**Speed Blog 1—April 14, 2017**

At this point, our group, Data in Recycling, is testing the various information sources as we develop our idea further, with the intent of drilling down and distilling our project into a manageable form. One major step we have taken is to search re3data.org for relevant

repositories. By “relevant,” we mean similar repositories that include potentially useful protocol and curation and presentation policies. Some potential datasets identified include:

The Comparative Agendas Project (<http://www.policyagendas.org/>)

The History Data Service (tp://hds.essex.ac.uk/)

Environment and Climate Change Canada (Water Data) (<http://www.nationalarchives.gov.uk/webarchive/archiving-datasets.htm>)

We identified these repositories as useful due to their presentation methods, subject matter or curation policies. We intend to each write a paragraph about one repository, outlining its strengths and weaknesses, and then meet to discuss our results and decide what elements do and do not serve our quarter project.

Regarding the in-class evaluation of the two data repositories, we took an overview of the repositories themselves, judging their aesthetic quality, transparency and usability, and read up on the repositories’ background. Our results are below:

***CKAN***

* Language of codebase: written in Python, available in other languages
* Current Release: 2.6.0 (November 2016), 2.7.0a? (found mentioned online, but not on website documentation)
* Current release under use: 2.3?
* API availability: yes
* Type of Data accepted: governmental, private or public, cartographic
* Users: national and local governments for a variety of countries, and paying organizations and communities

**Major Strengths:**

* Clean, intuitive design—this is a good interface
* Demo: easy to use
* Lots of information about how to use the repository
* Lots of analysis tools

**Major weaknesses:**

* Website doesn’t really say what CKAN *does*—what their focus is
* How does the tagging structure work?
* Somewhat chaotic; easy to get lost. How does one get started?

**Potential path to implementation:**

Search for related datasets for information. Upload our protocol for dissemination.

***Dataverse***

* Language of codebase: Java, SQL
* Current Release: 4.6.1
* Current release under use: unclear; possibly 4.6.1
* API availability: multiple open APIs, including search and data deposit
* Type of Data accepted: academic and governmental data
* Users: journals, governments, researchers, institutions

**Major Strengths:**

* Good, intuitive design—not too overwhelming
* Seems to be a strong community, with help easily to hand
* Mission and services explained simply

**Major weaknesses:**

* Information about software presented in bewildering GitHub format
* Not all the links seem to work
* How do you search the datasets?
* Some datasets seem difficult to reach; it’s hard to know how to exactly get hold of a dataset

**Potential path to implementation:**

Search for relevant information. Upload protocol for dissemination.