# Health Reporter

*Final Project Report*

After basic demographic data, health data is probably the most important collection of information for informing public policy, but it is still difficult to use. Inspired by Census Reporter, a tool that presents Census data in profiles that are tailored to journalists' needs, the Health Reporter project intends to develop a similar application to make health data easier for journalists to find and use.

The California Health Care Foundation funded this project with a goal of modifying Census Reporter to be able load health data, so we could explore how to improve journalists' access to a critical resource. This report summarizes the project and presents an overview of a proposal for the next stage.

The initial intent of the project was to start with the Census Reporter code base, but we quickly discovered that it was most suited data for the American Community Survey, and modifying it to work with the very wide variety of health data would be difficult. Fortunately, Civic Knowledge had an existing Open Source project, Ambry, which provided a suitable codebase. Using the Ambry Data Library, we imported a large, diverse health and social dataset produced by the California Department of Public Health, the Healthy Communities Indicators. Simultaneously, we interviewed journalists to better understand their data needs.

With the HCI dataset, we extended the Ambry Data Library with automatic visualizations, testing the application with another set of journalists. The outputs for this projects include:

- An analysis of the user's needs, *Providing Data For Data Journalism.*
- An overview of the interim technology demo.
- A demonstration application.
- The user interface design.
- This final report

The primary lessons of this project are:

- Journalists have four ways of using data in stories, and a tool like Health Reporter can address three of them.
- An indicator site with enough data to be useful will require innovative interfaces for search and browsing.
- Managing an indicator site with enough data to be useful would be expensive without efficient data management.

These three points are discussed in detail in the *Providing Data For Data Journalism* report, but summarized here.

*The Health Reporter project was funded with support from the California Health Care Foundation.*

California Health Care Foundation

## Informing, Supporting and Core Stories

Of the four ways Journalists use data, three are relevant to the Health Reporter application. For Informing stories, journalists are looking for a fact, often a single number, which Health Reporter can provide with profiles, similar to Census Reporter profiles. For Supporting and Core stories, the journalists is looking for a dataset, not a fact, so Health Reporter must also provide access to raw data.

When looking for data, the journalist either has a specific idea about what the data should be, such as “diabetes rates for men in Los Angeles” or has a more vague notion, such as “how are graduation rates changing over time?” In the first case, the best way to find data is a structured text search, like Google, but tailored for data. For the second case, the best model is a visual search, but because of the large amount of data the system must have, it requires a visual search that is not mentally taxing, and a technique called pre-attentive processing can make visual search much more efficient and accurate.

These lessons imply that future projects should develop a data library application with two linked search systems, one that uses text search, and another that uses pre-attentive processing. The text search system has already been implemented in Ambry, based on the results from a previous project exploring the use of metadata, so a future project need only explore pre-attentive visualizations.

## Cost-Effective Data Management

To be useful, Health Reporter will require a large database with a lot of datasets, and managing data is expensive. After initial development, we estimate that 80% of the cost of running Health Reporter would be data management. To make a system like Health Reporter sustainable, we have to tackle the cost.

To solve the cost problem, we’ve developed two solutions. For most datasets, the Ambry application can be used to create data packages in a concise, repeatable, distributed way, allowing us to reduce costs both through efficiency and using off-shore labor, students and interns. But, it would be even less expensive to have data producers create good data packages themselves, so we have also created a specification for metadata that anyone can use with little training and no additional software tools. If some of the datasets Health Reporter needs are produced with directly importable data packages, the data management for those datasets is zero, leaving more resources for our own staff to package other datasets.

## Next Projects

To capitalize on the first successful stage of this project, we propose two additional projects:

- **Health Reporter, Round 2**: Implement the pre-attentive exploration and refine text search in the Ambry Library, realizing the user interface described in the Health Reporter project.

- **Data Management**: Complete the Ambry Data Packaging design and test the process by training data producers to create data packages. The data packages created in this project will be loaded into the Health Reporter application.

As continuations of the Health Reporter project, these two projects have a substantial opportunity for social impact, far beyond the intersection of health and journalists. Our user interface research suggest significant improvements are possible in building interfaces tailored for data search and exploration, and we'll be working with an expert researcher in visual processing and perception to complete the user interface model. Our Open Source data management software can cut data management expense in half, and widespread adoption of our data packaging standards among data producers could make the data mangement of indicator websites virtually free. Combining these projects, with Ambry data packages fed into the Health Reporter search and visualization application, we can make California's, and the Nation's, health data easier to find and use.