Introduction to the Digital Public Library of America API

Unmil P. Karadkar
The University of Texas at Austin, School of Information
1616 Guadalupe St. Ste. 5.202
Austin, TX 78701-1213 USA
+1-512-471-9292
unmil@ischool.utexas.edu

Audrey Altman, Mark Breedlove, Mark Matienzo
Digital Public Library of America
c/o Boston Public Library, 700 Boylston St.
Boston, MA 02116 USA
+1-617-859-2116
{audrey,mb, mark}@dp.la

ABSTRACT
The Digital Public Library of America (DPLA) provides access to over 11 million objects from libraries, museums, and archives. In addition to serving as an open portal for cultural heritage, literature, art, and scientific materials, the DPLA provides access to extensive metadata related to these materials via an openly available, RESTful application programming interface (API). The open API enables third party developers to create targeted applications that enable new and transformative uses of the items indexed by the DPLA. This half day tutorial will introduce participants to the DPLA's data model, describe the API, explain how to retrieve data using the API, and how to work with the retrieved data using freely available software using both interactive and programmatic techniques.

Keywords
Digital Public Library of America; DPLA API; RESTful API

1. INTRODUCTION
The Digital Public Library of America (DPLA) is a large metadata aggregation that provides access to over 11 million objects from over 1,500 libraries, museums, and archives in the USA through its Web portal (http://dp.la) [1]. Items available via the DPLA continue to be hosted by their home institution, which contribute only the metadata to the DPLA. These items cover a variety of topics including textual materials, works of art and culture, records of America's heritage, and scientific data sets. The DPLA is committed to making objects openly available for engaging citizens, students, and scholars, much as the American public libraries have done for the last two centuries.

In this spirit of openness, the DPLA makes all its metadata publicly available via a Web-based application programming interface (API) in order to enable "software developers, researchers, and others to create novel environments for learning, tools for discovery, and engaging apps". This tutorial will introduce attendees to the DPLA API. Participants will learn how to use the

API to access DPLA data via the Web-based interface, to store this data on their disks, view and explore the data using open, freely available software such as OpenRefine, and finally, to write programs for advanced data manipulation.

JCDL '16, June 19-23, 2016, Newark, NJ, USA
ACM 978-1-4503-4229-2/16/06.
http://dx.doi.org/10.1145/2910896.2925428

2. FORMAT AND SCOPE
This will be a half day tutorial.

Prior to the tutorial, we will send instructions to participants for pre-installing the necessary software on their computers as well as assist them in installing the software on the day of the tutorial.

During the tutorial will introduce the participants to the DPLA API and train them to use the API to retrieve and manipulate data without programming as well as using the Python and JavaScript wrappers that simplify programming tasks. Attendees will be given small programming tasks that require fetching and manipulating DPLA data.

2.1 Aims
This tutorial aims to introduce the DPLA conceptual data model, DPLA API, JSON data format, and command-line as well as programmatic interfaces for retrieving DPLA metadata.

2.2 Learning Objectives
This tutorial will enable the participants to:

- Understand and use RESTful APIs in general and the DPLA API in particular
- Locate relevant information on the DPLA developer pages
- Read the DPLA data model documentation
- Retrieve DPLA metadata via a Web browser and a command-line interface
- Manipulate saved DPLA data using OpenRefine
- Retrieve DPLA data using a modern programming language such as Python, PHP, JavaScript, or Java.
- Use Python wrappers such as dpla_utils and DPyLA – for Python programmers
- Use the JavaScript wrapper for JS programmers
- Retrieve and merge data from multiple pages
- Select, sort and count data values programmatically

3. TOPICAL OUTLINE
Here is a tentative list of topics that the tutorial will cover:

- Introduction to the DPLA and the Web portal
- Introduction to the developer pages
- Introduction to the data model (Metadata Application Profile-MAP)
- Introduction to the API and RESTful architecture
- Requests and responses
  - Sending a simple request
  - Receiving a simple response
  - Hands-on examples

- Introduction to JSON
- Viewing retrieved data with JSON browser plug-ins
- Selecting, sorting, and searching within fields
- Pagination
- Faceting
- Hands-on examples of advanced requests
- Saving the retrieved data to disk
- Viewing and manipulating data with OpenRefine
- Introduction to programmatic data retrieval
- Introduction to Python and JavaScript wrappers (DPyLA, dpla_utils)
- Introduction to JavaScript wrapper

4. TARGET AUDIENCE

As this is the first tutorial that we are proposing to be held during a major conference, we are unable to estimate the level of interest in the community. We expect that a large segment of the JCDL attendees will be interested in the topic. The location of the conference and its proximity to several universities in the Washinton, DC-Philadelphia-NJ-NYC metropolitan areas will help us attract students from these institutions.

The target demographic for this tutorial includes practitioners and scholars in Digital Libraries who are interested in working with DPLA data or those who want to learn about RESTful APIs in general.

4.1 Prior Knowledge

This is an introductory tutorial. No prior knowledge is expected of the audience.

5. EQUIPMENT NEEDS

The instructors will need access to the internet and a projector. We will bring our laptop computers to host the presentation. We will encourage participants to bring their laptop computers in order to follow along or complete tasks in the hands-on components of the tutorial.

6. INSTRUCTOR BIOGRAPHIES

Unmil P. Karadkar (unmil@ischool.utexas.edu) is an Assistant Professor in the School of Information at The University of Texas at Austin. He situates his work at the intersection of digital libraries, human-computer interaction, and visualization. He studies data practices of researchers with an eye toward identifying unmet information needs. Based on an understanding of these needs, he designs software to support their evolving practices and evaluates the impact of this software on their work. In the School of Information, Unmil teaches courses in Digital Libraries, Visualization, and Metadata Generation for Large Datasets and has used as well as trained his students to use the DPLA API.

Audrey Altman (audrey@dp.la) is a Developer for DPLA. She works with the DPLA Technology Team to design, develop, test, integrate, support, and document user-facing applications and back-end systems; support content management policies, process, and workflows, and contribute to the development of new ones; and collaborate with stakeholders to contribute to strategic and tactical planning and implementation of content stewardship applications and technologies. Audrey previously worked as a web developer for Digital Research and Publishing at the University of Iowa Libraries, and for the University of Iowa Digital Studio for Public Arts & Humanities. She holds a Masters of Library and Information Science from the University of Iowa, a M.A. in American Studies from the University of Alabama and a B.A. in Theater from Albion College. While Audrey doesn't have a single favorite DPLA item, she is partial to zoological drawings like those of the White Heron, Squat Lobster, Snail, and Collared Hedge Hog.

Mark Breedlove (mb@dp.la) is a Senior Developer for DPLA, who contributes to the design and implementation of the organization's ingestion, API, and front-end website. His work has been concentrated in the development of the ingestion system, which moves data from providers to the DPLA datastore, and the development of the organization's new DevOps systems. He works closely with the Director of Technology and the Content team. Before coming to DPLA, Mark was the Technical Director at See.me, a social discovery website for artists, and the American Museum of Natural History, building scientific web applications and content management systems for its Science Division. Mark's current favorite DPLA item is Gunn's Domestic Medicine, or Poor Man's Friend.

Mark Matienzo (mark@dp.la) is the Director of Technology for DPLA. As Director of Technology, Mark is responsible for the overall technology vision for the DPLA and overseeing its implementation. Mark also serves as the primary technical contact for outside organizations, partners, and developers. Prior to joining DPLA, Mark worked as an archivist and technologist specializing in born-digital materials and metadata management, at institutions including the Yale University Library, The New York Public Library, and the American Institute of Physics, and participated in projects such as the ArchivesSpace open source archival management system and AIMS – Born Digital Collections: An Inter-Institutional Model for Stewardship. Mark received a MSI from the University of Michigan School of Information and a BA in Philosophy from the College of Wooster, and was the first awardee (2012) of the Emerging Leader Award of the Society of American Archivists. Mark's current favorite item in DPLA is Children in Goat Cart.

7. REFERENCES

[1] Digital Public Library of America – About. 2016. Accessed Jan. 5, 2016, http://dp.la/info/.