Julie Seifert

INLS 890

March 17, 2014

National Digital Libraries in the Czech Republic and the United States

Digitization and digital libraries present an opportunity for library professionals to

provide widespread access to library materials. In recent years, institutions in both the Czech

Republic and the United States have begun working to digitize materials of cultural significance,

and to provide access to these materials through digital libraries. In creating these valuable

resources, institutions face many challenges and difficult decisions. They must decide, for

example, what to feature in their digital library. They must also choose how many of these

digitized resources can and should be made available to the public, and they must decide whether

they have the legal right to do so. Two institutions, the National Library of the Czech Republic

and the Digital Public Library of America, stand out because of their unique approaches to

creating digital libraries. These institutions face interesting challenges in creating digital libraries

and have often found even more interesting solutions. They have developed unique tools for

creating digital libraries and have often taken a collaborative approach to creating digital

libraries, which allows their collections to be that much richer. Finally, both collections provide

access to valuable cultural heritage materials from their respective countries.

The National Library of the Czech Republic (NKP) began investigating digitization and

digitital preservation in 1992.[1] After joining the UNESCO Memory of the World Program, the

National Library of the Czech Republic began digitizing materials from its own collections, such

---

[1] Brian Rosenblum."Digital Access to Cultural Heritage and
Scholarship in the Czech Republic," Slavic & East European Information Resources,
9:1 (2008): 14

as manuscripts and old prints.[2] According to their website, one of the three main projects of the

National Library of the Czech Republic is the digitizaton of a signifigant part of the what they

call the *Bohemica* from the 19th and 20th centuries, in order to create what they call the National

Digital Library Project.[3] Bohemica refers to books issued in the Czech Republic, written in

Czech or relating to Czech lands. Another goal of the project is "the long-term preservation of

documents in a reliable digital repository, which will provide a space for the safe placement of

already digitised documents as well as the digital documents created or acquired in the [National

Digital Library] project and within other projects."[4] Finally, another goal of the project is

"Making digital documents accessible in a uniform, user-friendly interface with a high degree of

personalisation possible. From a single place, the digitised documents as well as paid online

databases will be accessible."[5] The NKP's digitization efforts have only grown since its first

joined the UNESCO program, and in 2005, it received the UNESCO Memory of the World Jikii

Prize for its contribution to the preservation of the world's cultural hertiage.[6]

     In undertaking a large-scale digitization project, such as the project at the National

Library of the Czech Republic, libraries must consider whether the project will meet the needs of

the users and whether the library has the resources and technical infastructure to support the

project, according to the Cornell Digital Imaging Tutorial.[7] Furthermore, after deciding to move

forward with a digitization project, the library must also decide what materials it will digitize.[8]

---

[2] Rosenblum, 14

[3] National Library of the Czech Republic, *National Digital Library Project*. http://www.ndk.cz/narodni-dk/national-digital-library/view?set_language=en (February 20, 2014).

[4] http://www.ndk.cz/narodni-dk/national-digital-library/view?set_language=en

[5] http://www.ndk.cz/narodni-dk/national-digital-library/view?set_language=en

[6] Rosenblum, 14

[7] Cornell University Libraries, *Moving Theory Into Practice: Digital Imaging Tutorial, Chapter 2: Selection.* http://www.library.cornell.edu/preservation/tutorial/selection/selection-02.html. (March 1, 2014).

[8] http://www.library.cornell.edu/preservation/tutorial/selection/selection-02.html

The library must consider legal issues and whether it is legally allowed to digitize materials.[9] The Cornell Digital Imaging Tutorial has several other suggested criteria for selecting materials for digitization, such as whether the material will lend itself to digitization, whether the material might be harmed by digitization, or if, from a preservation standpoint, it might benefit from digitization.[10] Cornell also suggests that librarians should consider whether it would benefit users to have the materials digitized, and whether the cost of digitizing the materials can be justified.[11] The Library of Congress has similar selection criteria, with some variations. For example, they suggest materials that are highly used might be considered first for digitization. They also give priority to "high-value, at risk materials."[12]

On its website, The National Library of the Czech Republic gives some insight into its selection process, and in many ways it is similar to those suggested by the Library of Congress and Cornell University. The project certainly prioritizes certain materials for digitization over others. The Library lists the documents receiving priority on their website. They are: "Documents of the 19th and 20th centuries which are most threatened by acid-paper degradation. The most user-interesting and most requested publications from 1801 to the present The most user-interesting historical documents until 1801 (including full-text translations). Documents freely accessible to the wider public or through Czech libraries. Other types of documents according to the criteria of significance, uniqueness, physical state, etc."[13] Of course, this begs the question of what is considered "interesting" and who gets to decide that, a question which is not addressed on the Library's website.

---

[9] http://www.library.cornell.edu/preservation/tutorial/selection/selection-02.html

[10] http://www.library.cornell.edu/preservation/tutorial/selection/selection-02.html

[11] http://www.library.cornell.edu/preservation/tutorial/selection/selection-02.html

[12] The Library of Congress, *Preservation Digital Reformatting Program: Selection Criteria.* http://www.loc.gov/preservation/about/prd/presdig/presselection.html. (March 1, 2014).

[13] http://www.ndk.cz/narodni-dk/national-digital-library/view?set_language=en

The digitized documents from the UNESCO project can be seen in the Manuscriptorium portal.[14] Manuscriptorium is a major project of the National Library of the Czech Republic.The Manuscriptorium project is funded by the National Library of the Czech Republic and a private, commercial firm, AiP Beronun.[15] Psohlavec and Karen write that "Manuscriptorium provides access to digitized historical contents held by the Czech National Library and twenty other institutions on the Internet."[16] Thus far, Manuscriptorium contains over 1,700 manuscripts, rare books and old maps, in total over 700,000 images.[17] The goals of Manuscriptorium are "to provide access to historical collections, to provide access and sharing of information on historical collections that were created by us," and "providing access and sharing information on trusted historical collections."[18] Manuscriptorium is a "union catalog" for rare books and manuscripts and also a repository of digitized manuscripts, rare books, and other rare documents.[19] Psohlavec and Karen believe that Manuscriptorium is evolving to the point where "the data created within the project are independent of any specific software products and systems."[20]

The project has a very collaborative focus. In addition to the documents from the National Library of the Czech Republic, Manuscriptorium database also contains materials from more than thirty Czech cultural heritage institutions, as well as some institutions from other countries.[21] For example, Poland, Croatia and Lithuania have all contributed materials to the

---

[14] http://kramerius-info.nkp.cz/digitalizace-v-nk/prezentace-prednesena-na-konferenci-digitale-quellensammlungen-erstellung-2013-archivierung-prasentation-2013-nutzungn-bratislava-8-7-2009
[15] Rosenblum, 15
[16] Stanislav Psohlavec & Vladimír Karen. "Digitised Manuscripts and Rare Prints in the Czech National Library and the Manuscriptorium Project." *New Review of Information Networking,* 12:1-2 (2006): 94
[17] Psohlavec & Karen, 95
[18] Psohlavec & Karen, 95
[19] Rosenblum, 15
[20] Psohlavec & Karen, 96
[21] Rosenblum, 15

database.[22] The National Library of Turkey has also joined the project and contributed more than 10,000 items.[23]

Additionally, the project is also fairly accessible to the public. As Rosenblum writes, "Metadata for the entire Manuscriptorium collection are available to the general public without restriction."[24] However, the full data, which includes all the records, not just the metadata, can only be accessed with an individual or institutional license.[25] An institutional license costs two-hundred dollars per year.[26] Partner institutions, however, receive free access to all the data and the database.[27]

The National Library of the Czech Republic is also involved in web-archiving, through a project called WebArchiv.[28] WebArchiv was launched in 2000 with the goal of providing a "comprehensive, ongoing archive of the Czech World Wide Web."[29] The project estimates that they will collect about 4 billion files, which will be stored in the Arc or WARC format.[30] "The project will produce...1572 files a day."[31] The data from these harvests will be ingested in batches, not daily, and currently the project has not decided on a long-term preservation plan for the data.[32] As they write, "so far the data from web archiving are only backed up, but no active preservation is in place."[33] The project is involved in automated, large-scale harvesting, which harvests the entire World Wide Web of the Czech Republic.[34] In addition to these activities, the

---

[22] Psohlavec & Karen, 96
[23] Rosenblum, 15
[24] Rosenblum, 15
[25] Rosenblum, 15
[26] Rosenblum, 15
[27] Rosenblum, 16
[28] Rosenblum, 16
[29] Rosenblum
[30] National Digital Library. *Web Archiving*. http://www.ndk.cz/webarchiv. (March 1, 2014).
[31] http://www.ndk.cz/webarchiv
[32] WebArchiv- Web Archive of the Czech Republic. *Overview of the WebAchiv project.* http://en.webarchiv.cz/information/. (March 1, 2014).
[33] http://en.webarchiv.cz/information/
[34] Rosenblum, 16

project also selectively harvests certain websites according to certain themes or to reflect certain events.[35] To complete this enormous project, the Czech National Library has partnered with the Moravian Regional library and the institute for Computer Science at Masaryk University, which is located in Brno. These institutions provide technical support to the project.[36]

Unfortunately, there is currently only limited online access to the WebArchiv content, because of legal restrictions in the Czech Republic. Rosenblum writes that, "According to individuals involved with the project, the Czech Legal Deposit Act does not cover Web pages, so Web publishers are not required to provide the NKP with content they publish electronically…"[37] Additionally, Czech copyright law bars the partner institutions from making the archived data available online without permission from the original publisher.[38] Therefore, as it stands now, users can only access the content on computers at the National Libraries and the Moravian Library.[39] The WebArchiv website adds that, although it cannot provide widespread access to the materials, it is still able "to harvest and store online documents to protect them from disappearing forever."[40]

The National Library of the Czech Republic is also involved in a project called Kramerius. According to the Library's website, "Kramerius is open-source program that operates using a content management system…It is designed for the presentation of digital images of the documents of the national cultural heritage through local networks and the Internet, in accordance with copyright law."[41] Kramerius was developed particularly with the idea of

---

[35] Rosenblum, 16
[36] Rosenblum, 16
[37] Rosenblum, 16
[38] Rosenblum, 16
[39] Rosenblum, 16
[40] WebArchiv: Web Archive of the Czech Republic, *Information.* http://en.webarchiv.cz/information/ (March 2, 2014).
[41] Kramerius System. http://kramerius-info.nkp.cz/. (March 2, 2014).

digitizing "more recent acid-paper materials," such as newspapers and journals.[42] The

development of Kramerius was funded by the Czech Ministry of Culture and the Academy of

Sciences Grant Agency.[43] The project began in 2000, and as of December 2006, it contains about

1.2 million pages of digitized materials, and is growing at a rate of about 400,000 pages per

year.[44]

The Kramerius project makes use of many tools in providing its users with access to

digital materials. Many of these tools were developed within the Czech Republic, and some were

developed specifically for this project. For example, the project makes use of a digitization

workflow management system called RDFlow.[45] According to a report on the JoinUp page,

"RDFlow is a digitization workflow management program linked to the Digitization Registry,

which is a system for keeping records of digitized documents and monitoring the digitization

process."[46] The Registry records all the digization activites of libraries in the Czech Republic. Its

purpose is to ensure that there is no duplicate digitization and that libraries do not digitize

materials that have already been digitizied by another instituation.[47] It is also meant to help

libraries collaborate with one another on digitzation, and is operated by the National Library of

the Czech Republic. [48]

Finally, the project uses a tool called ProArc. Developed by the Ministry of Culture in

2011, ProArc is a production and archiving system used to create digital documents that comply

---

[42] Rosenblum, 16

[43] Martin Lhotak. "Digitization and Digital Library Aggregation," Prague Summer Seminar (Prague, CZ), May 22, 2013.

[44] Rosenblum

[45] Lhotak, "Digitization."

[46] JoinUp, *Czech Digital Library and the Kramerius Open Source System,*
https://joinup.ec.europa.eu/software/studies/czech-digital-library-and-kramerius-open-source-system, (March 1, 2014).

[47] https://joinup.ec.europa.eu/software/studies/czech-digital-library-and-kramerius-open-source-system

[48] https://joinup.ec.europa.eu/software/studies/czech-digital-library-and-kramerius-open-source-system

with the standards set by the National Library of the Czech Republic.[49] ProArc also complies

with long-term preservation standards, such as OAIS and ISO.[50]

Like Manuscriptorium, the metadata from the Kramerius project are available freely

online to all users.[51] However, most of the actual digitized images can only be viewed through

computers at the National Library of the Czech Repbulic, because the digitized documents are

protected by copyright law.[52]

In 2006, the NKP worked to combine three of its main projects (Kramerius, WebArchiv

and Manuscriptorium) into one "administrative and technical infrastructure,"[53] known as the

Czech Digital Library. The Czech Digital Library project was funded by the Czech Ministry of

Culture and cost 20 million Czech Crowns (approximately 1 million USD).[54] The project began

in 2012 and is expected to finish in 2015.[55] The main goal of the project is aggregating all the

content from other digital libraries in the Czech Republic.[56] It offers many benefits to users. For

example, with just one search and one website, users can search through all digital data in all

digital libraries in the Czech Republic.[57] As Lhotak says, it allows for "one interface for full text

delivery." It is not just an index but "a dynamic loading of full text sources."[58] Its goal is to

provide "centralized digital sources in the Czech Republic, as defined in the 'Library

Development Strategy of the Czech Republic for 2011 to 2015,'' a strategy written by the Czech

---

[49] Lhotak, "Digitization."
[50] Lhotak, "Digitization."
[51] Rosenblum
[52] Rosenblum
[53] Rosenblum
[54] Lhotak, "Digitization."
[55] Lhotak, "Digitization."
[56] Lhotak, "Digitization."
[57] Lhotak, "Digitization."
[58] Lhotak, "Digitization."

Ministry of Culture.[59] The Czech Digital Library also provides data for international projects, such as Europeana.[60]

As for the question of how materials are selected for digitization, the answer is they are not. The Czech Digital Library is an aggregator for already digitized content. But what are the selection criteria for being featured in the Czech Digital Library? It appears that the only criteria are that the material be from a digital library within the Czech Republic.[61]

Many are impressed by the Czech National Library's ability to quickly adopt digitization techniques and digital library technologies. As Rosenblum writes, "In 1992, three years after the collapse of the communist government and in the midst of the split of Czechoslovakia into two countries, even as they were struggling with other pressing issues, such as poor storage conditions, incomplete cataloging of old volumes, and limited collections budgets, the NKP became one of the original participants in UNESCO's Memory of the World program, an early effort to explore digitization as a means of cultural heritage preservation."[62] However, as mentioned previously, an unfortunate fact is that access to many of these resources, as well as other digital resources, is limited due to legal restrictions. In particular, some are discouraged by the stance on open access in the Czech Republic, stating that "open access is not yet very well established in the Czech environment. There is no pressure from the government bodies and funding providers to support free accessioning of science and research outputs."[63]

In the United States, there have also been many efforts to digitize and make available online items of cultural and historical significance. Of particular interest is the Digital Public Library of America, a relatively new project. As Mitchell writes, "The DPLA is best described as

---

[59] Czech Digital Library, *About the Project*. http://www.czechdigitallibrary.cz/en/. (March 1, 2014).
[60] http://www.czechdigitallibrary.cz/en/
[61] Lhotak, "Digitization."
[62] Rosenblum, 14
[63] Ondrej Fabián, "Open access in the Czech Republic: an Overview," *Library Review*, 62: 4/5 (2013): 220

newly formed and undergoing transition."[64] Additionally, like many of the previously discussed

projects in the Czech Republic, this project is an attempt at a nation-wide digital library. It is

interesting to see how the project compares to the digital library projects in the Czech Republic,

which are also relatively new. These digital library projects provide similar services, but also

differ in many ways. Additionally, they have dealt with similar issues and raised similar

questions, and it is useful to see how the project in the United States has addressed these issues.

According to its website, the Digital Public Library of America "brings together the

riches of America's libraries, archives, and museums, and makes them freely available to the

world. It strives to contain the full breadth of human expression, from the written word, to works

of art and culture, to records of America's heritage, to the efforts and data of science. The DPLA

aims to expand this crucial realm of openly available materials, and make those riches more

easily discovered and more widely usable and used…"[65] This is certainly an ambitious mission.

Mitchell defines the DPLA as "a broad initiative geared toward the development of a unified

digital library of materials in the United States."[66] The DPLA also defines itself through what it

calls its "3 P's."[67] The DPLA is a "portal" which allows users to discovery all these valuable

resources.[68] It is a "platform" that allows people to use the digitized cultural heritage materials in

new ways, by making its API and data available.[69] Finally, the DPLA is "an advocate for a

strong public option in the twenty-first century." That is, they are an advocate for making

intellectual resources available in order to create a public that is intellectually engaged and full of

avid readers.[70] John Palfrey, who became involved in the DPLA early-on, writes that "the initial

---

[64] Erik T. Mitchell. "Three Case Studies in Linked Open Data." *Library Technology Reports 49* (July 2013): 30
[65] The Digital Public Library of America, *About DPLA*. http://dp.la/info/. (March 10, 2014).
[66] Mitchell, 31
[67] http://dp.la/info/
[68] http://dp.la/info/
[69] http://dp.la/info/
[70] http://dp.la/info/

idea is captured in a single sentence, to which the early disscusants [of the library] agreed: 'an

open, distributed network of comprehensive online resources that would draw on the nation's

living heritage from libraries, universities, archives, and museums in order to educate, inform

and empower everyone in the current and future generations'"[71]

The DPLA was formed in 2010 and officially launched in 2013.[72] The DPLA gets its

funding from a number of different organizations and agencies, including the Alfred P. Sloan

Foundation, the Arcadia Fund, The Institute for Museum and Library Services, the National

Endowment for the Humanities, and the John S. and James L. Knight Foundation.[73] The project

has gained more funding and support throughout its lifespan.[74] The DPLA was originally a

project of the Berkman Center for Internet and Society at Harvard University.[75] They received

their initial funding from the Alfred P. Sloan Foundation.[76] With these funds, the Berkman

Center "convinced an initial group of diverse stakeholders to discuss the project, exploring its

concepts, scope, architecture, memberships, cost, administration, and logistics."[77] The project

continued to grow into 2011, and then they received additional funding from the Sloan

Foundation, as well as the Arcadia Fund. Finally, on July 26, 2012 the National Endowment for

the Humanities awarded an additional million dollars to the project.[78] The DPLA has since

grown into its own entity, independent of Harvard University and the Berkman Center.[79] It is

based in the Boston Public Library, although it is also independent from that library, and consists

---

[71] John Palfrey, "A Digital Public Library of America?: Collective Management's Implications for Privacy, Private Use and Fair Use." *Columbia Journal of Law & the Arts 34* (Summer 2011): 838
[72] Mitchell, 31
[73] Mitchell, 31
[74] Paul S. Piper, "The Library's Future is Digital: HathiTrust and the Digital Public Library of America," *Online Searcher* 37.2 (March-April 2013): 22
[75] Piper, 22
[76] Piper, 23
[77] Piper, 22
[78] Piper, 22
[79] http://dp.la/info/

of a staff of eight people.[80] The Digital Public Library is free and open to all users.[81] The Digital

Public Library of America officially launched on April 18, 2013, after two and half years of

development and "many spirited discussions about mission, content, and design."[82] When it

launched, the DPLA provided access to data from eighteen partners, which amounted to over

3,200 collections and over 2 million records.[83] The number of partners and available materials

has continued to increase since the Library's launch.[84] As mentioned previously, the DPLA is

continually growing and changing and is still in a period of "transition."[85]

　　As mentioned previously, the DPLA pulls its content from other institutions. It is an

aggregator.[86] It does not actually hold any of the materials itself and instead acts as a portal

through which users can access materials from many institutions, all through this one platform.[87]

As Howard writes, the DPLA is "not meant to be a virtual equivalent of say, the Library of

Congress, a central storehouse for collections of images and texts. It's not in the business of

preservation. Instead the new digital library acts as a connector or superaggregator." Content

from the partner institutions remains with the partner institutions.[88] Instead, the DPLA harvests

the metadata about those objects and displays that metadata.[89] To do so, the DPLA "takes in

millions of records of items held by libraries, museum, historical societies and other cultural

institutions across the country – more than 1,100 so far. Then it standardizes the records'

---

[80] http://dp.la/info/
[81] http://dp.la/info/
[82] Mick O'Leary, "The Digital Public Library of America Opens Its Doors," *Information Today 30.7* (July-August 2013): 20
[83] Mitchell, 33
[84] Mitchell, 34
[85] Mitchell, 35
[86] Piper, 23
[87] http://dp.la/info/
[88] Micah Vandergrift, "The Digital Public Library of America: Details, the Librarian Response and the Future," In the Library with the Lead Pipe. http://www.inthelibrarywiththeleadpipe.org/2013/dpla/, (March 2, 2014).
[89] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/

metadata and uses it to point searchers toward items relevant to their interests."[90] This is similar

to the Czech Digital Library which does not hold content but instead aggregates it from many

other institutions.

The DPLA has two kinds of partners: the so-called "service hubs" and "content hubs."[91]

The service hubs collect records and sometimes digitize materials from smaller institutions in

their area, and also contribute their own materials.[92] They also host the content. According to

Howard, DPLA Director Dan Cohen has said that "the hubs help 'regularize metadata and get

content into shape, and host scanned and digitized content so that the DPLA doesn't have to.'[93]

The service hubs consist of nine major state and regional digital libraries or library collections,

such as New York's Empire State Digital Network, the Mountain West Digital library, the

Minnesota Digital Library, and the Kentucky Virtual Library.[94] This is in some ways similar to

the system used by the Czech Digital Library, which also aggregates content from other libraries

in the Czech Republic and abroad.

As mentioned previously, the DPLA also partners with what it calls "content hubs."[95]

These are organizations that contribute more than 250,000 unique records to the Library. Unlike

service hubs, they do not collect or host data from other organizations. Some content hubs

include the Internet Archive, the Harvard University libraries, the libraries at the University of

Virginia, ARTstor, the HathiTrust at the University of Michigan, and the New York Public

Library.[96]

---

[90] Jennifer Howard, "The Digital Public Library of America: Young but Well Connected," *The Chronicle of Higher Education,* Dec. 9, 2013. (http://chronicle.com/article/Digital-Library-of-America/143489/).
[91] *The Chronicle of Higher Education,* Dec. 9, 2013
[92] *The Chronicle of Higher Education,* Dec. 9, 2013
[93] *The Chronicle of Higher Education,* Dec. 9, 2013
[94] *The Chronicle of Higher Education,* Dec. 9, 2013
[95] Piper, 23
[96] *The Chronicle of Higher Education,* Dec. 9, 2013

There is a symbiotic relationship between the DPLA and its partners. The partners provide the data for the Library, and the Library, in turn, leads users to its partners.[97] This is because, as mentioned previously, the DPLA is simply a portal for discovery.[98] When users want to view the materials, they are led to a contributor's website.[99] As Howard writes, "[DPLA Director] Mr. Cohen likens the system to an ecosystem in which water flows from pond to stream to ocean in a continuing cycle."[100] This is especially beneficial for smaller institutions. Many contributors of the DPLA have seen increased traffic to the websites since the launch of the DPLA, and participation in the project has led to increased visibility.[101] The wide range of contributors also appeals to users. As Cohen says, "'I think one of the reasons people are liking DPLA is you can find material from a small rural archive alongside things from the Smithsonian…'"[102]

In addition to providing users with access to materials, the DPLA also serves as a platform "through which data and services can be accessed programmatically (APIs)."[103] This means that the public can create applications based on the API of the DPLA, and many have already chosen to do so.[104] According to Howard, the DPLA's API has received 1.7 million hits so far.[105] The Library's websites feature many of the applications that have been built thus far. It is the hope of the DPLA that these applications will be used for further learning, education and

[97] *The Chronicle of Higher Education,* Dec. 9, 2013
[98] http://dp.la/info/
[99] Piper, 22
[100] *The Chronicle of Higher Education,* Dec. 9, 2013
[101] *The Chronicle of Higher Education,* Dec. 9, 2013
[102] *The Chronicle of Higher Education,* Dec. 9, 2013
[103] Mitchell, 31
[104] Mitchell, 32
[105] *The Chronicle of Higher Education,* Dec. 9, 2013

discovery.[106] It appears that the Czech Digital Library does not have a similar feature, and does

not make its API available.

Similar to the Czech Digital Library, the Digital Public Library has grappled with

copyright law and other legal issues in its mission to make public the nation's valuable resources.

In its beginnings, the DPLA had high hopes for making all kinds of information widely

available. John Palfrey, one of the early members of the project, wrote that "what is and can be

great about the digital age" is the ability to spread information through networks.[107] As he said,

"I start with the general posture that it is a good thing for us to try to make more information,

especially that which is published in books, available broadly to many people."[108] However, the

DPLA ultimately chose to include only material that is in the public domain, rather than copy-

righted content, which does not include many published books.[109] In fact, many have noted that

the DPLA contains few books compared to similar projects, such as Project Gutenberg, and is

focused more on cultural heritage materials.[110] As O'Leary writes, "Having copy-righted content

would have created insurmountable obstacles to complete and free access, so the DPLA has

chosen to be a public domain collection, with a mission to curate 'openly available materials,'

according to its website."[111]

In addition to deciding whether they will include public domain or copyrighted content,

digital libraries must also decide on the focus of their collection. What kind of materials will they

feature, and what will they leave out? It appears that, in the case of the Czech Digital Library,

any material may be included as long as it is from the Czech Republic. The DPLA is similar in

---

[106] http://dp.la/info/
[107] Palfrey, 838
[108] Palfrey, 828
[109] O'Leary, 20
[110] Piper, 22
[111] O'Leary, 20

that it has a wide scope. As O'Leary writes, "DPLA does not have a defined collection development policy. Instead, it is a diverse mélange that represents the variety of content in its partner's collections. The simplest way to describe all of this content is to call it 'Americana,' which offers all the depth, richness and sometimes quirky variety that the term suggests."[112] As a result, the DPLA's collection is incredibly diverse. The collection features books, manuscripts, newspapers, articles, and other documents. It also includes photographs, maps, and archival records. "The oldest items are photos of prehistoric art; the most recent are 108 items from 2013."[113]

As an aggregator of an incredibly large amount of content, the DPLA relies heavily on metadata. Metadata allows the Library to organize its contents and, more importantly, it allows users to more easily discover the materials. Therefore, metadata is extremely valuable to the DPLA, and it might benefit from having what O'Leary calls a "robust metadata standard." This would mean asking its partner institutions to comply with strict metadata rules, and to create large amounts of metadata, with many fields, for each of the items they contribute.[114] However, it might be difficult for partner institutions to adhere to these strict standards. The DPLA standard could also conflict with the metadata practices at the partner institutions, and some institutions have their own standards and special practices for metadata.[115] As O'Leary writes, "The emergent standard has the basic discovery elements, but its implementation is uneven."[116] A full metadata record for the DPLA includes title, creator, location (of origin), owning institution, providing institution, creation data, a brief description, format type, subject headings,

---

[112] O"Leary, 20
[113] O'Leary, 20
[114] O'Leary, 20
[115] O'Leary, 20
[116] O'Leary, 20

rights category, thumbnail, and link to the item at its holding institution.[117] However, not every record has an entry for each of these fields.[118] For example, sometimes the creation date is not known.[119] O'Leary also notes that there is not much normalization or standardization among the language used for subject headings. For example, there is the heading "snowboarding" and "Snowboards" and "Snowboarders."[120]

In regards to user discovery, the search system also poses some issues. The Digital Public Library of America has a basic search system, in which results can be limited by owner, partner institution, date, location, subject and format. However, O'Leary notes that "there is no advanced search option in which these elements could be specified." [121] It is not clear how and if the DPLA will improve its search system in the future.

Another concern in regards to the DPLA is the privacy of users. Palfrey suggests that the DPLA might "need to authenticate users into the system and to track their usage in ways more extensive than in the traditional analog-world library model," such as tracking what books users interact with. But individual users might feel that this is a threat to their privacy."[122] As it stands now, users can create an account but they are not required to do so.[123] These free accounts have some benefits to the user, such as allowing them to bookmark materials of interest.[124]

The DPLA has also struggled to form ties with the professional librarian community. Some librarians are wary of the project and are worried that it may draw patrons away.[125] The

---

[117] O'Leary, 20
[118] O'Leary, 20
[119] O'Leary, 20
[120] O'Leary, 20
[121] O'Leary, 20
[122] Palfrey, 840
[123] http://dp.la/info/
[124] http://dp.la/info/
[125] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/

use of the word "public" in the title has been controversial.[126] The fact that it is called a public library "has led some in the [librarian] profession to dismiss it outright, or at least question its motives. The Chief Officers of State Library Association went so far as to issue a resolution that the name be changed."[127] After all, the project was originally based at Harvard, and librarians were not involved in the initial development, to the frustration of many librarians.[128] Additionally, the project did not have a physical space, like public libraries usually do. As Piper writes, "Public libraries are many things to many people, and a large percentage of these are social. Library as space has always been a critical function of a public library, as is its role within a community. Public libraries host numerous events, discussions, and training."[129]

Dan Cohen, the Director of the library, wrote that "the inclusion of 'public' in the title is important in defining the role of this organization in our country's mind. People understand that the items in a public library's collection belong to them and are available for their use."[130] Cohen wants the public to view the Digital Public Library of America in the same way, and he believes that public libraries engender trust, localness, relevance, and familiarity.[131] It was therefore a conscious decision to call the library "Public," in order to signal to the public that the library is theirs.[132] "It is not a public library in the same way the Brooklyn Public Library is," writes Vandergrift, "yet the goals and hopes of the organization are the same." [133] However, it remains to be seen how, if at all, the DPLA will address the social function discussed by Piper.[134]

---

[126] Piper, 22
[127] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/
[128] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/
[129] Piper, 22
[130] Dan Cohen. "The Digital Public Library of America: First Things First," http://www.dancohen.org/2011/10/20/the-digital-public-library-of-america-first-things-first/ (March 10, 2014)
[131] http://www.dancohen.org/2011/10/20/the-digital-public-library-of-america-first-things-first/
[132] http://www.dancohen.org/2011/10/20/the-digital-public-library-of-america-first-things-first/
[133] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/
[134] Piper, 22

Furthermore, in an effort to mitigate concerns among librarians, the DPLA emphasizes that it is not meant to replace public libraries, but to serve as another tool for learning and discovery, and that it's meant to direct people towards libraries.[135] The DPLA hopes that librarians will see that they both (the DPLA and public librarians) share similar goals. Additionally, a recent million dollar grant from the Gates foundation may help public librarians become more comfortable with the DPLA.[136] The grant will finance a training program for public librarians which will give them more digital skills so that they can more easily work with DPLA materials.[137]

Another complaint that some have about the Digital Public Library is that it is not as diverse as it could be, and not as inclusive. As Howard says, "In spite of its rapid growth, the Digital Public Library of America has a lot more expanding to do if it is going to live up to its name."[138] Many parts of the country do not have service hubs and are therefore not represented in the library, so the DPLA hopes to add more hubs in the future.[139] The DPLA also hopes to add more variety in terms of what kind of cultural institutions and objects are featured in the collection.[140] According to Howard, "Mr. Cohen would like to see more audiovisual material represented, for instance, along with more recent books, although how to deal with copyright remains a question mark."[141]

Additionally, some have claimed that projects such as the Digital Public Library of America represent only the intellectual interests of the "elite," and do not truly serve the whole

---

[135] http://www.inthelibrarywiththeleadpipe.org/2013/dpla/
[136] *The Chronicle of Higher Education,* Dec. 9, 2013
[137] *The Chronicle of Higher Education,* Dec. 9, 2013
[138] *The Chronicle of Higher Education,* Dec. 9, 2013
[139] *The Chronicle of Higher Education,* Dec. 9, 2013
[140] *The Chronicle of Higher Education,* Dec. 9, 2013
[141] *The Chronicle of Higher Education,* Dec. 9, 2013

of America.[142] Rothman wonders whose "heritage" exactly we are collecting, and whether it represents a "multicultural, multiethnic America." Furthermore he adds that, "as priceless as our "heritage" is, shouldn't the library also be a repository for the collected wisdom and recollections of the great unwashed of modern America?" Rothman argues that the library should not include just scholarly works, but also non-academic books and multimedia. As he says, "Simply put, let the content be both formal and informal, dynamic and static, popular and academic, cultural and directly practical (culture in the end can be extremely practical, beyond its intrinsic value—if nothing else, in business activities such as marketing and Web design)." He also wonders whether a public library that is online is truly accessible to all, if all do not have access to the internet. Furthermore, he questions whether a library such as the DPLA, which is supported by philanthropic organizations, will really be responsive to the demands of the public."[143]

As you can see, whether in the Czech Republic or the United States or somewhere in between, the creation of a national digital library is a complex undertaking. The libraries must decide on a collection development policy, in order to determine what materials will be included in the digital library, and what will not be. After all, although digital libraries do not face the same space limitations as public libraries, it would still not be beneficial for them to collect every digital item ever created. Furthermore, digital libraries aim to make materials widely accessible, and to provide patrons with access to a large, diverse number of items. However, these goals can often be hindered by copyright laws, and digital libraries grapple with these legal issues in different ways. Furthermore, digital libraries must also decide whether they will work with partner institutions, and what the relationship will be with these institutions. Will the institutions provide the content? Will the digital library assist its partners with digitization and metadata

---

[142] David H. Rothman, "It's Time for a National Digital-Library System," *The Chronicle of Higher Education*, Jan. 24, 2011 (https://chronicle.com/article/Its-Time-for-a-National/126489/).
[143] *The Chronicle of Higher Education*, Jan. 24, 2011

creation? Additionally, in regards to metadata, digital libraries must also determine how they will organize this large volume of materials and how, if at all, they will implement metadata standards. However, despite the challenges, digital libraries also have the potential to provide great value to their patrons. For example, the Czech Digital Library preserves and provides access to valuable historical materials from the Czech Republic, as well as more recent materials, such as websites. The Digital Public Library of America allows users around the world to access materials from the New York Public Library as well as the Minnesota Historical Society. These digital libraries give their users unprecedented access to intellectual materials, and therefore may well be worth the challenges.