

Memorandum



TO: Citizens of Santa Fe, NM
FROM: Tom Johnson tom@jtjohnson.com
**SUBJECT: Draft of It's The People's Data initiative
position paper**
DATE: May 2015

- A video introduction to the project is at <http://bit.ly/1eM5Vof>
- Download the It's The People's Data mindmap [here](#)
- If you are interested in participating in the project, please click [here](#).

Background

A handful of scholars and concerned citizens gathered informally in Sebastopol, Calif. in December 2007 to consider the state and role of data collected by governments in the United States. The result of that session was "...a set of eight open-government data principles that inaugurated a new era of democratic innovation and economic opportunity."¹

"The objective...was to find a simple way to express values that a bunch of us think are pretty common, and these are values about how the government could make its data available in a way that enables a wider range of people to help make the government function better," Harvard Law School Professor Larry Lessig said. "That means more transparency in what the government is doing and more opportunity for people to leverage government data to produce insights or other great business models."²

Access to **The People's Data** is the feed stock for democratic participation, economic planning, innovation, growth, philanthropy and greater efficiencies in government at all levels.³ Open Data engenders greater confidence in government, a higher level and

¹ <http://www.paristechreview.com/2013/03/29/brief-history-open-data/> and <http://fcw.com/Articles/2014/06/09/Exec-Tech-brief-history-of-open-data.aspx?Page=2>

² <http://fcw.com/articles/2014/06/09/exec-tech-brief-history-of-open-data.aspx>

³ See [Insight, inspiration, and intel: what governmental open data offers for entrepreneurs of all kinds](#)

generation of analytic journalism and fact-based decision-making on the part of government, industry and commerce and social services.

We are in the initial nanoseconds of the Digital Age, only the fourth phase in the 2-million-year history of humanity where new tools and methods of communication are driving changes of unprecedented scope, speed and cultural, intellectual and societal impact. Adjusting to these dynamic and technological changes is, for some, difficult; for others they are fascinating and challenging, offering new possibilities in all aspects of life. But implementing the new technologies is easy compared to the cultural changes necessary if we are to flourish in the Digital Age.

While the City of Santa Fe has and is making efforts to [post some](#) of The People's Data, those efforts to date fall short of the progress being made in other cities.⁴ Citizens still labor under the weight of personal visits, with paper in hand, to the downtown office of the city's attorney office [to request public records](#).⁵ If the desired data can be found, citizens are charged not-insignificant fees to retrieve and copy that data.⁶ Some of the desired data may, however, be downloaded from the city's web site. But in many (most?) cases, those documents/data exist as PDF files. As such, it requires much effort, time and money to extract both text and quantitative data from those files to conduct any analysis.⁷ Therefore, citizens are being taxed twice for The People's Data: initially to collect, file and store the data, but then again when the citizen has to extract the data from files in order to analyze it. The solution? Post all data on the city's site in its original file format and create links for direct downloads, "direct" without having to file Inspection of Public Records requests.

It's The People's Data initiative is intended to make sense of the changing information environment and enlist citizens in a community effort to identify and retrieve the data they have paid local government to collect, to learn new analytic skills to tease information out of that data for a myriad of purposes and then to communicate the results of that analysis to a variety of audiences in the community.

We anticipate working with organizations such as the New Mexico Foundation for Open Government, the League of Women Voters, the Republican and Democratic

⁴ See Kerrigan, Heather. "8 of the Best City Financial Transparency Sites." Jan 1, 2015 <https://www.govloop.com/community/blog/best-city-financial-transparency/>

⁵ As of 3 June 2015, the link on the City Attorney's Transparency page to [New Mexico Inspection of Public Records Compliance Guide](#) was not functional.

⁶ "The Public Records Policy includes a [Schedule of Fees](#)."

⁷ The City of Santa Fe's FY 2014/15 budget, for example, is only available as a PDF file. Yet it is obvious that the tables in that document were originally prepared as an Excel or some other spreadsheet file. Consequently, great effort must be made to extract that budget data to facilitate any analysis. <http://bit.ly/1bMPCoN> An advanced Google search indicates that the city has posted no .xls files online.

parties, neighborhood and homeowner's associations, various professional and economic groups and the Society of Professional Journalists.

Objectives

- Make the data collected by the government easily and quickly available to taxpayers in its original and most fine-grained format.
- Change the culture of public employees and politicians so they will embrace the idea that making the people's data available is a primary mission and responsibility of government at all levels.⁸ E.g. Change job title from "Public Records Custodian" to "Public Records Provider."
- Enhance the community service mission of journalism institutions.
- Teach citizens (and government employees) how to identify, retrieve, clean, analyze and present data and the information it holds.
- Maximize the openness of the initiative to make participation easily accessible, to show process and progress and foster broad collaboration.
- Make Santa Fe the first small city to gain national and international recognition as championing It's The People's Data
- Produce a print and online "cookbook" of the process used to create an **It's The People's Data** initiative. This book will be distributed to local media in the U.S. and internationally.
- Build a easily accessible meta database of the city's databases.

Strategies

- Use the audience reach of the local journalism organizations (newspapers, broadcast, professional organizations) to publicize the existence of the **It's The People's Data** initiative.
- Recruit interested participants from the broadest possible spectrum of political, commercial, non-profit organizations and government employees in the city and county.
- Create a sophisticated online site to manage the initiative calendar, training modules, participants' profiles, demonstrate the openness of the project, and attract suggestions from non-local sources. This will also demonstrate our validity to funding individuals and agencies.

⁸ As of June 3, 2015 the City of Santa Fe "Request Public Records" [web page](#) has a dead link to the NM IPR Compliance Guide, and the "public records custodian" does not permit electronic submission of a public records request.

- Recruit experienced teachers of data analysis⁹ and create 12-18 training sessions focusing on:
 - Navigating the initiative task management web site. (Possibly Bitrix24.com)
 - Theory of the Data In → Analysis → Information Out process
 - Database structure, file types and retrieval, data cleaning, basic analytic tools, information presentation and data visualization, online presentation, social media technology as promotion and presentation tools
- Build contact database of local organizations and individuals who could be interested in initiative
- Meeting in the journalism organization's offices bi-weekly for progress reports and training sessions.¹⁰

Tactics

● Phase I

- Recruit participants with regular house ads in print and broadcast announcing the initiative, initial meetings and posting web site
- Make presentations 4-6 weeks before initial meeting to local organizations¹¹ about the initiative, its objectives and seek collaboration. Gather individual contact information with post cards
- Promote online form for interested persons to indicate contact information. See <http://bit.ly/1dUk6a8>

● Phase II

- Select one or two cooperative city departments to use as proof-of-concept models for conducting a database/data census
 - Fine-tune methodology, e.g. questionnaire/recording sheets, best time to interview; data entry process
 - Train data census takers
- Conduct a census of data collected by city and county government agencies. The census will describe in some detail:
 - the form of the data (ink-on-paper or digital),
 - the type of data (text, quantitative, geographic, images, audio, video, etc.)
 - database schema or code sheets (i.e. variables, past and present definitions)

⁹ e.g. Journalists, statisticians, real estate appraisers, financial analysts, investment advisors, forensic auditors, economists and other social scientists, computer scientists, GIS professionals, graphic and user interface designers, social media professionals.

¹⁰ These sessions will not be gratis. There shall be a moderate fee charged, just enough to let participants know that we expect their regular and dedicated participation, but not so much that would discourage participation.

¹¹ Include political, elected officials, religious, non-profits, environmental, social services, commercial, professional, education, neighborhood and homeowners,

- of the fields); field type (limited/delimited/text/alpha-numeric)
 - genealogy of the data (who or what called for the collection of the data; who and how was and is it collected; who and how is the data verified/validated; who can edit the data; who has access to the data; how is the data used and by whom)
 - Location of the data (physical and online)
 - Contact person in government to request access and download
- Parallel with data census, conduct training sessions focusing on:
 - How to use team management sites
 - Bitrix24.com
 - Xmind.net
 - [Google Drive Documents](#)
 - Fundamentals of file types
 - Fundamentals of databases
 - Data retrieval and storage
- **Phase III**
 - **Extensive review and assessment of progress and method**
 - How to improve online presence?
 - What are we missing?
 - What are the roadblocks?
 - Could the process be improved?
 - What can we do to assist the public employees with their tasks?
 - Are we reaching the community?
 - Can the project be scaled up to the county level?
 - **Training**
 - Data cleaning (Excel Basics; Open Refine Intermediate, etc.)
 - Excel Basics: Sorting; hidden columns; index column; beginner mistakes; linking worksheets;
 - Basics of data visualization;
 - Basics of mapping with Esri plug-in; creating maps with other tools
 - **Publication**
 - Begin writing and design of cookbook site and manual

Potential Participants

- Journalism Organizations
- FOI funders
- Local foundations and related non-profits

- Santa Fe Community Foundation
- Political action groups
- Environment interests groups
- Labor organizations
- Chamber of Commerce

Metrics and Milestones

- **Phase I**

- Recruitment:
 - TK¹² presentations to community groups
 - TK print and TK broadcast announcements
 - TK conversations on Mary-Charlotte's program on KSFR
 - 10 committed individuals representing TK organizations
- Training
 - Lesson plans (Show, Do, Teach) for initial 12 weeks. PowerPoint, handouts, exercises
- Prototype of databases of databases web site. (Socrata?)
- Prototype of It's The People's Data website
- Financial
 - Pledges of ~\$25k in kind or cash
 - Agreement with [New Mexico Community Foundation](#) to handle accounting
- **End of Phase I: Go-No-Go decision**

Staffing (Start small; expand as needed)

- Project Director:

Responsible for executing and updating 12-month project plan. Reports regularly to advisory board. Coordinates promotion of initiative and recruitment of participants. Schedules meetings, training and instructors. Oversees data census and creation of our database of city/county databases. Maintains collaborative task management tool, i.e. Bitrix24 or other. Manages financial accounts and records.¹³

Desired applicant: Some experience in administrating analytic/data journalism projects or news research or database design and management.

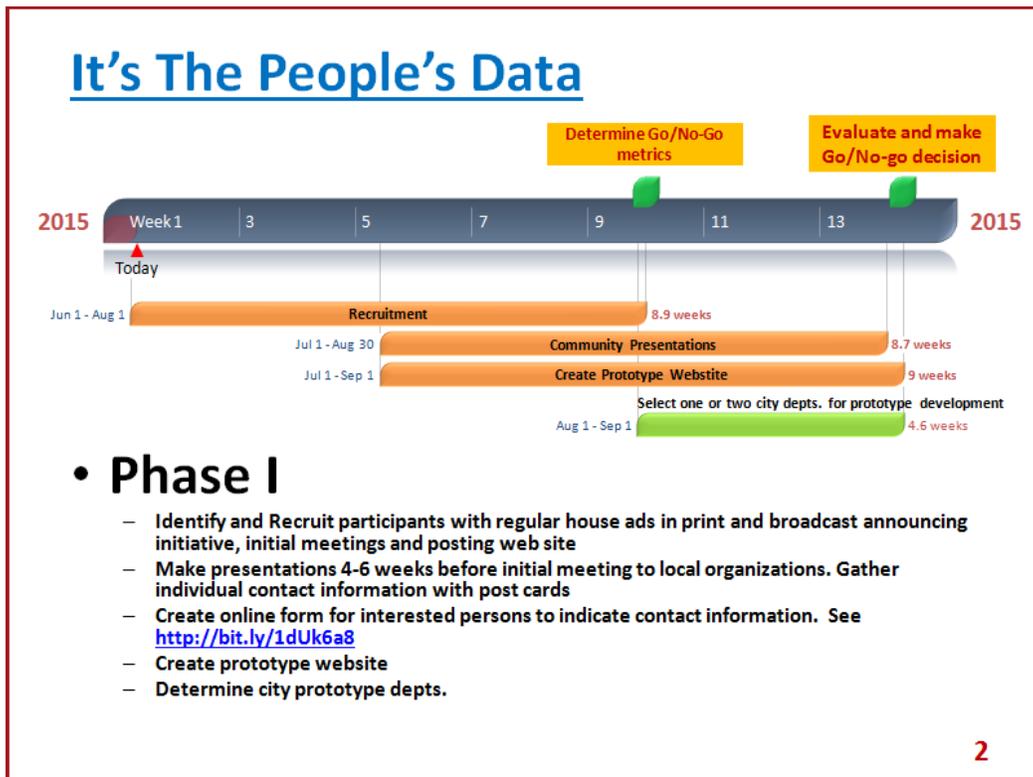
Initially half-time position. Hourly or salary?

¹² "TK" is journalese for "to come."

¹³ In conformance with the project's objective to promote transparency, effort will be made to livestream all meetings and to post all financial records and activities on the project website.

- Database designer/consultant:
Responsible for either design and coding of project's database of databases OR liaison with vendor(s) such as Socrata or Tableau. Hourly?
Desired applicant: Ideally, news research/librarian experience. Online portfolio of database creation and administration. Rich knowledge of data-cleaning and validating techniques, GIS tools, SQL and probably Python and Javascript.
- Website Designer:
Works with database designer and marketing/social media coordinator to create and maintain uniform look-and-feel of user interface. Hourly.
- Marketing/social media director
Coordinates broad multi-media publicity efforts to (a) recruit participants; (b) promote training sessions; (c) assist with fundraising, development. Hourly.

Draft Timeline



It's The People's Data



- **Phase II – Capture and analyze**

- the form of the data (ink-on-paper or digital),
- the type of data (text, quantitative, geographic, images, audio, video, etc.)
- database schema or code sheets (i.e. variables, past and present definitions of the fields); field type (limited/delimited/text/alpha-numeric)
- genealogy of the data (who or what called for the collection of the data; who and how was and is it collected; who and how is the data verified/validated; who can edit the data; who has access to the data; how is the data used and by whom)
- Location of the data (physical and online)
- Contact person in government to request access and download

- **Train**

- Data extraction; cleaning, basic analysis, basic mapping

3

It's The People's Data



- **Phase III – Extensive review, expanded data collection, publishing**

4