ScraperWiki

PROJECT	GRANTEE	INNOVATION	GRANT
ScraperWiki	ScraperWiki	New, journalist-specific features within an existing tool to collect, store, and publish data from across the web	**$280,000**

As more institutions make data available online, the potential to increase data-driven journalism grows immensely. Unfortunately, accessing and processing data is often tremendously difficult, particularly for journalists and citizens with limited programming skills. Few tools exist to help journalists find, access, and process obscure or complex data. ScraperWiki, a start-up in Liverpool, England, sought to develop journalist-specific features to facilitate the collection of information from across the web.

THE INNOVATION

ScraperWiki allows users to collect, store, and publish public data, with a tool called a scraper. The data they scrape is then made available for others' use. ScraperWiki was awarded News Challenge funding to add a data on demand feature that is specifically tailored for journalists, and to accelerate adoption of the platform in the United States by hosting journalism data camps. The new feature would allow journalists to request data sets and be notified of changes in data that might be newsworthy, and it would allow them to place data embargos that keep information private until after their stories break. ScraperWiki ultimately aims to allow journalists to produce richer stories and data visualizations by providing them with the means for accessing updated, aggregated public data.

IMPLEMENTATION

The first goal of ScraperWiki's News Challenge award was to add new, journalist-specific features to its existing platform. These features included a tool for on-demand, rapid-response data scraping; custom, private scrapers for a fee; assistance in developing public data sets; and a data alert service that notifies journalists about changes in relevant data sets. ScraperWiki released its first iterations of these features in late 2011, and it used customer feedback and A/B testing to drive ongoing development.
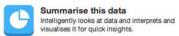
The second goal of ScraperWiki's News Challenge grant was to accelerate the adoption and use of ScraperWiki throughout the United States by hosting journalism data camps in New York, NY; St. Louis, MO; Washington, DC; San Mateo, TX; and San Francisco, CA. ScraperWiki kicked off its program of events in January 2012. To conduct these data camps, ScraperWiki partnered with newspapers and organizations such as *The New York Times*, the *Chicago Herald Tribune*, the Sunlight Foundation, ProPublica, Investigative Reporters and Editors, Spot.Us, the Centre for Investigative Journalism, Code for America, and the Online News Association. It also received support from journalism schools in its targeted states. During these events, journalists learned more about ScraperWiki and coding, and programmers learned more about ScraperWiki and data scraping.

A number of challenges impeded ScraperWiki's ability to gain newsroom adoption. First, the project team found that newsrooms often had very little money and were unwilling to pay for data services. The team also encountered confusion about the platform's free versus paid services. In addition, ScraperWiki found that its levels of newsroom adoption were low because its platform required a level of programming expertise and technical skill that few journalists and media professionals held. Often, journalists with coding experience also had their own internal tools for data mining and scraping. In response, the team decided to develop a new, more user-friendly platform that would address

Super-charge your data analysis with tools including…

Search for tweets
The simplest way to collect tweets matching a search term over time.

Upload a spreadsheet
Transforms and extends your data beyond the limitations of a spreadsheet.

Code in your browser
Write scripts that transform data or scrape the web, in any language you like.

Summarise this data
Intelligently looks at data and interprets and visualises it for quick insights.

Query with SQL
Get under the hood and do powerful analysis of your data. Then deploy via our JSON API.

View on a map
Get a simple, effective, and zero-configuration bird's eye view of your data.

Get a free account to try them all out

these challenges and also allow more technical users to code within the browser application. It also introduced a community plan option, which allows users to get, clean, visualize, and analyze up to three data sets for free and to upgrade to a premium account if they agree to publish their data to Datahub.io, the Open Knowledge Foundation's CKAN government catalogue.

ScraperWiki released the beta version of its new platform in July 2013 and closed its original system. For journalists, the new ScraperWiki provides built-in data tools for specific tasks, such as scraping and searching for tweets, uploading spreadsheets, and summarizing large volumes of data through data visualizations. The new platform also allows more technical users to write their own code using GitHub, SSH, or the programming language of their choice.

ScraperWiki continues to experience challenges in selling its platform and services to news organizations, where data would be used to support development of editorial content to create deeper, more compelling experiences for news consumers. It has experienced more success, however, in selling the tool to government agencies and corporate media clients. In this context, *The Guardian*, Channel 4 News, the UK Cabinet office, the United Nations Office for the Coordination of Humanitarian Affairs, Informa Agra, TraderMedia, and others have purchased subscriptions to the ScraperWiki platform.

REACH AND OUTCOMES

ScraperWiki succeeded in building versions of its software for use by journalists and news organizations. However, it has struggled to gain traction with journalists because few newsrooms are willing to pay for data services and because use of the platform previously required significant technical ability. The release of the new, more user-friendly system in July 2013 and the introduction of a free community plan subscription model were attempts to address this challenge and to make ScraperWiki more useful for investigative journalists. As of February 2014, only about 200 data journalists have registered for the tool. ScraperWiki as a whole, however, averages around 3,000 new registered users a month, many of whom are corporate media clients and government agencies.

By March 2014, journalists from *The Guardian*, the *Chicago Tribune*, and the *Texas Tribune* had used ScraperWiki to produce stories. *The Guardian* used ScraperWiki in a front-page story revealing corporations and interest groups that channeled money to buy influence among UK's parliament. Reporters used ScraperWiki to collect data located on different services from registers across parliament, and to aggregate it into one source table which could be viewed in an automatically updated spreadsheet or document. ScraperWiki has yet to achieve similar levels of traction within the United States, but it has succeeded in sparking connections and collaboration between over 500 journalists and developers through its data journalism camps. As the result of a connection made during ScraperWiki's data journalism camp, two attendees Brian Ableson and Michael Keller have gone on to collaborate on various open news projects. The pair produced an interactive news app published in the Daily

Beast displaying the size and locations of mass-shootings within the United States.

Today, ScraperWiki continues to promote its platform to both journalists and corporate communications clients, relying on its consulting work and managed services in other industries to help sustain the development of its tools for investigative journalism. ScraperWiki continues to market the services of its media tool pack to journalists, and it plans to conduct additional market testing throughout 2014.