

Midwest Big Data Hackathon

September 16-17, 2017

Iowa City IA, USA

bigdata.uiowa.edu



Datasets + Tools

- [U.S. Government's Open Data](#) - Provides open data in various topics such as agriculture, health, climate, energy, etc.
- [CUAHSI Resources and Tools](#) - Provides national-scale data services and resources for hydrology and climate.
- [Twitter REST API's](#) - Provide programmatic access to read and write Twitter data.
- [Yelp](#) - 2.7M reviews and 649K tips by Yelp's users and user businesses.
- [BestBuy Open Product Data](#) - Programmatic access to BestBuy's product and store data.
- [The New York Times API's](#) - Programmatic access to The New York Times' content. It includes articles, books, events etc.
- [The Guardian API's](#) - Programmatic access to The Guardian's content.
- [GDELT Database](#) - Most comprehensive database of human society.
- [MIT Human Dynamics Lab Datasets](#) - Several mobile data sets that contain dynamics of several communities of about 100 people each.

- [Million Song Dataset](#) - Collection of audio features and metadata for a million contemporary popular music tracks
- [Wikipedia Dumps](#) - Wikipedia database dumps in any language.
- [Google Books Ngram Dataset](#) - The dataset that used in Google Books Ngram Viewer.
- [International Cancer Genome Consortium](#) - Dataset that includes information about cancer patients and their genomes.
- [GeoNames Data](#) - Geographical database covers all countries and contains over eleven million place names.
- [NOAA Ocean CO₂ Dataset](#) - More than 30 years worth of surface ocean CO₂ observations.
- [NOAA Climate Datasets](#) - Datasets that are provided by NOAA on climate.
- [The Movie DB API](#) - Allows programmatic access to TMDb that includes information about movies, tv shows, actors and actresses.
- [US Patent Citation Network](#) - Consists of all the utility patents granted between January 1, 1963 and December 30, 1999. totaling 3,923,922 patents
- [HAZUS](#) - Consists of hazard data, boundary map data, and a proxy for the general building stock (GBS) in the continental United States, Hawaii and the U.S. held Territories.
- [ACT](#) - Test events dataset

Project Ideas

PROJECT: PROTOTYPE BIBFRAME 2.0 MODELING FOR LIBRARY INFO SPOTLIGHT "OPERA PLANET"

Brief explanation:

Library linked data promises to meet libraries' need for agility in content delivery and user engagement. This project will demonstrate the prototype of BIBFRAME 2.0 modeling for work, instance, item, agent, topic, etc. from the local bibliographic records in Alma and external data

sources. The prototype will represent library info spotlight of operas in Opera Planet, a collection of opera books, videos, sound recordings, streaming media, etc. interwoven into user's online experience using Alma, LC MARC to BIBFRAME Transformation Service, RDF Translator, RDF Validator, RDFa Validator, Apache Jena Fuseki Server, etc.

The initial data set consists of six library bibliographic data in linked data model using BIBFRAME 2.0, representing an opera titled "The Marriage of Figaro." The sample files are in RDF/XML, JSONLD, RDFa, and N3 formats. More info about BIBFRAME 2.0 can be found in the overview of the LC BIBFRAME 2.0 Model, available from <http://www.loc.gov/bibframe/docs/bibframe2-model.html>

Expected results:

- Make library info spotlight of operas into linked data model using BIBFRAME 2.0 and extended vocabularies, visible to the web, and easily engaged with library users.

Knowledge Prerequisite:

- Bachelor's degree in computer sciences or related field or equivalent combination of education and experience;
- 3 years' minimum Web applications or systems development experience;
- Proficiency with programming languages such as AngularJS2.0, PHP, Ruby, Python, jQuery, HTML, JavaScript and CSS;
- Knowledge of triple stores, SPARQL, and RDF;
- Experience performing data transfers utilizing software library or language APIs; and
- Working knowledge of Apache, Amazon Web Services, etc.

Attachments: [Detailed Project Description](#)

Mentor: Amanda Xu / University of Iowa Libraries

PROJECT: CUAHSI PROJECT IDEA 1

Brief explanation:

Synthesize information from multiple data sources to show health of, or change in, a region of the US. This would require teams to collect and aggregate data from CUAHSI and non-CUAHSI sources and develop an easy to use, map-based, interface for citizen users to view the results.

Attachments: [Detailed Project Description](#)

PROJECT: CUAHSI PROJECT IDEA 2

Brief explanation:

A Mobile alert application for flood conditions as predicted by the National Water Model. This would require teams to develop an efficient algorithm for collecting, processing, and search large datasets to identify regions of high flood potential and present the findings via a map-based interface. This could result in a prototype of a app that citizens would use to determine if their county is at risk of flooding.

Attachments: [Detailed Project Description](#)