Library Hi Tech

Low-barrier-to-entry data tools: creating and sharing humanities data

Linda L. Rath

Article information:

To cite this document:
Linda L. Rath , (2016),"Low-barrier-to-entry data tools: creating and sharing humanities data", Library Hi Tech, Vol. 34 Iss 2 pp. 268 - 285
Permanent link to this document:
http://dx.doi.org/10.1108/LHT-07-2015-0073

Low-barrier-to-entry data tools: creating and sharing humanities data

Linda L. Rath
Newman Library, Baruch College, The City University of New York (CUNY), New York, New York, USA

Received 15 July 2015
Revised 16 October 2015
12 December 2015
Accepted 10 January 2016

Abstract

Purpose – The purpose of this paper is to determine whether TAMS Analyzer and Viewshare are viable free and open source software data sharing and creation tools for those with limited funding and technological skills.

Design/methodology/approach – The participant observer method was used to collect experiential evidence while applying the tools to a collection of text-, image-, and video-based digital cultural records.

Findings – TAMS Analyzer was found to be a low barrier to entry tool for those with coding and qualitative data analysis experience. Those with general experience will be able to create datasets with the support of manuals and tutorials, while those with limited experience may find it difficult to use. Viewshare was found to be a low barrier to entry tool for sharing data online, and accessible for all skill levels.

Research limitations/implications – TAMS Analyzer supports Mac and Linux platforms only, so a low-cost software recommendation was made for those in Windows environments.

Practical implications – Librarians can use these tools to address data access gaps while promoting library digital collections.

Social implications – With a greater understanding of data tools, librarians can be advisors, collaborators, agents for data culture, and relevant participants in digital humanities scholarship.

Originality/value – The research evaluates both the capabilities of the tools and the barriers to using or accessing them, which are often neglected. The paper addresses a need in the literature for greater scrutiny of tools that are a critical component of the data ecology, and will further assist librarians when connecting scholars to tools of inquiry in an environment with limited funding and technical support.

Keywords Digital libraries, Academic libraries, Data, Qualitative research, Digital humanities, Data services

Paper type Technical paper

Introduction

Data and analysis tools are critical for scholarly inquiry. The growth of digital data and the proliferation of tools enable new approaches of exploration and the emergence of new fields of study. Collectively, this data-intensive approach is known as the fourth paradigm of scientific discovery and scholarship, following the empirical, theoretical, and computational models (Lynch, 2009). Within humanities scholarship, the field of digital humanities has emerged. Digital humanists engage in text mining and cultural analytics to examine digitized and encoded texts, images, audio, videos, and other objects available for humanistic inquiry. Computational processing tools are used to reveal details of source materials, uncover patterns and anomalies, quantify attributes, and create visualizations to assist with knowledge creation. Unfortunately, access gaps to resources and skills exist and they impede scholarly pursuits. The National Science Board's (2005) long-lived digital data collections report found, "Not all researchers have equal access to the resources and expertise necessary to create and operate a digital

Library Hi Tech
Vol. 34 No. 2, 2016
pp. 268-285

0737-8831
DOI 10.1108/LHT-07-2015-0073

data collection. The need is especially apparent at the level of an individual investigator developing a research collection" (p. 23). While data and tools are more prevalent today, access gaps continue to impede both researchers and librarians due to lack of funding, skills, training, technical support, and common or shared data infrastructure (Borgman, 2015; Posner, 2013). The barriers encountered when addressing the access gaps differ by discipline. Librarians and humanists must be strategic and resourceful to address the economic, technological, and cultural barriers hampering humanities research. One solution is the use of low barrier to entry tools, affordable tools requiring minimal training or cost, to create and share data. For librarians advising or collaborating with self-funded or minimally funded researchers in the humanities, free and open source software (FOSS) data tools are attractive. The online digital research tools (DiRT) directory is useful for identifying tools, but the many choices can be overwhelming or confusing. Investing time with trial and error and disappointing results can be avoided if library literature included more evaluations of data tools and their uses in various, domain-specific contexts. Tools are a vital component in the data ecology and need to be examined and scrutinized (Zuiderwijk *et al.*, 2014). Beyond listing software features, evaluations can determine whether a tool's barrier to entry is worthwhile, and an achievable and practical solution for librarians and scholars with limited funding or technical skills.

Using a case study approach, this paper explores TAMS Analyzer, a qualitative data analysis tool, and Viewshare, an online digital collections display platform, as complementary FOSS tools to create and share humanities data. A review of the literature discussing the academic data landscape, humanities data, and library data services provides context and six criteria for evaluating the tools. The process of turning non-structured, qualitative data from text, image, and video files into structured data and shared datasets will determine the degree to which TAMS Analyzer and Viewshare are low barrier to entry data exploration and sharing tools supporting humanities scholarship.

Literature

The academic data landscape

While the volume and variety of digital data created from the scholarly process is growing, the contexts of data creation impact how frequently or easily it can be accessed, shared, discovered, reused, and preserved. Anderson's (2004) "long tail," a concept describing the internet market's demand curve to illustrate the availability and distribution of a few, high-demand and popular items versus the cumulative variety and low-demand of obscure items, is regularly applied in library literature to describe variables of scholarly data distribution and the variety of data available in terms of big data sciences, such as astronomy and physics, and small sciences data, such as arts and humanities (Borgman, 2012, 2015; Fitzpatrick, 2012; Heidorn, 2008; Wallis *et al.*, 2013). The data of big sciences are well funded, consistently structured, shared among large teams for long-term projects, created for longevity, reusability, and transferability, and are easier to find since datasets are placed in established domain- specific or university repositories (Borgman, 2015; Wallis *et al.*, 2013). Often, knowledge creation in big sciences is funded by grants requiring data to be publicly available. This results in a large volume of demand for these datasets, even though they represent a fraction of the total data output. Small sciences data tends to have nominal funding or is self-funded, and is structured for local projects with small teams or for individuals (Wallis *et al.*, 2013). Discovering small sciences data can be difficult since there are few, well established, domain-specific data repositories (Reilly, 2012). Additionally, data may not be in a format acceptable to deposit in university repositories, and may only be posted on personal websites. Since the volume and sharing of data produced by each project is small and infrequent, it is mostly neglected. This results in data that is difficult to find, and less frequently reused (Heidorn, 2008). Considering the cumulative volume of long tail data represents a majority of scientists (Heidorn, 2008), more attention should be given to small sciences data. Rectifying long tail data issues has the potential to increase discoverability, use, repurposing, and professional acknowledgment through the citing of data. Long tail researchers need better tools to create, manage, access, and share data (Wallis *et al.*, 2013), and scholarly literature should increase publications investigating data exploration tools (Tiwari *et al.*, 2014) to assist scholars with identifying appropriate tools.

Humanities data

The economic, technological, and cultural barriers for long tail humanists are plenty. Funding allotted for humanities research tends to be minimal, and scholars are nominally supported by their institutions or are self-funded (Fitzpatrick, 2012). Humanities-focussed data repositories are not well established (MacMillan, 2014), and websites of individual humanists sharing their data may be difficult to find, not well managed, or out of date (Wallis *et al.*, 2013). Librarians have inefficient funding, training, time, support, and professional incentives to take on additional roles as advisors, collaborators, and contributors for digital humanities research, or to fully integrate digital humanities activities into library workflows (Posner, 2013). This creates a landscape where costly datasets and software tools are out of reach for individual humanists or too expensive for underfunded libraries; free and appropriately structured data may be available, but is difficult to locate or is undiscoverable; technological skills are insufficient to reformat accessible data into the scholar's needed format or structure; and libraries have insufficient resources or skills to be proactive agents of data culture (Thessen and Patterson, 2011). However, if given the capabilities to successfully create data, humanists are most willing of all scholars to share their data with the public (Akers and Doty, 2013), contributing to the overall growth of data. Additionally, libraries strive to increase the visibility and use of their collections and services, connect researchers with resources and tools, remain relevant, and demonstrate value.

The expansiveness and complexities of humanities data are discussed and captured in the literature. Borgman (2009) conveys a grand view by stating, "Any record of human experience can be a data source to a humanities scholar" (p. 33), and further stresses how humanists rely heavily on data generated from records, versus data generated from observations, models, or experiments. Examples of cultural records can include newspapers, correspondence, transcripts or recordings of speeches, photographs, interviews, statues, and websites. Others insert human influence and process into definitions by describing humanities data as "[…] selectively constructed, machine-actionable abstraction representing some aspects of a given object of humanistic inquiry" (Schöch, 2013, para 8) and "[…] organized difference presented in a form amenable to computation put into the service of Humanistic inquiry" (Padilla and Higgins, 2014, p. 325). Owens (2011a) emphasizes how data are an authored work and multifaced object. Data are constructed artifacts, interpretable and created from selected items and encoding choices for an audience to process and manipulate. Data are an information source that holds evidentiary value. The decisions made during manufacturing, processing, and interpreting data can shape the evidence.

When humanists reuse data, they should consider the author and context of the data, along with their own authoring decisions when reusing it.

When it comes time to create data for inquiry, humanists recognize a need for ingenuity. There is a sense that digital humanists are willing to work with the resources at hand employing a do-it-yourself (DIY) approach (Owens, 2011b). When possible, using tools with low barrier to entry design would be welcomed and ideal (Bailey and Owens, 2012). When addressing issues related to maximizing data visibility, humanists are encouraged to turn their long tail data into discoverable, reusable, and cited data. Establishing partnerships and collaborations with libraries is recommended (Hagood, 2012) to share expertise, create domain resources and repositories, discuss relevant technologies, use common standards and tools, and to address common challenges (Borgman, 2009). To further push for a critical mass of usable data, Laney (as cited Schöch, 2013) promotes moving from small, clean, labor intensive datasets to creating clean, "big data" in the humanities with "volume, velocity, and variety" (para 19) through automatic annotation and the shared labor of crowdsourcing.

Academic libraries and humanities data

Few academic libraries offer data management services (MacMillan, 2014) and most are inadequately equipped (Gelfand and Tsang, 2015), yet they do foresee providing some data services in the future and are open to opportunities. When data services are offered, libraries must be selective or focussed. The range of data services is dependent on budgets, staffing, and skills of librarians. Current services provided by libraries include creating web guides assisting with data finding, advising with data management, providing metadata expertise, preparing data for repositories (Tenopir *et al.*, 2012), customizing small datasets for research, and providing access to primary sources and making them interactive (Kamada, 2010).

Recommendations for improving data services emphasize the activities of liaisons and their established relationships with faculty (Kim, 2013; MacMillan, 2014; Nicholson and Bennett, 2011). Liaisons with subject expertise can identify potential collections for digitization and encoding (Tiwari *et al.*, 2014), provide domain relevant data services, and should receive professional development to provide next level data services (Kim, 2013). Another recommendation includes the establishment of institutional subject or domain repositories (Akers and Doty, 2013; Tenopir *et al.*, 2012). Additional recommendations target library collections, and visibility strategies. Promoting library digital collections as humanities data (Padilla and Higgins, 2014) can encourage use of collections in exploratory ways, and differentiate a library (Dempsey *et al.*, 2014). Additionally, libraries should advocate a culture of open data (Nicholson and Bennett, 2011), optimize library websites and collections for search engines (Arlitsch, 2014), and publish outside of traditional venues. Libraries establishing open and public digital data collections that are findable via commonly used search engines increase their visibility and use. Using this "giving it away" approach (Fitzpatrick, 2012) can both confirm and showcase the value of the library with both local and external users.

Selection and background of TAMS Analyzer and Viewshare

TAMS Analyzer and Viewshare were selected as the tools for this case study. TAMS Analyzer was selected for coding the unstructured text-, image-, and video-based files to create structured data and datasets. Viewshare was chosen as the platform to display and share the datasets online. This process replicates the data creation and sharing activities a humanist or librarian would perform when using digital cultural records as source documents. Both tools were selected after searching the online DiRT directory and reviewing literature discussing computer assisted qualitative data analysis software (CAQDAS) and platforms for sharing digital collections. Both tools met the requirements of being a free or low-cost tool; supported text, image and multimedia files; and supported more than one platform (Windows, Mac, Linux, and web-based). TAMS Analyzer is a free download supporting Mac and Linux platforms, and Viewshare is a free, web-based hosted service accessible after creating an account. While other tools such as Coding Analysis Toolkit, CATMA, QDA Miner Lite, and RQDA are free, they do not support image and multimedia files. Dedoose, a low-cost, web-based tool with subscription pricing, was not selected since it does not support image files or the capabilities to code sections of images or visual content within text files or PDFs.

TAMS Analyzer is CAQDAS. The "Text Analysis Markup System" was developed when the software market did not serve Mac users with a CAQDAS program. TAMS Analyzer was designed by a scholar for ethnographic research to identify themes in field notes and interviews (Weinstein, 2006), and has been used in anthropology, education, marketing, and other social sciences research analyzing varying documents of interest to respective investigators. The program allows investigators to extract, analyze, save, and export coded information, and create memos during the process. TAMS Analyzer was selected since it supports text, image, and multimedia files, and allows one to assign codes to passages of texts or sections of images, audio, video, and PDFs. Multiple investigators can code a project with "signed tags," identifying individual coders with the passages or sections they coded. TAMS Analyzer can generate reports with code counts, tables of code comparisons, and create concept maps graphically representing relationships of the codes used in the project. Investigators can search across all documents in a project, sort and filter results, and save results in tab delimited files. Saved results can be exported to Microsoft Excel as a simple spreadsheet.

Viewshare is a web application used to create interactive interfaces to digital collections, and is administrated by the National Digital Information Infrastructure and Preservation Program at the Library of Congress. Viewshare was designed for librarians as a low barrier to entry tool for managing and providing access to cultural heritage materials, and is open to anyone wanting to use the platform to share collections, or data, online (Algee *et al.*, 2012), once approved for an account. Viewshare allows users to import collections from Microsoft Excel spreadsheets, XML MODS, JSON, OAI records, or CONTENTdm to their Viewshare account. Once uploaded, one can generate "views," or interfaces, of the collections. Views can be in the form of maps, timelines, scatter plots, bar charts, tables, lists, pie charts, and galleries (thumbnail images). Viewshare supports multiple formats and URLs, and can augment data. For example, if uploading a Microsoft Excel spreadsheet containing geographic names, Viewshare can augment the geographic place names into latitude-longitude coordinates to create interactive maps. The interactive views can be searched, filtered, and sorted when exploring the collections. The views, or datasets, can be shared, embedded into websites, and saved to the clipboard in multiple formats (Exhibit JSON, tab separated values, semantic wikitext, RDF/XML, BibTex, and generated HTML).

When paired together, TAMS Analyzer and Viewshare can be complementary partners for creating and sharing datasets. Librarians and scholars can turn unstructured text documents, images, and media files into structured data through coding in TAMS Analyzer. The codes, along with the coded text or data related to coded sections of media files, can be exported to Microsoft Excel as a simple spreadsheet. The spreadsheet is then uploaded into Viewshare where an interactive interface to the data can be created. Once published, the dataset can be shared or embedded to make it easily accessible and findable online. TAMS Analyzer and Viewshare were designed for slightly different purposes, but using both in this DIY manner may be a viable way for librarians and humanists to use resources on hand.

Research method and design

The participant observer case study method was chosen for collecting experiential evidence of the process. A log was created to collect observations and notes as qualitative data. Text-, image-, and video-based cultural records were chosen since they are heavily explored in humanist inquiry.

The cultural records chosen for the case study were comprised of digital text files representing annual schedules for genre film festivals, marketing images from festival websites, and online videos containing interviews with festival programmers and attendees. The content of the text files consisted of film festival information (location, dates), a schedule of film screenings and special events, attributes of individual films, and special events information. The image files represented logos, posters, and website headers from festival websites, and the video files contained discussions related to festival programming and attendance. The goal was to create online, interactive displays allowing users to filter or search datasets by festival name, geographic location of the festival, and year of the festival; film title, director(s), year of film release, production country, and short or feature length format; content of festival images; and content of festival videos. Patterns in genre film festival programming could be identified when filtering or searching the data, or a reference source of programmed films could be generated. Additionally, patterns in visual film festival marketing images, programming strategies, and attendee preferences could be identified through content analysis. Another requirement was the ability for others to download and share datasets, so data would be available for their own exploration and scholarship needs.

The research questions:

RQ1. To what degree is TAMS Analyzer a low barrier to entry data tool for the creation of structured data for humanities scholarship?

RQ2. To what degree is Viewshare a low barrier to entry data tool for the creation of an online dataset for humanities scholarship?

The evaluation criteria were based on economic, technical, and cultural barriers identified in the literature review.

Evaluation criteria:

(1) cost comparison;

(2) support/help documentation;

(3) skills required;

(4) sharability and findability;

(5) time required and labor;

(6) professional rewards and incentive; and
(7) format support.

The steps in the process:

(1) identify text-, image-, and video-based, unstructured cultural records, and the excerpts or sections within the records to be coded;
(2) create structured data by coding selected excerpts of texts, images, and videos from the cultural records using a customized coding scheme in TAMS Analyzer; and
(3) share dataset by publishing online using Viewshare.

Identify text-, image-, and video-based, unstructured cultural records, and the excerpts or sections within the records that will be coded

In total, 70 text files, 30 image files, and 10 video files were identified for coding. The files represent annual schedules for 12 genre film festivals running in the USA and Canada between 2001 and 2013, marketing images from film festival websites, and videos posted on social media discussing film festivals. The inaugural years and length of festivals varied, so the text files represented a cumulative 70 years of schedules. The text selected for coding in the documents identified film information (title, director, country of release, year of release, length of film, feature or short format), and special events (awards, contest, discussion/panel, games, guests, lounge, party, and performances). The sections of images selected for coding represented visual elements and content depicted in logos, posters, and website headers. The content was coded to identify text within the images, colors, geographic references, and commonly depicted visuals such as body parts (eyes, ears, hearts, mouths), weapons, and monsters or creatures associated with horror, science fiction, and fantasy film festivals. The clips of video selected for coding represented quotes highlighting programming strategies, and festival experiences of attendees.

Create structured data by coding selected excerpts of texts, images, and videos from the cultural records using a customized coding scheme created in TAMS Analyzer

The text, image, and video files were imported into TAMS Analyzer as rich text format (.rtf), JPEG, or MP4 files. An initial file (init.rtf) was created to establish context codes. These codes are assigned to establish the context for each file and included the film festival name, festival city, festival state/province, festival year, and the original source of the document (website URL, paper, etc.). Image files included the type of image (logo, poster, and website header).

Each file was assigned their respective context codes, and then the content within the files were coded using a hierarchal coding scheme. To assign codes to the text files, the source text was selected and then a code was chosen from a custom list. Film titles were assigned the code of (film), the director's name was assigned the code (film > director), the year of release was assigned (film > year), the country of production was assigned (film > country), and the length of the film was assigned (film > length). Short films were given the same structure, but used the "short" code instead of "film." Special events were identified using their respective codes. Codes were formatted into different colors to visually differentiate them from the source text of the document and between other codes. The feature film codes were formatted green and the short film codes were formatted orange. Special events were assigned various colors. Color coding helped to identify coding errors when visually scanning the text (Figure 1).

While coding files, code reports were generated to provide code counts to help identify errors. In some instances, this identified passages of text that were coded multiple times, with the wrong code, or were overlooked. Once coding was complete, the documents were searched by codes, sorted, and filtered. This helped to identify and correct text errors, such as typographical errors and extraneous empty spaces. An iterative process of searching codes and exporting results to Microsoft Excel created four spreadsheets. The four spreadsheets included a film festival directory, a list of all feature length films programmed at the 12 film festivals, a list of all short length films programmed at the 12 film festivals, and all films (feature and short length) programmed at a specific festival, Fantastic Fest in Austin, TX. Using data sorting options in Microsoft Excel provided another opportunity to review data and correct any errors. The spreadsheets were saved using the .xls extension, since Viewshare does not support .xlsx files.

A similar process was performed to code the image and video files, although the interface of the coding windows and the method to select areas of images or clips of videos when coding differed (Figure 2).

For images, rectangular shapes were drawn over selected areas by dragging the cursor, then appropriate codes were chosen from a custom list and applied to the rectangular area. Coders can type additional data and comments into fields to be associated with the code and selected rectangular area. Textual elements within the images were assigned the code "Text" and a description of the text was typed into the "data" and "comments" field. Other visual elements were coded by content and

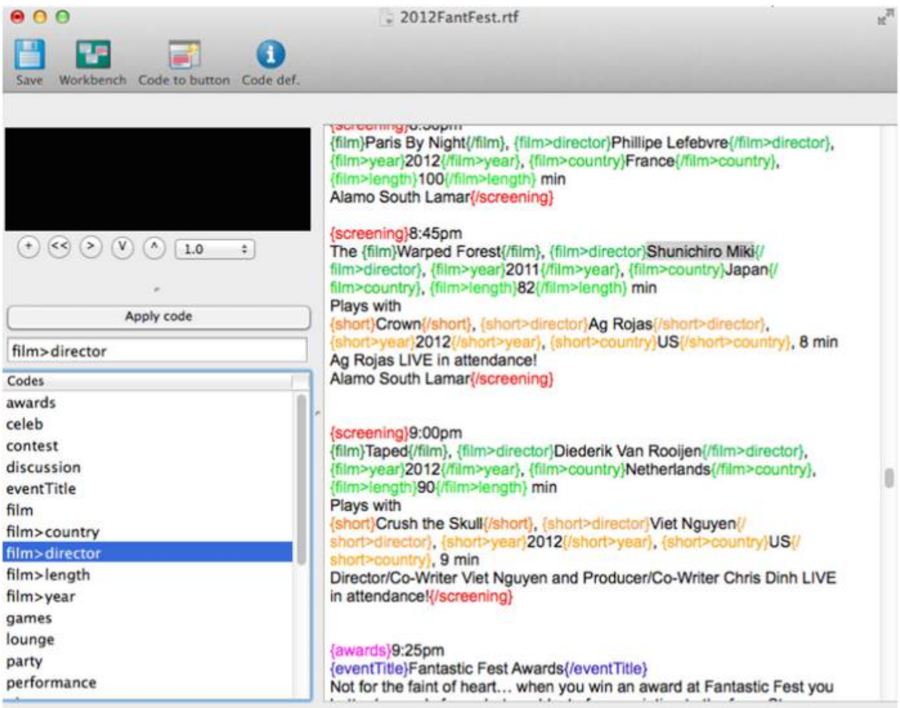

Figure 1.
TAMS Analyzer window for coding a text file

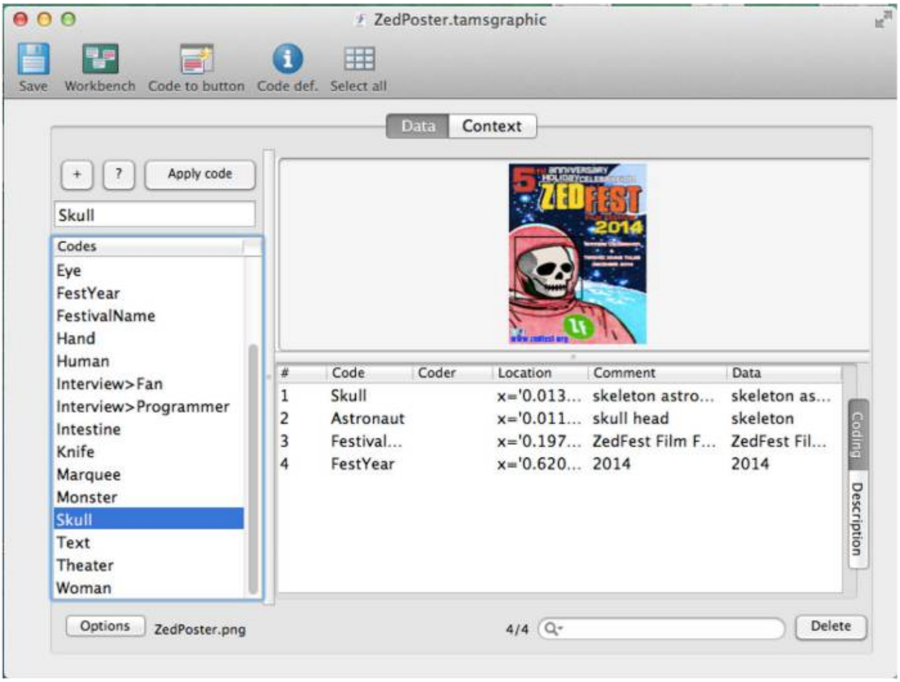

Figure 2.
TAMS Analyzer window for coding an image file

assigned codes such as "Monster," "Human," "Eye," "Hand," and "Camera," and were similarly described in the "data" and "comments" fields. Rectangular areas for codes can overlap, allowing an area to be assigned multiple codes or have overlapping codes. Searching and filtering by codes across files and generating code reports assisted with identifying errors or areas in need of coding.

The same process was used for the video files, although the coding window differs and allows for video playback, moving forward and backward within a video file, and a button to create insertion points when selecting clips within the videos to code. After applying codes to selected clips, data and comments were added into fields to be associated with the clip and code. Clips can overlap with one another, allowing a section of video to be coded multiple times and be associated with more then one code. Codes assigned to video clips identified a person as a "programmer," "fan," "interviewer," and "filmmaker" along with topics such as "film," "panel," "party," and "competition." Quotes were transcribed into the data field of their respective codes and clips, along with additional descriptive text. Searching and filtering by codes across files and generating code reports assisted with identifying errors or clips in need of coding.

An iterative process of searching codes and exporting results to Microsoft Excel created one spreadsheet related to genre festival images and one for genre festival video interviews. Timecode data associated with the coded video clips (beginning and ending of the clip) were included in the spreadsheet (Figure 3).

Share dataset by publishing online using Viewshare

The Microsoft Excel spreadsheets were uploaded into the Viewshare account. Six views were built based on the six spreadsheets. The Edit window was used to

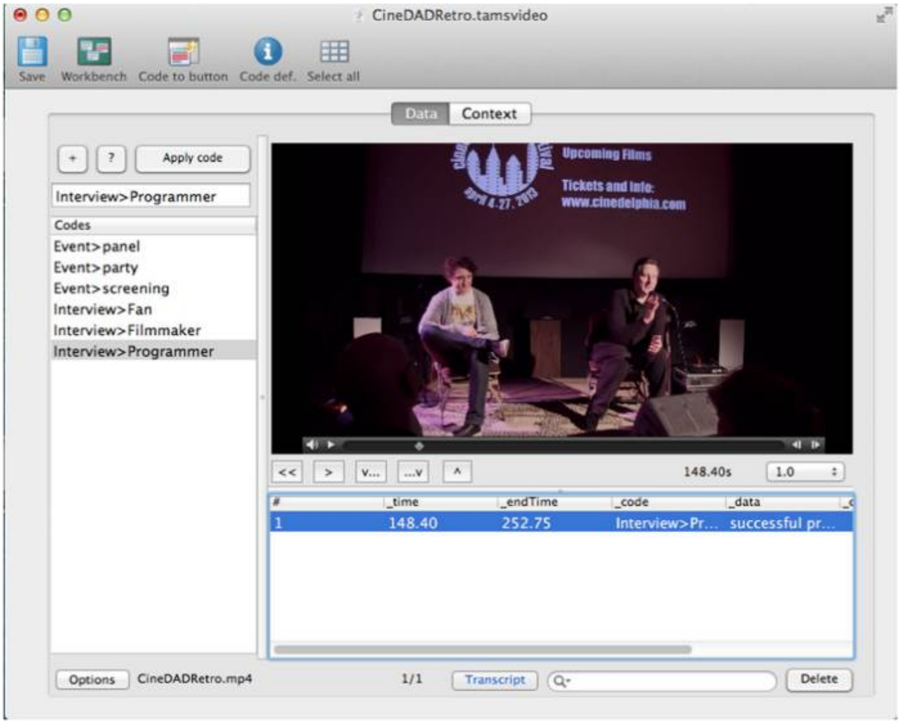

Figure 3.
TAMS Analyzer window for coding a video file

describe the data, assigning labels and data types. Fields related to geographic cites, dates, and URLs were augmented so they could be used in interactive maps, timelines, and link to the original source of the online image or video (Figure 4).

Using menus and drag and drop options, views were created to represent the data in various, interactive displays. The maps, tables, lists, scatter plot, pie chart, and bar chart displays were chosen. Widgets were added to allow searching and filtering of data through a search box, tag cloud, sliders, selecting a range, or selecting an option from a list. Text was added to the views providing a description and scope of the data, and how to cite the data. This was an iterative process, and after each change and update, the view was saved (Figure 5).

Using a preview screen helped with deciding which type of interactive views and filtering options were most appropriate for the data. Once ready, the datasets were made public and the URLs were shared. The options to embed on another website were tested, and worked successfully (Figure 6).

When typographical errors were noticed, they were corrected in the source spreadsheet and the view was refreshed. If it is necessary to make major revisions, one may want or need to upload a new spreadsheet as a source document for the view. While this is possible, the view will need to be recreated.

Views for the image and videos included URLs to the source of the online images and videos. For the view created with the image file codes, the source URL was augmented during the edit process when importing the data into Viewshare to include a thumbnail image in the data display (Figure 7).

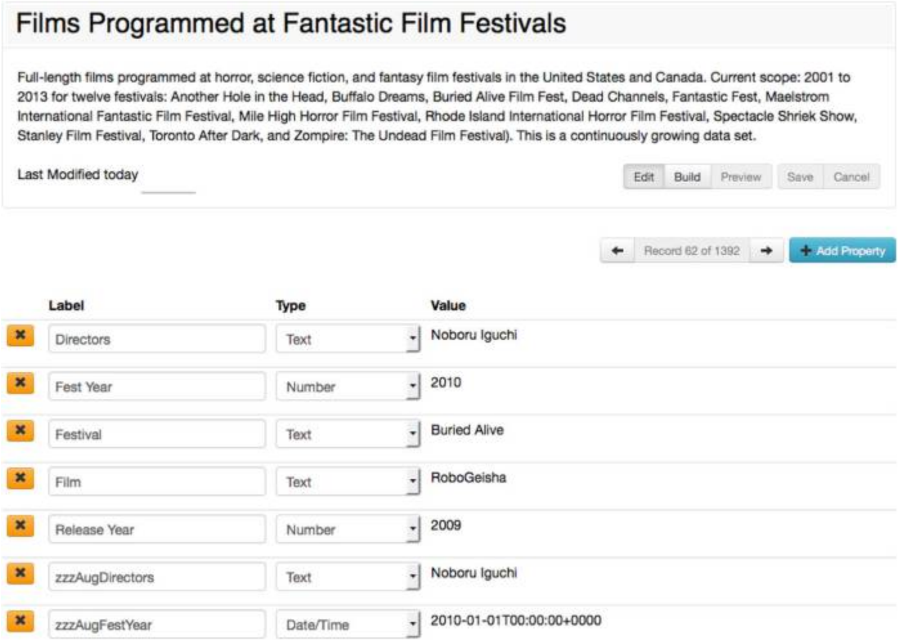

Figure 4.
Viewshare window to describe imported data

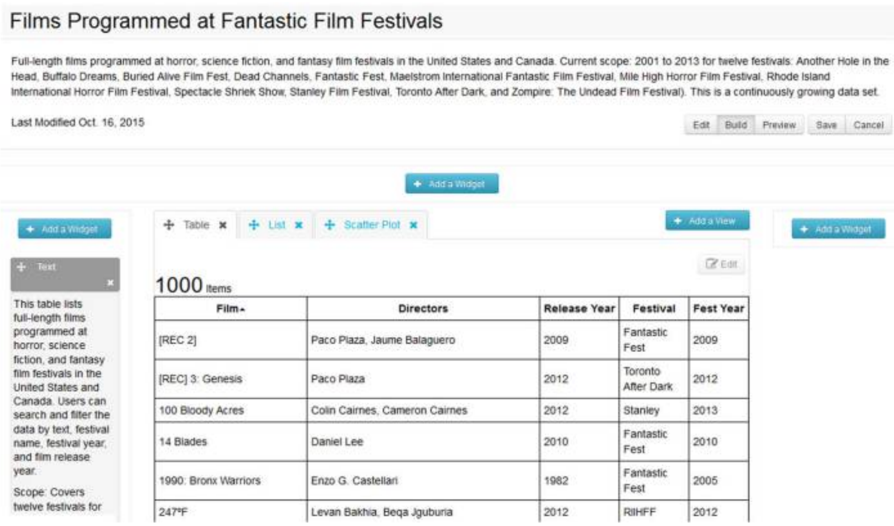

Figure 5.
Viewshare build window to create views of data

Findings for TAMS Analyzer

To what degree is TAMS Analyzer a low barrier to entry data tool for the creation of structured data for humanities scholarship?

Cost comparison. The DiRT registry and literature were consulted to identify qualitative data analysis tools with similar coding functions as TAMS Analyzer. Pricing for a single

Figure 6.
View on Viewshare depicting data being saved as tab separated values

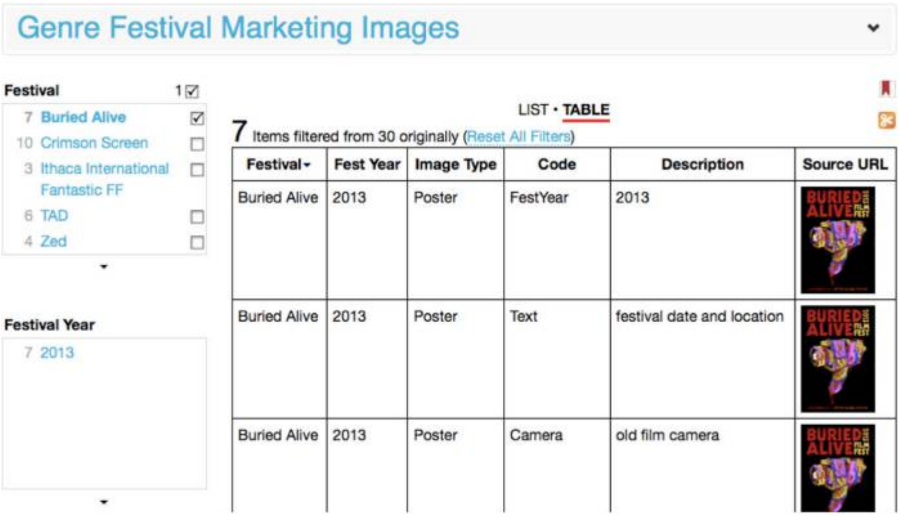

Figure 7.
View on Viewshare depicting a dataset created from image files

user, educational license was identified for 11 additional software programs (see Table I). The price for ATLAS.ti is $670 for either a Windows or Mac license, HyperRESEARCH is $499, MAXQDA is $495, Transana Professional is $350, and NVivo is $690 for Windows and $570 for Mac. Considering TAMS Analyzer is free, there is a substantial difference for a self-funded, individual. These tools do have additional features and capabilities that may make the price worthwhile, depending on one's needs. The web application Dedoose is another possibility to consider. The price for an individual is $12.95 per month, or $155 per year. If only short-term access is needed, this is a cost friendly option. Dedoose

Table I.
Computer assisted qualitative data analysis software

CAQDAS	Price	Platform support	File support
ATLAS.ti	$670	Windows, Mac	Text, images, PDFs, audio, video
Coding Analysis Toolkit (CAT)	Free	Web-based	Text
CATMA	Free	Web-based	Text
Dedoose	$12.95/month $155/year	Web-based	Text, PDFs, audio, video
HyperRESEARCH	$499	Windows, Mac	Text, images, audio, video
MAXQDA	$495	Windows, Mac	Text, images, PDFs, audio, video (tables)
NVivo 11 Pro	$690	Windows	Text, PDFs, images, audio, video, spreadsheet, web data
NVivo for Mac	$570	Mac	Text, PDFs, images, Audio, video, spreadsheet, web data
QDA Miner Lite	Free	Windows	Text, PDFs
RQDA	Free	Windows, Mac, Linux	Text
TAMS Analyzer	Free	Mac, Linux	Text, images, PDFs, audio, video
Transana Professional	$350	Windows, Mac, Linux (partially)	Text, image, audio, video

may be a budget-friendly option for Windows users not supported by TAMS Analyzer and not needing to analyze image files.

Support/help documentation. User guides and screen shots are available on the TAMS Analyzer website and in the download files. Video tutorials are posted on YouTube and Vimeo, and updates are posted on the TAMS Analyzer Facebook page. The developer of TAMS Analyzer, Matt Weinstein, created the documents and videos. These are useful guides, and thorough. The manual was helpful with setting up the project with initial files, importing documents, explaining project windows, how project files work together, how to create a code list, and how to apply codes. When discussing coding schemes and advanced coding features and searching, the presentation of the material was not as clear. This section may be difficult for those with little qualitative analysis and coding experience or new to CAQDAS. Several readings and practice projects may be needed. Using TAMS Analyzer for MAC OSX, a the manual created by Tabitha Hart while at the University of Washington's Center for Social Science Computation and Research, was helpful in conjunction with the Weinstein manual, and presented coding aspects in a clear presentation for a novice. It is recommended to read the manuals, thoroughly, and watch tutorials before beginning a project.

Skills required. Basic file management skills are needed, along with comfort using multiple windows for a project. Coding the documents was a simple task involving the selection of text, area of an image, or video clip, and then selecting a code from a list. If one has experience with creating coding schemes, TAMS Analyzer is easy to use. Those with little experience will need additional time or may consider learning more about coding schemes, particularly if complicated schemes are needed for a project. Simple coding schemes will be easier for novices, but must be appropriate for the dataset requirements.

Sharability and findability. TAMS Analyzer does not have built-in, seamless functions supporting the sharing of reports online. One could copy, save, and export files, and then upload into websites or e-mail, if needed.

Time required and labor. The time needed to code documents depends on the number of files to be coded, the length of the documents, the number of codes being used or the amount of the files being coded, and the degree of familiarity the coder has with the files. The amount of time generating reports, analyzing codes, and exporting data to create Microsoft Excel spreadsheets depends on project parameters, as well. Large projects with lengthy files and heavy coding can be labor intensive. Coding 70 text files required more time than was estimated for the project. The text files representing schedules for large film festivals were lengthy and heavily coded, demanding nearly double the estimated time. Coding video files required multiple viewings of video clips during the coding and transcribing process, necessitating slightly more time than was estimated. Since TAMS Analyzer supports multiple coders for projects, labor can be shared.

Professional rewards and incentive. The reward from this project includes learning new skills and tools, gaining greater familiarity with coding a variety of file formats, and developing management and workflow strategies for coding projects. Knowledge gained was transferred to faculty members and students when discussing research requiring the coding of videos and text documents for anthropology and communication courses.

Format support. TAMS Analyzer was developed to code text-based files, and uses RTF files. Microsoft Word documents must be converted to RTF files. Support for additional file formats include images, PDFs, video, and audio files can be coded. Image files must be in jpeg format. Of note, codes for video and audio files are added to a log, or transcript, and linked to the video/audio file through time codes. The audio and video files are not directly coded, as they are with other coding programs. Multimedia files need to be in a format supported by Quicktime (AIFF, WAV, MP3, MP4, and etc.).

Evaluation of TAMS Analyzer

TAMS Analyzer is a low barrier to entry data tool for those with general familiarity to extensive experience with qualitative data analysis and coding, basic file management skills, and when used for a small- to medium-sized project. TAMS Analyzer was developed to code and analyze text-based documents and multimedia files, and provides the necessary functions to accomplish this task. Liaisons and scholars with subject expertise are good candidates for coding due to their familiarity with data needs of the domain. The scalability of the software to support large-scale projects with many files should be tested. TAMS Analyzer can be used as a stand-alone product if a researcher only needs to code and analyze materials, and not publish datasets online. Those with limited-to-general qualitative analysis and coding experience should prepare before attempting a project. Watching tutorials and reviewing research methods guides is recommended, and would make TAMS Analyzer more of a medium barrier to entry data tool. Those new to qualitative analysis and coding may find TAMS Analyzer difficult to use, unless the project is small and coding scheme is simple. Unfortunately, TAMS Analyzer does not support those in Windows-only environments.

Findings for Viewshare

To what degree is Viewshare a low barrier to entry data tool for the creation of an online dataset for humanities scholarship?

Cost comparison. The DiRT registry and literature were consulted to identify online platforms for visualizing collections with similar capabilities as Viewshare. There are a variety of cultural heritage collection platforms, and there are numerous individual applications dedicated to a specific visualization such as timelines, word clouds, and maps. However, none provide the capabilities to present data in multiple views while also providing the capability to share or download the dataset as a whole or as a subset after being filtered. A cost comparison was not made. Viewshare is a valuable multi-purpose and accessible tool.

Support/help documentation. The support documentation for Viewshare is excellent. Online guides and tutorials are helpful and presented well. Directions are clear and easy to follow. Help can be found in the online forums and knowledge base, as well. When needed, contacting support via an online submission form resulted in quick responses, usually within 24 hours.

Skills required. Basic file management skills are needed to use Viewshare, and comfort with uploading files, and using menus. The initial set-up of describing the collection data and augmenting data are essential for creating views, but is clear and easy to accomplish, especially with the step-by-step instructions available.

Sharability and findability. The Viewshare datasets were published online and made public, successfully. All datasets are findable when using Google, Bing, and DuckDuckGo search engines. The capabilities to embed datasets into webpages worked when tested. The clipboard function was tested to save views of data as tab delimited values and generated HTML, successfully. The bookmarking function to save the current view of datasets were successful, as well. All sharing tasks are easy to perform by selecting button icons.

Time required and labor. Labor included loading a spreadsheet, or collection of data, into the Viewshare account, describing the data for the views, and then creating views of the data using menus and drag and drop selections. A view of data can be accomplished within minutes, if one is prepared and familiar with the interface. When experimenting with different views and options, the process will take longer. For evaluation, six views were created. The first view took the longest due to unfamiliarity and experimentation. The fifth view was a quick process.

Professional rewards and incentive. The reward was learning a web application that can be used for data exploration and sharing, and knowledge can be transferred to scholars. The knowledge gained from this process can be used to share cultural heritage digital collections, as well. Additionally, datasets created for this paper were added to an online genre film festival resource, and could result in increased visibility and use of the datasets on Viewshare and heighten awareness of the film festival resource.

Format support. Viewshare was designed to host cultural heritage digital collections, so it supports a variety of collection types. Besides supporting simple Microsoft Excel spreadsheets, the XML MODS, JSON, OAI records, ad CONTENTdm data types can be uploaded and made into views. Fields types for the records can be text, URL, image, data/time, location, number, audio URL, and video URL. Data in a Viewshare view can be saved in Exhibit JSON, tab separated values, semantic wikitext, RDF/XML, BibTex, and generated HTML formats, supporting standard scholarly and repository formats.

Evaluation of Viewshare

Viewshare is a low barrier to entry data tool to make humanities data findable, accessible, and sharable online. Anyone with basic file management and computer skills can create views to publish data, after becoming familiar with the interface, options, data setup, and menu selections. Those new to creating datasets or views to cultural heritage digital collections will find the user guides and tutorials helpful, and most likely all that is needed to successfully publish. The datasets created for this paper represented discrete text and visual content within text-, image-, and video-based cultural records. If one wants to promote library digital collections as humanities data, where the records themselves are the data, this can be accomplished by creating a spreadsheet including URLs to the items in the library collection. The scalability of Viewshare should be tested depending on the data types. It works well with small to medium collections. Large datasets may create long webpages, crowded views, and become cumbersome to use or visually unappealing. Color choices are limited, and bar or pie charts with many items may repeat colors. This makes differentiating data difficult. If one only wants to create a word cloud or a timeline as access points to data, investigating dedicated tools should be considered. However, Viewshare excels with providing options on how data can be explored in different ways. Providing different views to the same data may lead to new interpretations or uses of data.

Discussion and conclusion

Pairing TAMS Analyzer and Viewshare accomplished the goal of creating, publishing, and sharing humanities data. The DIY approach utilizing the resources at hand, in ways they were not necessarily designed, provided a free and workable solution. Free is open to interpretation, since it does take time to complete the process, and coding can be labor intensive. A librarian's time is limited, so understanding the scope of a project is important to budget time responsibly. Stressing quality control activities throughout the process is necessary. For this evaluation, regularly generating code reports, code and data comparisons, and manipulating data helped to identify typographical and data formatting errors impacting analysis, which were rectified. The potential of the datasets to be published on other venues is notable. Datasets can be shared or embedded into an online community data collection or used to create a domain repository.

For those with general coding experience and familiarity with qualitative analysis, TAMS Analyzer and Viewshare are low barrier to entry tools. For librarians and other scholars with little domain or research methods experience, TAMS Analyzer will take time and effort to learn and use.

This evaluation was successful for text-, image-, and video-based cultural records. Coding audio files was not evaluated, a limitation of the assessment. Since any cultural record is open to humanistic inquiry, further research should evaluate the pairing of TAMS Analyzer and Viewshare when records consist of audio and other items that may need data structured in a different manner. Further recommendations for research include a comparative analysis with proprietary qualitative data analysis tools, the scalability of TAMS Analyzer and Viewshare, an evaluation of TAMS Analyzer and Viewshare when intricate coding schemes are necessary, an assessment of visibility and use of datasets published on Viewshare, how Viewshare and published datasets can be used with other data tools, copyright issues related to published datasets, and software accessibility for those with disabilities.

Data appears to be everywhere and seemingly easy to obtain and manipulate. In a sense, the current environment promoting Open Access and Linked Data culture along with rapid development of new web applications shapes this notion. Librarians tend to have a better understanding of what data are available, where to find it, and barriers to accessing it. Librarians tend to have a clearer understanding of the labor, skills, and time required to make data freely available and when data needs to be created. When this happens, librarians should be prepared to identify and assist with connecting researchers to data and data tools. The evaluation of tools in the literature is key to assist with advisement, and becoming agents of data culture. Understanding domain relevant data barriers and solutions can create a clearer path to access, creating, and sharing. Through technology transfer, data literacy initiatives, and promoting domain data culture, humanities data can grow and be more visible and accessible. When librarians are proactive contributors and collaborators, the visibility of their services and collections increase. New demand and new ways to explore the services and collections can be a result, and confirm the value and relevancy of libraries while promoting scientific discovery and scholarship.

References

Akers, K.G. and Doty, J. (2013), "Disciplinary differences in faculty research data management practices and perspectives", *The International Journal of Digital Curation*, Vol. 8 No. 2, pp. 5-26.

Algee, L., Bailey, J. and Owens, T. (2012), "Viewshare and the Kress collection: creating, sharing, and rapidly prototyping visual interfaces to cultural heritage collection data", *D-Lib Magazine*, Vol. 18 Nos 11/12, p. 3, available at: www.dlib.org/dlib/november12/algee/11algee.html (accessed December 19, 2014).

Anderson, C. (2004), "The long tail", *Wired*, Vol. 12 No. 10, pp. 170-177, available at: http://archive.wired.com/wired/archive/12.10/tail.html (accessed October 13, 2015).

Arlitsch, K. (2014), "Being irrelevant: how library data interchange standards have kept us off the internet", *Journal of Library Administration*, Vol. 54 No. 7, pp. 609-619.

Bailey, J. and Owens, T. (2012), "From record to data with Viewshare: an argument, an interface, a design", *Bulletin of the American Society for Information Science and Technology*, Vol. 38 No. 4, pp. 41-44.

Borgman, C.L. (2009), "The digital future is now: a call to action for the humanities", *Digital Humanities Quarterly*, Vol. 3 No. 4, available at: http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html (accessed December 19, 2014).

Borgman, C.L. (2012), "The conundrum of sharing research data", *Journal of the American Society for Information Science and Technology*, Vol. 63 No. 6, pp. 1059-1078.

Borgman, C.L. (2015), *Big Data, Little Data, No Data: Scholarship in the Networked World*, MIT Press, Cambridge, MA.

Dempsey, L., Malpas, C. and Lavoie, B. (2014), "Collection directions: the evolution of library collections and collecting", *Portal: Libraries and the Academy*, Vol. 14 No. 3, pp. 393-423.

Fitzpatrick, K. (2012), "Giving it away: sharing and the future of scholarly communication", available at: www.plannedobsolescence.net/giving-it-away/ (accessed September 9, 2014).

Gelfand, J.M. and Tsang, D.C. (2015), "Data: is it grey, maligned or malignant?", *The Grey Journal*, Vol. 11 No. 1, pp. 29-40.

Hagood, J. (2012), "A brief introduction to data mining projects in the humanities", *Bulletin of the American Society for Information Science and Technology*, Vol. 38 No. 4, pp. 20-23.

Heidorn, P.B. (2008), "Shedding light on the dark data in the long tail of science", *Library Trends*, Vol. 57 No. 2, pp. 280-299.

Kamada, H. (2010), "Digital humanities: roles for librarians?", *College & Research Libraries News*, Vol. 71 No. 9, pp. 484-485.

Kim, J. (2013), "Data sharing and its implications for academic libraries", *New Library World*, Vol. 114 Nos 11/12, pp. 494-506.

Lynch, C. (2009), "Jim Gray's fourth paradigm and the construction of the scientific record", in Hey, T., Tansley, S. and Tolle, K. (Eds), *The Fourth Paradigm: Data– Intensive Scientific Discovery*, Microsoft Research, Redmond, WA, pp. 177-184.

MacMillan, D. (2014), “Data sharing and discovery: what librarians need to know”, *The Journal of Academic Librarianship*, Vol. 40 No. 5, pp. 541-549.

National Science Board (2005), “Long-lived digital data collections”, available at: www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf (accessed November 4, 2014).

Nicholson, S. and Bennett, T.B. (2011), “Data sharing: academic libraries and the scholarly enterprise”, *Portal: Libraries and the Academy*, Vol. 11 No. 1, pp. 505-516.

Owens, T. (2011a), “Defining data for humanists: text, artifact, information or evidence?”, *Journal of Digital Humanities*, Vol. 1 No. 1, available at: http://journalofdigitalhumanities.org/1-1/defining-data-for-humanists-by-trevor-owens/ (accessed January 10, 2015).

Owens, T. (2011b), “The digital humanities as the DIY humanities”, available at: www.trevorowens.org/2011/07/the-digital-humanities-as-the-diy-humanities/ (accessed November 4, 2014).

Padilla, T.G. and Higgins, D. (2014), “Library collections as humanities data: the facet effect”, *Public Services Quarterly*, Vol. 10 No. 4, pp. 324-335.

Posner, M. (2013), “No half measures: overcoming common challenges to doing digital humanities in the library”, *Journal of Library Administration*, Vol. 53 No. 1, pp. 43-52.

Reilly, S. (2012), “The role of libraries in supporting data exchange”, paper presented at IFLA World Library and Information Conference, Helsinki, August 11-17, available at: http://conference.ifla.org/past-wlic/2012/116-osswald-en.pdf (accessed November 5, 2014).

Schöch, C. (2013), “Big? Smart? Clean? Messy? Data in the humanities”, *Journal of Digital Humanities*, Vol. 2 No. 3, available at: http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/ (accessed January 10, 2015).

Tenopir, C., Birch, B. and Allard, S. (2012), “Academic libraries and research data services: current practices and plans for the future”, available at: www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf (accessed November 5, 2014).

Thessen, A.E. and Patterson, D.J. (2011), “Data issues in the life sciences”, *ZooKeys*, Vol. 150, pp. 15-51, available at: http://zookeys.pensoft.net/articles.php?id=3041 (accessed November 5, 2014).

Tiwari, R., Chand, B.B. and Ramesha (2014), “Open access to research data: sharing data for efficient research”, *Indian Journal of Library and Information Science*, Vol. 8 No. 3, pp. 239-246.

Wallis, J.C., Rolando, E. and Borgman, C.L. (2013), “If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology”, *PLoS ONE*, Vol. 8 No. 7, p. e67332.

Weinstein, M. (2006), “TAMS Analyzer: anthropology as cultural critique in a digital age”, *Social Science Computer Review*, Vol. 24 No. 1, pp. 68-77.

Zuiderwijk, A., Janssen, M. and Davis, C. (2014), “Innovation with open data: essential elements of open data ecosystems”, *Information Polity*, Vol. 19 Nos 1/2, pp. 17-33.

About the author

Linda L. Rath is the Graduate Services Librarian at Newman Library at Baruch College, The City University of New York. Her research explores the roles, tools, and approaches of librarians assisting with knowledge production and cultural heritage participation. She administers the Fantastic Film Festivals: USA and Canada website http://fantasticfilmfests.com, a PSC-CUNY Research Award Recipient project. Linda L. Rath can be contacted at: linda.rath@baruch.cuny.edu