Introduction

Copy whatever from proposal
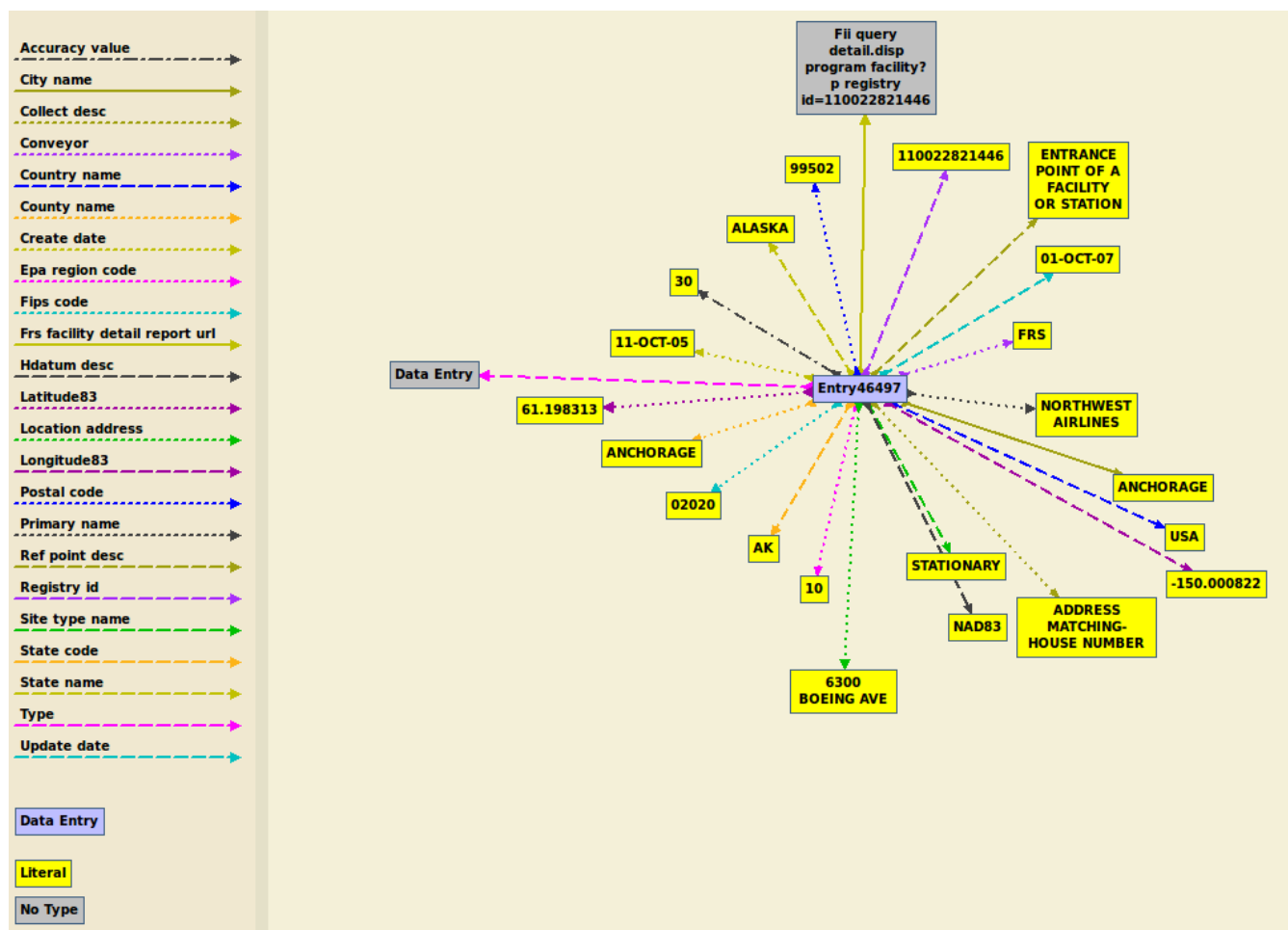
Target Audience

Copy whatever from proposal

Description of Data Sources

The datasets we use are Datasets 1022 through 1080 from Data-gov. These datasets represent the data collected by the Environmental Protection Agency Facility Registry Service for each of the 59 states and territories of the United States.

Although the data is given in n-triple format, it's not really in a linked format, but rather CSV-style format, and in fact is the output of csv2rdf on several CSV files which were obtained from the EPA-FRS.

Each DataEntry class instance is a row in the original CSV, and corresponds to the collected data about one facility, where each property is a literal from each field in the row.

Each row may provide information about a site's location, at different granularity levels, such as state, or street address, or even longitude and latitude; about the type of site; and dates at which the row was created and updated. Not all information is present for every entry in the data set.

Data Integration

We integrate the data sets into a single one using a bunch of stuff.

If a facility entry is missing a state, it's assumed to be located in the state for that data set.

If a facility entry is missing a type <http://data-gov.tw.rpi.edu/vocab/p/996/interest_type>, it is given "unspecified" type.

Data Product Results

We made some map web page with google visualization api.

Summary