Zenodo Data Archive

Irtiza Delwar, Michael Culhane, John Sizemore, Gil Turner
Client: Dr. Seungwon Yang
Instructor: Dr. Edward A. Fox
CS 4624 Multimedia, Hypertext, and Information Access
Virginia Tech
Blacksburg, VA 24061
April 27, 2017
Goals at the beginning of the semester:

- Set up a data archive application (Zenodo) on a server running CentOS 7 operating system
- Configure and populate archive with datasets
- Create detailed Admin / User Guide
- Add additional scripts for bulk uploads

👍 All goals completed!
Timeline

Figure: Gantt Chart representing the timeline of our project
Design

Tools Used:
- Linux CentOS
- SSH
- Zenodo
- DSpace
- PostgreSQL
- Python
- Zenodo Developer API

Data Flow Design
Refinement

Python Scripts

Zenodo Developer API

Single Upload

Bulk Upload

Python 2 v Python 3

Delete?

```python
>>> data = {
    ...     'metadata': {
    ...         'title': 'My first upload',
    ...         'upload_type': 'poster',
    ...         'description': 'This is my first upload',
    ...         'creators': [{'name': 'Doe, John',
    ...                         'affiliation': 'Zenodo'}]
    ...     }
    ... }
>>> r = requests.put('https://zenodo.org/apideposit/depositions/%',
    ...                   params={'access_token': ACCESS_TOKEN},
    ...                   data=json.dumps(data),
    ...                   headers=headers)
>>> r.status_code
200
```
Uploading different types of data

Search for different data sets

Create communities and upload to them

Export Metadata

Create additional scripts
<table>
<thead>
<tr>
<th>Guides</th>
<th>Administrator Guide</th>
<th>General User Guide</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Installation</td>
<td>Account Creation</td>
</tr>
<tr>
<td></td>
<td>Configuration</td>
<td>Search</td>
</tr>
<tr>
<td></td>
<td>API Information</td>
<td>Upload</td>
</tr>
<tr>
<td></td>
<td>Adding Scripts</td>
<td>Communities</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Metadata Export</td>
</tr>
</tbody>
</table>
Demo

Local Zenodo Installation

Upload a dataset
  Basic information
  Licensing information
Search for a dataset
Exporting metadata

Automated Python Scripts

Single upload
Batch upload
Publishing option
Lessons Learned

• Planning Effectively for a Large Technical Project
  • Short term
  • Long term
  • Unforeseen issues
    • Port issues
• The Importance of Communication
  • Long distance communication
  • Dynamic project goals in accordance with client’s needs
• Attention to Detail
Acknowledgements

Dr. Edward A. Fox
Instructor: CS4624 Multimedia, Hypertext, & Information Access
Professor, Department of Computer Science
Virginia Tech

Dr. Seungwon Yang
Client
Assistant Professor, School of Library and Information Science and Center for Computer Computation and Technology
Louisiana State University


