DATA-FIRST MANIFESTO

Shifting Priorities in Scholarly Communications

Clifford Anderson
Associate University Librarian
Vanderbilt University
@andersoncliffb
DATA - FIRST MANIFESTO

1. Down the API
2. Interfaces are perspective
3. Publication/credit:
   a. Minimal computing
   b. Preservation
   c. Sustainability
   d. Integrity
4. Publish now!!

- Watermark data?
- Unstructured data, denormalized data?
- What about utility?
- How does this data resource allocation?
- Coordination with a big boss
Data-First Manifesto

In this spirit of the Agile Manifesto, we are seeking to develop more agile, collaborative, efficient, and transparent practices in digital scholarship. To those ends, we propose emphasizing:

- **Data Serializations** over Databases
- **Application Programming Interfaces** over Graphic User Interfaces
- **Curating Data** over Summarizing Information
- **Citing Datasets** over Referring to Results

In the words of the Agile Manifesto,

“That is, while there is value in the items on the right, we value the items on the left more.”
DATA SERIALIZATIONS
OVER DATABASES
**HeardLibrary / corpus-baudelaire**

An experiment in collaborative TEI text encoding — Edit

- [911 commits](#)
- [1 branch](#)
- [0 releases](#)
- [18 contributors](#)

Branch: master

<table>
<thead>
<tr>
<th>Branch Name</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>master</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This branch is 49 commits ahead, 1 commit behind CliffordAnderson:master.

<table>
<thead>
<tr>
<th>Ramona2020 committed on GitHub</th>
<th>Merge pull request #191 from AngelaSutton/master</th>
<th>Latest commit f1336f7 8 days ago</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editions</td>
<td>A UNE MENDIANTE ROUSSEY 1868 Edition</td>
<td>2 years ago</td>
</tr>
<tr>
<td>Misc</td>
<td>moved two files</td>
<td>a year ago</td>
</tr>
<tr>
<td>WebApp</td>
<td>Refactored corpus module</td>
<td>a year ago</td>
</tr>
<tr>
<td>fleurs-du-mal</td>
<td>Merge pull request #191 from AngelaSutton/master</td>
<td>8 days ago</td>
</tr>
<tr>
<td>license.md</td>
<td>Create license.md</td>
<td>a year ago</td>
</tr>
<tr>
<td>readme.md</td>
<td>Update readme.md</td>
<td>2 years ago</td>
</tr>
</tbody>
</table>
Application Programming Interfaces over Graphic User Interfaces
OPenn contains complete sets of high-resolution archival images of cultural heritage material from the collections of its contributing institutions, along with machine-readable descriptive and technical metadata. All materials on OPenn are in the public domain or released under Creative Commons licenses as Free Cultural Works. Please see specific collections and documents for applicable license terms.

<table>
<thead>
<tr>
<th>Name</th>
<th>Last modified</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collections.html</td>
<td>2017-02-15 19:01</td>
<td>16K</td>
</tr>
<tr>
<td>Data/</td>
<td>2017-02-17 14:48</td>
<td></td>
</tr>
<tr>
<td>ReadMe.html</td>
<td>2017-02-18 18:16</td>
<td>11K</td>
</tr>
<tr>
<td>Repositories.html</td>
<td>2017-02-18 18:16</td>
<td>16K</td>
</tr>
<tr>
<td>TechnicalReadMe.html</td>
<td>2017-02-18 18:16</td>
<td>87K</td>
</tr>
<tr>
<td>robots.txt</td>
<td>2015-09-10 14:54</td>
<td>24</td>
</tr>
</tbody>
</table>
Curating Data
over Summarizing Information
CHARLES BAUDELAIRE
UNE MICRO-HISTOIRE
CHRONOLOGIE BAUDELAIRENNE

LIBRAIRIE JOSE CORTI
1987
Citing Datasets over Referring to Results
http://bioimages.vanderbilt.edu/

What is Bioimages?
Bioimages is:
- a collection of high quality images of living organisms.
- a set of tools for learning about plants and ecoregions.
- a biodiversity database
- a demonstration of biodiversity informatics best-practices.

How do I use Bioimages?
You can explore Bioimages by clicking on images, buttons, or links starting with those on the left side of this page. Usually, if you click on a thumbnail or image it will take you to a higher quality image and additional information about that image or something related to it.

All images on the Bioimages site are available under a license no more restrictive than Creative Commons Attribution-Noncommercial-Share Alike (CC BY-NC-SA). That means that they can be used for non-commercial purposes without asking permission and with proper attribution.

All image and organism metadata are available without restriction (i.e. under CC0 - No rights reserved). The data are available at the Bioimages GitHub repository.

http://datamanifesto.info/