DPLAFest 2016 Hackathon Resources


Schedule

- **Thursday, April 14** (Dining Room A, Library of Congress Madison Building)
  - 10:45-11:00: Brief recap and intros
  - 11:00-12:30: Time to work on hackathon projects
  - 12:30-1:45: Lunch (feel free to continue working!)
  - 1:45-5:00: Time to work on hackathon projects
  - 6:30-8:30: Reception (National Archives Rotunda Galleries)
- **Friday, April 15** (Innovation Hub, National Archives - Pennsylvania Avenue entrance)
  - 9:30-9:45: Brief recap and intros
  - 9:45-12:45: Time to work on Hackathon projects
  - 12:45-1:45: Lunch (feel free to continue working!)
  - 2:00-2:45: [Developer Showcase](#) (National Archives McGowan Theater)

Resources

- [Hackathon Resources Slide](#)
- [DPLA Developer Showcase Signup](#) for lightning talks (Friday at 2 pm)
- IRC: Freenode.net, #dpla channel (use [http://webchat.freenode.net/](http://webchat.freenode.net/))
- Tools to help you:
  - [Postman](#) and [JSONView](#) for Chrome (install at least Postman)
  - [RESTClient](#) and [JSONView](#) for Firefox (install at least RESTClient)
- DPLA developer documentation
  - [Introduction to the DPLA API presentation](#)
  - API codex: [http://dp.la/info/developers.codex/](http://dp.la/info/developers.codex/)
  - Demo & temporary API Key: **219d3d0c9a96c56c1798ea298aac12f4**
  - Get your own key: [http://dp.la/info/developers/codex/policies/#get-a-key](http://dp.la/info/developers/codex/policies/#get-a-key)
- **DPLA MAP**
- **DPLA App Library**: [http://dp.la/apps](http://dp.la/apps)
  - Submit your app to the App Library by emailing us at apps@dp.la or through the contact form.
- Creating a GitHub Repository (optional way to share your code)
  - [https://help.github.com/articles/create-a-repo/](https://help.github.com/articles/create-a-repo/)
• National Archives Catalog API
  ○ GitHub: https://github.com/usnationalarchives/Catalog-API
    ■ Documentation
    ● Search and export
    ● Posting contributions
  ○ API: https://catalog.archives.gov/api/v1
  ○ Interactive API sandbox: https://catalog.archives.gov/interactivedocumentation
  ○ API Reference Guide (DRAFT)
  ○ Questions: developer@nara.gov
  ○ Catalog UI: https://catalog.archives.gov/
  ○ National Archives Developer Hub: https://www.archives.gov/developer/

• Historypin API developer documentation
  ○ http://www.historypin.org/resources/docs/api/site/index.html
  ○ Slides from today. Most read-only operations do not require an API key -- write operations do, however, and let us know if you need some details on how to get your API key / secret key from your Historypin user profile page. Ping @bfk for more info. In general, the Historypin API can do everything that you can do as a user at Historypin.org, and generally uses the same URL syntax as normal Historypin page URLs.

• Europeana API
  ○ slides from today
  ○ http://labs.europeana.eu/api
  ○ Search API: http://labs.europeana.eu/api/search
  ○ Search fields: http://labs.europeana.eu/api/query/
  ○ Record API: http://labs.europeana.eu/api/record
  ○ Data fields: http://labs.europeana.eu/api/data-fields
  ○ Showcased open datasets in Europeana: http://labs.europeana.eu/data
  ○ Europeana also has SPARQL endpoints (via data.europeana.eu, working this week) and an OAI-PMH repository server available (not working this week!)

• NYPL API
  ○ Slides from today
  ○ Digital Collections
  ○ http://api.repo.nypl.org/
  ○ Sign-up is required to use the NYPL API. For questions during the hackathon, tweet with the hashtag nyclAPI or email shawnaverkamp@nypl.org, and someone will help you out.
  ○ API Wrappers:
    ■ Javascript: https://github.com/NYPL-publicdomain/api-client
    ■ Ruby: https://github.com/timwaters/nypl_repo
Notes

- **Wednesday, April 13** (Mumford Room, 6th Floor, Library of Congress Madison Building)
  - Introduction to APIs workshop
  - 1:30-1:45 Intros (Jon Voss @jonvoss)
  - 1:45-3:00 Presentations on cultural heritage APIs
    - DPLA (Audrey Altman and Tom Johnson @no_reply)
    - National Archives (Jason Clingerman)
    - NYPL (Shawn Averkamp @saverkamp)
    - Europeana (Antoine Isaac @antoine_isaac)
    - Historypin (Breandán Knowlton @bfk)
  - 3:00-3:15 Followup panel Q&A for presenters (facilitated by Scott Williams @moltude, DPLA)
  - 3:15-3:30 Break/informal discussion
  - 3:30-4:00 Project ideas from the audience (facilitated by Mark Matienzo @anarchivist & Mark Breedlove)
  - 4:00-4:45 Informal Q&A, helping people get set up with API keys, trying to convene project teams/groups
  - 4:45-5:00 Wrapup; plus/delta; recap of plans for Thursday and Friday

Potential Seed Ideas:

- **DPLA**
  - Review DPLA (and other) app libraries
  - API SDK work (++ tom) or tools to work with data (++ tom; for blazegraph)
  - Documentation
  - Visualizations: # of search results by hub

- Intro API curriculum (easy application, step to step; documentation; lesson plan)

- Build an app that shows pictures from a location
  - Parsing location information out of metadata (jonvoss: would love to do some of this on a specific WWI dataset)

- Populate transcription field with OCR text from National Archives Catalog API
  - OCR text: @extractedText in API data
  - PUT in transcription:
    https://catalog.archives.gov/api/v1/id/521451/transcriptions?text=
  - Possible targets:
    - Department of Defense Directives, 2009

- **Audience ideas**
  - Historypin - Parsing/Refining WWI location data from photographs - would be ni
  - Deep learning for geolocation determination
- Tensorflow?
  - Immersive virtual reality experiences
    - VR app about African-American veteran
  - Who’s on First for geographic name reconciliation
  - Image recognition
    - Face recognition
      - Candidate collections:
        - NYPL -- Friedman-Abeles Photos (Performing Arts)
        - NYPL -- Portrait File Collection (17th-20thc. historic figures)
  - Identify interesting images
  - Identify generic images
  - DPLA data dump - moving it into a high throughput triplestore
    - Analysis of Metadata in aggregate
      - Supporting statistical analysis and researchers with DPLA data
      - Europeana: http://144.76.218.178/europeana-qa/ (from Péter Kiraly)
      - Generalized statistical framework?
  - Identifying texts in DPLA
  - Content cross-promotion
    - OutBrain/LibBrain
  - Maps
    - Sanborn Maps - NYPL → Historypin?
      - NYPL request for public domain Sanborn Maps: search query for Publisher="Sanborn Map Company":
        http://api.repo.nypl.org/api/v1/items/search?q=%22sanborn%20map%20company%22&field=publisher&publicDomainOnly=true&per_page=500
      - From that results list, add UUID to this syntax:
        http://api.repo.nypl.org/api/v1/items/item_details/7b02f031-4418-dc8c-e040-e00a180646c5
      - You should see all available image links and link back to Digital Collections (mapping to Historypin template)
    - Creating a 3D environment for content harvested by DPLA partners. Demo for Google Cardboard available at http://labs.cogapp.com/tc
    - Autocomplete for Getty vocabularies and other controlled vocabularies?
      - Immersive Environment notes:
        - Stereographs are a great source for 3D. E.g. http://stereo.nypl.org/
**Metadata QA Analysis**

- **JSON Schema Validation**
- Europeana quality assessment experiment:
  - Statistics for quality assurance: [http://144.76.218.178/europeana-qa/](http://144.76.218.178/europeana-qa/) (from Péter Kiraly)
  - SHACL rules and reports for EDM:

- Google Analytics concept: [https://www.umbrasearch.org/contributing-institutions](https://www.umbrasearch.org/contributing-institutions)
  - Track items per data provider
  - Track engagement location (did a user click on “view original” in search, on a full record page, etc)

**HACKATHON RESULTS**

- **StoryHub**
  - Create wonderful stories from collections available on NARA, NYPL, Historypin, DPLA, Europeana
  - Does real-time API queries on selected sites (default = search all APIs)
  - A search-result based story is created on-the-fly
  - Only stories and narratives get stored on StoryHub’s server, objects are real-time API calls
  - How-to and more details at the bottom of the home-page
  - User: demo1 / 123456 for the login
  - We’ll enable account registrations and new UI/UX templates next week … after we get some sleep :)