI Computer personnel. Back-up servers are housed separately from the main servers.

9. Once the research project is completed, what is the long-term plan for archiving, managing, and making the metadata and dataset(s) available? What steps will you take to prepare the data for sharing (e.g., labeling missing data, standardizing measures statistical disclosure limitation methods)?

We will maintain the data on School of Information servers until we have completed most of our analysis and publication. At that time we will prepare the data for submission to a repository by reviewing the survey data and codebook for inconsistencies and adding any variables which we created in analysis to the initial codebook and making sure all the data are properly identified. We will again ensure that all data have the appropriate metadata in the file headers.

10. Identify where you will be depositing research dataset(s) and metadata into:

a) an institutional repository:

Name: URL:

b) a subject specific research community digital repository:

Name: URL:

c) or some other publicly accessible repository:

Name: URL:

Does this repository enforce any access restrictions? Yes (If yes, describe.) or No

The TLE is adept at working with sensitive data as it collected records of practice that involve underage children. Therefore it has developed protocols for dealing with these data and ensuring that researchers are informed of the issues and agree to abide by the policies.

ICPSR has access restrictions on confidential information and processes for researchers to apply to use these materials.

If so, how will they be mitigated to allow the public free access to these data? Detail the experience this repository has in managing research datasets and metadata with similar attributes? What preservation and backup procedures does this repository use?

In both cases the repositories have access restrictions to protect the privacy of the participants in the research. Both also have processes to ensure the research has knowledge of how to handle sensitive data.

ICPSR has the data seal of approval and has been managing social science data for 50 years. ICPSR has a strong preservation regime which is detailed at <http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/preservation/policies/index.html>.

The Teaching and Learning Exploratory is a new repository but is developing robust metadata and ontologies for

11. When and how frequently will you review this data management plan? How will the implementation be monitored?

We will refer to the data management plan as we collect, analyze, and maintain the data. Better data management will assist us in analysis and dissemination of products also.

We will do self-monitoring of the implementation of the plan and report on progress / issues in our reports to IMLS.

After data collection is over in year 3, we will begin negotiations with our intended data repositories to ensure we turn over well described data for ingest that contains all the elements of a good submission information package.