

uments. Publishing Excel spreadsheets is the most typical example of a 2-★ data-publishing step.

★ ★ ★ *use an open format.* This means avoiding propriety formats such as Excel, and instead using open formats like .csv file or Open Office format. Typically simple format-transformation steps might suffice to get from a 2-★ to a 3-★ rating.

★ ★ ★ ★ *give a URL for each data item.* This is the first step where web semantics is making its appearance: use a URI to denote every data-type, data-item, and data-property. This allows others to link to the published dataset.

★ ★ ★ ★ ★ *Finally, link out to shared vocabularies.* For example, use the DBPedia term for city when talking about the class of all cities, and similar terms for names of entities and properties. This use of external vocabularies truly enables the interlinking of datasets across different sites.

6.4.2 Adoption

Although much of the data on <http://data.gov> has been (and still is) only in 2-★ format, it turned out not to be hard to move this data up in the above ranking system. In a period of eight months, students at the Rensselaer Polytechnic Institute managed to “web-enable” 6.4 billion data-items from <http://data.gov>, and (more significantly), they created applications for such diverse projects as plotting postal service expenditure against performance, plotting wilderness fires against agency budgets (to measure effectiveness), covering interstate migration from information on tax forms, plotting family income against Medicare claims, comparing agency budgets across three public budget datasets, and plotting social networks of people visiting the White House. Although impossible to establish a causal link, the number of citizen appeals under the Freedom of Information Act have reportedly dropped substantially since the launch of <http://data.gov>.

Encouraged by this success, other governments have started to follow this exam-