The Colorado Data Commons   9/19/2011

EXECUTIVE SUMMARY

Overview
The Colorado Data Commons will unlock and untangle key socioeconomic datasets currently locked within public agency databases or in hard-to-use formats, providing one-stop access for journalists, policy-makers and citizens working to understand their communities. Public information is too often inaccessible to people who need it. The Data Commons would change that. More than a simple data repository, the Data Commons will be an actionable tool to unlock data and put it to work in communities, connecting datasets with people who can use them to affect change.

These examples highlight the urgent need and the effect the Data Commons will have by unlocking public information:

> *After lawmakers exempted some schools from regulation, community organizer Patty Lawless tried to find out how those schools performed. The data she needed were public, but inaccessible. Patty had to get printouts of each school's performance and retype them into a spreadsheet.*

> *Reporters at the Denver Post wanted to examine pay in Jefferson County, where the treasurer was accused of attempting a kickback. An official later told I-News' Burt Hubbard they put County financials on-line in a format designed intentionally to stymie reporters. Burt helped the Post discover that bonuses ballooned 700%.*

> *Hunger Free Colorado knew the state's summer food program reached a fraction of the kids who needed it. The Piton Foundation overlaid poverty data with food program locations, and revealed gaps in service. By making public information usable, 200,000 more meals reached hungry kids in Colorado.*

The Data Commons would solve information needs such as these and have a measurable impact on the community by removing barriers, bridging gaps, and surfacing new information needs.

Anticipated Outcomes
Our community – and eventually others, too – will have access to more useful news and information. Communities will have a way to voice their information needs, an infrastructure to help meet those needs, and a network to inspire community engagement and action.
The Data Commons will "democratize data" by increasing access and creating an environment in which to build public data literacy -- strengthening Colorado's information ecosystem.
Residents and organizations working on community issues will have access to data on community health, which will allow them to make better decisions about how to deploy resources to have maximum impact on the issues they care about most.
By incorporating the Data Commons in an ongoing project called Vital Communities, we will demonstrate how residents can immediately use data to solve problems. Students at a local high school suspect punishment is linked to race, which causes turmoil and affects learning. Parents, students, and staff hope a restorative justice program will help. Vital Communities will draw from the Data Commons to track student suspensions and expulsions, and determine whether student achievement increases as a result of their work.

Measuring Success
The social value and impact of the Data Commons may be best measured by identifying the ways in which it affects the work of organizations and residents in the community, as they seek to address issues such as hunger, poverty, education, crime and health. By improving their capabilities, the Data Commons will have a demonstrable impact on these social issues.

Our approach to creating and measuring impact will be based on the situated logic model approach, which is a way of building an impact model that clearly establishes links between use of the Data Commons and community change efforts. During the discovery process we'll ask stakeholders how they plan to accomplish their goals, and help them describe it through a logic model or theory of change. Then we'll help them 1) determine how the Data Commons can advance their work, and 2) develop a concrete description of the various types of impact and how the Data Commons contributed to this impact.

Partners

This project builds on The Denver Foundation's strong relationships with these potential partners:

I-News: A statewide nonprofit news service that received a KCIC grant last year and produces content for more than 20 media partners, I-News will develop stories using the Data Commons and training for public use of the data.

The Piton Foundation: An operating foundation that works to strengthen low-income communities, Piton provides a data bank of information on poverty in Metro Denver, and has been working for 20 years to provide information to residents about their communities. They would house and staff the Data Commons.

The Civic Canopy: A network of partners working to build strong communities through deep civic engagement, Civic Canopy will facilitate groups of residents who are using the Data Commons.

Sustainability

Because the Data Commons will be housed at the Piton Foundation, which already invests in an ongoing Data Initiative, once the Commons is built it will have some of the infrastructure needed for sustainability.

We will be exploring earned income revenue streams such as licensing, custom platform development, and custom information analysis. For example, a foundation seeking to make a difference in early childhood education could track where their investments have gone in the past decade as related to third-grade reading levels in the adjacent school districts. If they don't have the staff to do this analysis, they could hire Data Commons staff to do so.

Requirements

The project will unfold in four stages:

Planning and Discovery
First, we will surface information needs of the local community. We will engage citizens, foundations and organizations, journalists, civic web developers, data providers and government agencies and conduct usability studies on how these groups might interact with the Data Commons. Specific pilot projects such as I-News, Vital Communities, and the Children's Corridor Atlas will help describe key features and requirements. We will identify solutions to the barriers that keep government agencies from making their data more accessible. We will build a strong data taxonomy based on themes such as health, education, economic development and environment. We will also explore the Data Commons' potential as a platform for crowd-sourced research and journalism.

Scoping and Building
Following our user needs study assessment, we will identify key Data Commons features and contract skilled web developers to build them <u>or leverage an existing platform</u>. Important features include: a public data repository allowing for metadata capture, linking datasets, and multi-year comparisons; tools that remove barriers to information dissemination, making it easy for data providers to address privacy concerns through tiered access and geographic aggregation; and static and dynamic data visualization tools integrated with social media such as Facebook and Twitter, allowing users to share their data analyses.

We will build an API that will allow the development of dynamic applications. Potential applications include: a customizable data dashboard for community groups to visualize and track change indicators related to their focus issues; a community mapping tool for users to view the local assets and data for customized geographies; and a portal for public agencies to easily upload and share data with community groups.

Dataset Gathering
Prior to the releasing the Data Commons, we will organize an intensive dataset collection and cleaning effort. Building on the catalog Piton's Data Initiative already curates, we will deepen our store of information from local sources and yield new datasets to combine and use in unprecedented ways.

Public Roll-out
This project will bring previously inaccessible data to the public through platforms that extend the Data Commons; through ad hoc use in skilled media reports by partners such as I-News; and through direct citizen access and exploration. Roll-out activities will include data training targeted by skill-level from novices to programmers and the development of "power" users who can become field-trainers helping their communities engage with the Data Commons.

Core Elements

Element 1 - Data Repository - A web-based warehouse for social data sets that is searchable and

supports importing and exporting of data sets. Sub-requirements include:

a. Data standard for multiple data types (quantitative and qualitative, media, geospatial)
    a. A standard format (e.g. JSON, XML) for data and meta data
    b. Format conversion
       (for example see: Mr. Data Converter http://shancarter.com/data_converter/)

b. Data search and taxonomy
    a. By geography
       i. Custom geographies
          (for example, see: WSJ Census Map Maker http://censusmapmaker.com/, Wellbeing Toronto http://map.toronto.ca/wellbeing/)
       ii. By topic/theme
          (for example, see BuzzData http://buzzdata.com/, ManyEyes http://www-958.ibm.com/)
       iii. By year

c. Data uploading
    a. In multiple formats (from Excel, CSV, Google Docs, etc.)
    b. Spatial dataset uploading (shapefiles, kml, etc.)
    c. By administrators and general users

d. Data export/download
    a. In multiple formats (Excel, CSV, KML, etc.)
    b. Customizable download (indicators, extent, timeframe, etc.)

Element 2 – Tools for Data Visualization and Analytics – In addition to a portal to access data, the Commons will have features that allow data-savvy users to interact with data sets through visualizing them. These features will be lightweight as we anticipate most Data Commons users will have their own preferred visualization techniques [*Note: This is something we need to validate during our usability studies.]*

a. Mapping/GIS
b. Graphics/analytics – bar charts, line graphs, pie charts, bubble charts, etc.
c. Data dashboard with auto updates when new data is available
d. Connectivity/use of existing data visualization platforms? (Tableau Public, Google Docs, Google Fusion Tables, etc.)

[Note: We are interesting in incorporating existing open source visualization technologies. We do not want to compete or reinvent the wheel. The goal of the Commons is data sharing and collaboration.]

Element 3 – Sharing of Data Sets, Building a Community of Data Users – The power of the Data Commons lies in the ability of data users to connect with each other. People and agencies supplying data sets can see how their data is being used, and people using data can see what others are doing with it. The Data Commons will encourage users to sharing data sets and data work.

a. Social media integration
    a. Export/share data sets and visualizations to social networks
    b. Export data sets and visualizations as reports, PDFs, sharable and printable formats
    c. Embed data sets and visualizations (i.e. produce embed code like YouTube)
    d. Multi-lingual considerations
b. Forum/data community tools
    a. Discussion boards
    b. Data requests/wish-lists

c. Best practices and data tutorials
d. Internal sharing – view what other users have done with data sets in the commons; "close the loop" by linking data users with data set providers

Element 4 – Open API – The Data Commons will support an open API for web and app developers to build extensions from available data sets. (see Civic Apps Portland http://www.civicapps.org/)

a. Extensible and open source
b. Clear documentation and support for developers (see jQuery documentation and tutorials http://docs.jquery.com/)