Science Stories: Using IIIF and Wikidata to Create a Linked-Data Application

Katherine Thornton and Kenneth Seals-Nutt

Yale University
katherine.thornton@yale.edu

Abstract. Images capture human attention. When digital images circulate on the web, people often gather information from the visual content alone. There may be embedded metadata in an image, perhaps automatically generated by the digital camera the photographer used. People do not often embed descriptive metadata into digital images they share online. The standards for image application programming interfaces (APIs) that the International Image Interoperability Framework (IIF) created provide guidance for how to communicate descriptive, technical, and administrative metadata along with images on the web. We will demo an application in which we combine images and structured data to tell stories. We tell science stories in this application related to the lives of women involved in scientific research, many of these stories have not yet been told on the web, we aim to raise their profile.

Science Stories is a linked data application powered by structured data. The facts about the lives of the women whose science stories we are telling are syndicated from the Wikidata knowledge base. Content stored in Wikidata is machine-readable linked open data, and is published under a Creative Commons Zero license. We added references to the statements we contributed to Wikidata to published source materials so that others can quickly locate additional information.

We demonstrate how free software and open standards can be leveraged to create visually appealing, interactive, information experiences which allow people to extend science communication to additional social spaces on the web. Images are an important component of scientific communication. Images draw human viewers in, and captivate interest. Science communication of the twenty-first century allows us to reconnect images with descriptive, technical, and administrative metadata in a web presentation, and to combine these images with structured statements in the web of linked data, backed by references to published sources. This application combines structured data from Wikidata, and images published in conformance to the IIIF specifications. This novel combination of conceptual and technical infrastructure allows us to explicitly connect these science stories to reference materials.

Keywords: Wikidata, scholarly communication, linked open data, International Image Interoperability Framework

1 Introduction

Science Stories brings scientific work into social spaces. Communicating scientific information in spaces where people can discover it in-context creates a jumping-off point for further exploration. For an institution with cultural heritage collections, there are many scientific resources held in libraries, archives, museums and galleries that are not yet available on the web. We provide a web application that leverages linked open data, the International Image Interoperability Framework, the Wikidata knowledge base, and semantic web technologies to demonstrate a vision of what getting scientific work products and scientific information into social spaces can do.

2 Description of the App

We developed a web application named Science Stories for this project[1]. The application is based on the Model-View-Controller paradigm [1]. We used Express[2], a Node.js framework and the Handlebars[3] templating engine to render HTML in combination with Bootstrap[4], PagePiling.js[5] and the jQuery library[6]. We generate manifests for the images in our application via our custom Science Stories API. We use Mirador[7] to display IIIF manifests in the application. For app-specific content, we use a Postgresql[8] database in combination with Sequelize[9], the Object-Relational-Mapping for Postgresql with Node.js. The code for the application is available via GitHub[10]. Supporting the content of our application, we built a High-Performance Cantaloupe image server to serve static media from our Amazon Web Services Bucket instance containing our raw content and locally developed IIIF manifests. For those wanting to produce content inside our application, we have created an annotation server that performs as a filter between existing manifests and our application view. Authors enter into the annotation tool, where they can markup media to provide context describing their contents. To enhance this functionality, we developed a client-side JavaScript function that allows users to simply label annotations by their Wikidata Identifiers in the format, wdAnnotation(Q123), and when the manifest is then presented in a Science Story, the client will convert the label into a rich knowledge graph card that quickly aggregates data about the specified object in real time using the Wikidata API for structured quick facts information, Wikimedia Commons for

[1] http://www.sciencestories.io
[2] https://expressjs.com/
[3] http://handlebarsjs.com/
[4] https://getbootstrap.com/
[5] https://alvarotrigo.com/pagePiling/
[6] https://jquery.com/
[7] http://projectmirador.org/
[8] https://www.postgresql.org/
[9] https://www.npmjs.com/package/sequelize
[10] https://github.com/ScienceStories/application

a descriptive image, and the Wikipedia API for the summary to provide some free-text descriptors of the object. The Science Stories application also offers a full graphical user interface tool for authors to create, draft edit, and populate Stories all from an easy-to-use web-based Story Builder where the API converts the elements of the form into a well-structured JSON object inside our database.

Stories are presented as a series of moments, and each moment is illustrated by an image, or set of images, or related media files. Users can progress through the moments of a story, or may choose to branch off and follow pathways linked from media objects themselves. This interaction is enabled by the reuse of URIs from the Wikidata knowledge base.

3 Semantic Web Powered Science Stories

The data that we curated is now in Wikidata for reuse by anyone. All of the SPARQL queries we wrote are available for reuse[11]. These queries will become valuable as people continue to add data to Wikidata which will cause these queries return additional results.

These items are available for future enrichment- as institutions that have not yet published info about their collections on the web decide to do so, these items are ready and waiting for them, allowing for structured statements to accumulate productively.

Of the women affiliated with Yale University in some scientific capacity prior to 1969, many of them did not have Wikipedia articles or Wikidata items about them before the start of our project. Information that exists in print and has not yet been published on the web requires more time-intensive curation than information that is already on the web in some form. More people will now have access to information about these women than ever before.

The pathways created within Science Stories can be used to guide researchers. Links to persona papers from archival collections, primary and secondary sources, and media files related to these individuals will reduce some of the labor in the initial stages of research, potentially allowing people to explore more deeply.

Structured data connected to references indicating provenance, such as the data we curated for the the Science Stories project, could support the work of exhibit staff in libraries, archives, galleries and museums. Digital exhibitions can present images that will draw viewers in, and the metadata presented along with the images provide pathways for further exploration of collection material and historical, social, or scientific context that support sensemaking. Exhibit staff need tools to support the work of creating digital exhibits, we demonstrate how combinations of free software tools and semantic web technologies can be combined to create exhibits with high-resolution images and deep zoom and the work of cataloging, metadata, and curatorial staff can be easily leveraged to present descriptive metadata that bridge the "semantic gap" of images that are not self-describing [2].

[11] `https://github.com/ScienceStories/SPARQL`

4 Conclusion

We promote the presence of science stories and scientific data in social spaces. Young people gain exposure to new ideas from participating in social spaces on the web. Scientific outreach that extends to social media may attract the next generation of scientists.

Our system demonstrates the value to users of aggregating metadata from different institutional collections. Yale University's rich collection of cultural heritage materials are located and managed by the many libraries, archives, museums and galleries across campus. As more institutions begin to digitize collection materials and to publish metadata for collection material on the web, we can create systems that allow for interoperability between different metadata schemes and styles. For users this means that they will no longer need to repeat the same search across multiple databases or websites to find material related to their research questions. This also helps to raise the profile of the institution's collections as a whole because it will minimize the risk that a user might overlook collection material that is administratively separated by contextually connected.

Connecting documents, media files, and unambiguous identifiers for people, places, things and ideas is possible through web-base publishing integrating technologies of the semantic web. The infrastructure of our project serves as a reference model. We use free software, open standards, and freely-licensed content in order to model open science as a practice that maximizes the potential for worldwide audiences to engage with resulting content.

5 Acknowledgements

We would like to thank our colleagues at Yale University Library. We appreciate the support of Holly Rushmeier, David Skelly, Larry Gall, Eric Lazo-Wasem, Patrick Sweeney, Nicole Palffy-Muhoray, Tim White, Emmanuelle Delmas-Glass, Michael Appleby, Alison Clemens, the IIIF community, the Wikidata community, and the Wikicite community.

References

1. Krasner, G.E., Pope, S.T., et al.: A description of the model-view-controller user interface paradigm in the smalltalk-80 system. Journal of object oriented programming **1**(3) (1988) 26–49
2. Shotton, D.: Data webs for image repositories. In Wouters, P., Beaulieu, A., Scharnhorst, A., Wyatt, S., eds.: Virtual Knowledge: experimenting in the humanities and the social sciences. MIT Press (2012) 98–101