CITY OF OAKLAND

AGENDA REPORT

TO: DEANNA J. SANTANA
CITY ADMINISTRATOR

FROM: Scott P. Johnson

SUBJECT: Open Data System Implementation Report

DATE: June 20, 2012

City Administrator Approval

Date 6/28/12

COUNCIL DISTRICT: City-Wide

RECOMMENDATION

Staff recommends that the Oakland City Council accept this informational report from the Office of the City Administrator on the cost and benefits of the implementation of an Open Data System and the steps necessary to implement an Open Data System.

EXECUTIVE SUMMARY

Online open government initiatives—sometimes referred to as "Gov 2.0"—are being launched by many national, state-level and local governments worldwide. Driving most online open government strategies is Open Data. The idea behind Open Data is that information collected and held by government should be freely available to use and re-mix by the public.

By opening our data and becoming a more transparent and participatory City we begin to:

- Promote Civic Engagement
- Improve Decision and Policy Making
- Increase Government Transparency and Improve Access to Public Information

For fairly low overhead, cities all over the country, who have adopted Open Data policies are driving innovation around tools that address their communities wants and needs. Cities that have embraced open data have seen software applications, quality research and action-oriented ideas emerge from the community that address issues around finance, parking, transportation, entertainment, public health and sustainability. They have also seen more public trust in municipal government and increased civic participation and collaboration.

Like most government agencies facing diminished resources, the City of Oakland can no longer afford to do business as usual. Over the past five years, the City has lost nearly a quarter of its staff. Yet, constituents increasingly expect a more open, participatory and responsive

government. Despite the worst financial crisis since the Great Depression, the City is now seeing modest growth. To continue on the path towards economic stabilization, it will require an investment in tools that will help the City better engage with citizens, increase efficiency and make more informed decisions, without putting additional burdens on existing staff.

The biggest costs to the City will be in the development of an Open Data portal. We are reviewing all options including:

- Building in-house with City staff using open source technology (CKAN and Dmpal)
- Hiring a third-party developer to build a platform and then maintain the platform in-house on City servers
- Purchasing a subscription with a leading Open Data platform provider to host, serve, visualize and sync our data

In all cases, the City can implement an Open Data system by the end of 2012 for under $50,000. Factors that will influence this decision will be staff time needed and long-term sustainability.

OUTCOME

While governments have for some time made statistical datasets (e.g. aggregated Census information) available publicly—originally in published books and tables, then on disks, and, for the past 10-15 years, downloadable or viewable on public-facing websites—the recent open government movement is unique in that it applies a standards-based approach to the release of government datasets to the public.

Data made public under open govemment standards adheres to stricter, reviewable guidelines than datasets of the past. Data will be provided at one "portal" address on the City of Oakland's website, rather than spread over dozens of agency, department and division web pages.

"Metadata"—information about the dataset—will include information about the fields in the dataset, how often the dataset will be updated or refreshed, the source of the data and who published it. In addition, open govemment data is presented in a way that makes it possible for developers of web-based or mobile applications to sync to the dataset to feed current and even real-time data to the application as it is used.

New technology provides the opportunity for improved City services and increased engagement with citizens. By opening up data, the City is paving the way for increased engagement with citizens through web, smartphone and SMS text software applications (apps).

Today's empowered citizens expect the same level of service from their govemment that they receive from private-sector organizations, such as retailers and financial institutions. They want

24-hour access to information and services through the communications channel and device of their choice.

In order to engage all citizens, the City must provide them with multiple choices for connecting to City government, including face-to-face communications, telephone, online portals, email and social media.

Implementation and Planning

The City of Oakland has already taken steps to embrace Open Data. Many datasets are already posted online, in various formats including Excel files, PDF and in tables on web pages. Staff has begun to identify these datasets and who they are maintained by. An internal multi-agency working group has been established to help drive the process and to complete the steps needed to implement an Open Data system. The steps to implementation are:

- Identify datasets that are available, and who they are maintained by
- Prioritize datasets based on levels of public interest and the value of the data to the community
- Review data for accuracy, clean up the data and redact confidential information
- Tie the data to an automated script that collects the data as it is updated making the process as seamless as possible for employees
- Create an online data "portal" to publish the datasets to a central location that allows for the data to be viewed, visualized or synced
- Publish data and launch the open data initiative
- Gather feedback, analyze user downloads and syncs and continue locating new datasets to post online

The initial portal will not be comprehensive of all City data. Staff will start with "low hanging fruit," easily available data in digital and publishable formats. Staff will then prioritize data based on the value of the datasets, outlined by the needs and expectations of citizens and developers. As additional data becomes available in digital and publishable formats, it will be automatically tied to the portal.

Identifying and Prioritizing Data

In an Open Govemment Data Benchmark Study released this year by Socrata, Inc., researchers identified high-value data categories from three perspectives: citizens, govemment employees and civic application developers.

Stakeholder groups were formed and asked to rate the relative importance of 15 categories of public data, based on their perceived value. The top five datasets from each category are below:

Citizens	Government Employees	Civic Application Developers
Public Safety *(crime data, food inspection)*	Gov. Services *(where, when, how services can be accessed)*	Public Safety *(crime, food inspection)*
Financial *(e.g. Gov. expenditures)*	Environmental *(e.g. air/water quality)*	Transportation *(e.g. parking, transit, traffic)*
Accountability *(e.g. campaign finance)*	Public Safety *(crime data, food inspection)*	Property *(property taxes, zoning maps)*
Education *(e.g. school tests)*	Legislative *(e.g. voting records)*	Environmental *(e.g. air/water quality)*
Gov. Services *(where, when, how services can be accessed)*	Regulatory *(e.g. disclosure for regulated industries)*	Census *(Population, Economy)*

Using this study as a guide, the Department of Information Technology and the Office of the City Administrator met with a group of "data owners" throughout the City to brainstorm City owned relevant datasets that fall within these categories..

As a follow up to this, staff sent out a survey to those that attended the meeting, as well as to those who could not make it to follow up on the readiness of these identified datasets to be published, the location of the data, confidential fields that would need to be omitted and to determine accuracy of the data. With this additional information acquired, staff will be able to prioritize data to be published and set goals to determine what needs to happen before the rest of the data is launched.

This list does not represent a comprehensive list of all of the data that we collect and track within the City. As staff continue this Open Data inifiative, additional data will be revealed and new data will begin to be collected. In building the portal, staff will make it easy for new datasets to be idenfified and published on a regular basis.

Open Data Portal

Like most govemment agencies facing diminished resources, the City of Oakland can no longer afford to do business as usual. Over the past five years the City has lost nearly a quarter of our

staff. Yet, constituents expect a more open, participatory and responsive govemment. Despite the worst financial crisis since the Great Depression, the City is now seeing modest growth. To continue on the path towards economic stabilization, it will require an investment in tools that will help us better engage with citizens, increase efficiency and make more informed decisions, without putting more weight on existing staff.

Therefore, staff members are reviewing all options for an Open Data Portal for cost, staff-time and intemal infrastmcture needs. Three viable options have been identified.

- Build in-house with City staff using open source technology (CKAN and Dmpal)
- Hire a third-party developer to build a platform and then maintain the platform in-house on City servers
- Purchase a subscription with a leading Open Data platform provider to host, serve, visualize and sync City data

Criteria being used to evaluate an open data portal solutions are as follows:

- Directory and search capabilities to find data
- Client-side dataset analysis, visualization and mapping
- API enabled
- Community participation and moderation portal
- Social Media Integration
- Datasite Metrics & Analytics

In addition, City staff are in the process of speaking with other cities and agencies that have implemented Open Data portals in these three ways, to determine the pros and cons of the respective options.

Timeline for Implementation

Task	Timeframe to Implement
Identify Available Datasets	Ongoing
Prioritize Datasets	Ongoing
Review Data for Accuracy	Ongoing
Choose Open Data Platform	June 2012 – September 30, 2012
Build Open Data Platform	October 1, 2012 – November 30, 2012
Create Participation and Data Policies	October 1, 2012 – November 30, 2012
Load Initial Datasets into Platform	December 2012
Launch Open Data Platform	January 2013
Gather Feedback and Analytics	February 2013 -- Ongoing
Expand on Datasets Available	Ongoing

BACKGROUND

On April 30, the Oakland City Council passed a resolution supporting an Open Data initiative introduced by Councilmembers Schaaf and Kaplan. Direction was given to staff to retum to the Council with a report on the cost and benefits of implementation and an analysis of the steps necessary to implement an Open Data system.

Many cities and govemment entities have opened up their data and established policies, standards and best practices around the use of Open Data, Including the United States Govemment. President Obama signed the Memorandum on Transparency and Open Govemment on his first full day in office, ushering in a new era of open and accountable govemment to bridge the gap between the American people and their govemment. To date, 30 U.S. states and 15 cities have launched Open Data initiatives.

In October 2007, 30 Open Govemment advocates met to discuss how government could open up electronically-stored govemment data for public use. The conference, sponsored by the Sunlight Foundation resulted in eight principles that, if implemented, would empower the public's use of government-held data.

Since then, the Sunlight Foundation updated and expanded upon the list and identified ten principles that provide a lens to evaluate the extent to which govemment data is open and accessible to the public. The list is not exhaustive, and each principle exists along a continuum of openness.

1. Completeness
Datasets released by the govemment should be as complete as possible, reflecting the entirety of what is recorded about a particular subject. All raw information from a dataset should be released to the public, except to the extent necessary to comply with federal law regarding the release of personally identifiable information. Metadata that defines and explains the raw data should be included as well, along with formulas and explanations for how derived data was calculated. Doing so will permit users to understand the scope of information available and examine each data item at the greatest possible level of detail.

2. Primacy
Datasets released by the govemment should be primary source data. This includes the original information collected by the govemment, details on how the data was collected and the original source documents recording the collection of the data. Public dissemination will allow users to verify that information was collected properly and recorded accurately.

3. Timeliness
Datasets released by the govemment should be available to the public in a timely fashion.

Whenever feasible, information collected by the government should be released as quickly as it is gathered and collected. Priority should be given to data whose utility is time sensitive. Realtime information updates would maximize the utility the public can obtain from this information.

4. Ease of Physical and Electronic Access
Datasets released by the government should as accessible as possible, with accessibility defined as the ease with which information can be obtained, whether through physical or electronic means. Barriers to physical access include requirements to visit a particular office in person or requirements to comply with particular procedures (such as completing forms or submitting FOIA requests). Barriers to automated electronic access include making data accessible only via submitted forms or systems that require browser oriented technologies (e.g., Flash, Javascript, cookies or Java applets). By contrast, providing an interface for users to download all of the information stored in a database at once (known as "bulk" access) and the means to make specific calls for data through an Application Programming Interface (API) make data much more readily accessible. An aspect of this is "findability," which is the ability to easily locate and download content.

5. Machine readability
Machines can handle certain kinds of inputs much better than others. For example, handwritten notes on paper are very difficult for machines to process. Scaming text via Optical Character Recognition (OCR) results in many matching and formatting errors. Information shared in the widely used PDF format, for example, is very difficult for machines to parse. Thus, information should be stored in widely used file formats that easily lend themselves to machine processing. When other factors necessitate the use of difficult to parse formats, data should also be available in machine friendly formats. These files should be accompanied by documentation related to the format and how to use it in relation to the data.

6. Nondiscrimination
"Nondiscrimination" refers to who can access data and how they must do so. Barriers to use of data can include registration or membership requirements. Another barrier is the uses of "walled garden," which is when only some applications are allowed access to data. At its broadest, nondiscriminatory access to data means that any person can access the data at any time without having to identify him/herself or provide any justification for doing so.

7. Use of Commonly Owned Standards
Commonly owned (or "open") standards refers to who owns the format in which data is stored. For example, if only one company manufactures the program that can read a file where data is stored, access to that information is dependent upon use of the company's processing program. Sometimes that program is unavailable to the public at any cost, or is available, but for a fee. For example, Microsoft Excel is a fairly commonly used spreadsheet program which costs money to use. Freely available altemative formats often exist by which stored data can be accessed without

the need for a software license. Removing this cost makes the data available to a wider pool of potential users.

8. Licensing
The imposition of "Terms of Service," attribution requirements, restrictions on dissemination and so on acts as barriers to public use of data. Maximal openness includes clearly labeling public information as a work of the govemment and available without restrictions on use as part of the public domain.

9. Permanence
The capability of finding information over time is referred to as permanence. Information released by the govemment online should be sticky: It should be available online in archives in perpetuity. Often times, information is updated, changed or removed without any indication that an alteration has been made. Or, it is made available as a stream of data, but not archived anywhere. For best use by the public, information made available online should remain online, with appropriate version tracking and archiving over time .

10. Usage Costs
One of the greatest barriers to access to ostensibly publicly available information is the cost imposed on the public for access—even when the cost is *de minimus*. Governments use a number of bases for charging the public for access to their own documents: the costs of creating the information; a cost recovery basis (cost to produce the information divided by the expected number of purchasers); the cost to retrieve information; a per page or per inquiry cost; processing cost; the cost of duplication etc. Most govemment information is collected for governmental purposes, and the existence of user fees has little to no effect on whether the govemment gathers the data in the first place. Imposing fees for access skews the pool of who is willing (or able) to access information. It also may preclude transformative uses of the data that in tum generates business growth and tax revenues.

The City of Oakland currently makes a very small amount of public data available on various City websites and in various formats. Much of this data is currently available in only proprietary and/or visual formats (e.g. PDFs, maps), and almost none of it is raw data. The data would greatly benefit from being posted in raw formats on a central portal that utilizes a web application programming interface that would permit application programs to request and receive public data sets directly from the web portal.

ANALYSIS

The City of Oakland collects a broad range of different types of data in order to perform its tasks. By law, the data collected by the City is public data, and therefore could be made open and available for others to use. In today's "data driven economy" the public increasingly expects that data be readily available and accessible when they want it. It is impossible to predict precisely

what the value of our data will be in the future. The nature of innovation is that developments often come from unlikely places. The untapped potential can only be unleashed if the City turns public data into open data.

There have already been several successful efforts by City of Oakland staff to engage with citizens through the use of free third party apps, social media and Open Data.

- Last year, the City of Oakland's Department of Public Works opened up its service request data and integrated with SeeClickFix, an online and mobile tool that helps residents report, track and monitor non-emergency problems, such as graffiti, illegal dumping or potholes.

- The Oakland Police Department integrated with Nixie and Twitter to send out neighborhood specific Crime Alerts.

- Citizens have long had the ability to track and monitor crime through websites such as crimemapping.com and oakland.crimespotting.org, due to the Oakland Police Department and the Department of Information Technologies' efforts to open Crime Data.

Oakland is a dynamic hub of innovation and visionary thinkers. The City of Oakland is fortunate to have creative, committed and tech-savvy residents and community partners. Through a combination of data and planning, the City of Oakland could spark innovation and allow for the creation of dozens more free or low-cost customized apps that would create value for citizens and promote civic engagement.

Code for Oakland

In 2011, the John S. and James L. Knight Foundation ("Knight Foundation") and the Federal Communications Commission ("FCC") sponsored an Apps for Communities challenge. The goals of the Challenge were to: make local public information more personalized, usable, and accessible for all Americans; promote broadband adoption, particularly among Americans who are less likely to be regular Internet users (including low-income, rural, residents on Tribal lands, seniors, people with disabilities, and the low digital/English literacy communities); and create better links between Americans and services provided by local, state, Tribal and federal governments.

In response, the City of Oakland worked with community partners to develop the first annual Code for Oakland event where 100+ civicaliy engaged developers, coders, designers, entrepreneurs and innovators came together and created apps that met the requirements for the Apps for Communities challenge. The city, along with other local partners gathered 90+ local datasets that could be used to help drive the applications.

Apps and ideas that came out of the Code for Oakland event addressed community needs around reentry, transportation, urban farming, neighborhood watch, healthy food systems and youth engagement. Txt2wrk, one of the apps created during the Code for Oakland event, went on to receive the third prize in the national competition.

Txt2wrk is an SMS text messaging and feature phone based app that connects job seekers and employers by providing text-to-speech delivery of job postings on any mobile phone. Job seekers receive text message alerts of new job postings, listen to job descriptions, and submit job applications, 24-hours a day, all without a connection to the intemet.

By continuing a relationship with this engaged and innovative community we encourage the growth of niche markets, and grow apps locally that address our needs.

Open Data Use

While there is often talk about app creation when speaking of open data, it is hardly the only use. Publishing data online could minimize public records requests managed by staff. Staff currently respond to approximately 100 public records requests a month. Some of the requests are so extensive that hundreds of documents are produced for a single request. By publishing regularly requested data and information online in a raw, already redacted format, we estimate that we could over time eliminate at least 20% of the staff time dedicated to responding to public records requests. This alone would free up staff to work on other things, and make the development of the Open Data System cost neutral.

Citizens, journalists and students will be able to view the raw data, mash it up with other data, and visualize it in new and unfiltered ways, allowing them to make more informed decisions, and tell better stories.

In addition, it allows for other City employees and local agencies to get a better sense of the types of data that other departments are collecting, encourages re-use, and reduces duplication of data collection.

PUBLIC OUTREACH/INTEREST

The adoption of Open Data improves transparency, access to public information, and improved coordination and efficiencies among agencies and partner organizations across the public, non-profit and private sectors. Software applications and tools that enable Citizens to access, visualize, and analyze public information will encourage Citizens to provide feedback on local issues.

Essentially, opening up data is just the first step — and arguably, a necessary step to ensuring that data can be reused, contextualized, and interpreted in meaningful ways. However, Public input and education around Open Data is just as key of an ingredient to the success of an Open Data initiative as opening up the data in the first place.

Staff has been working with an adhoc group of technology and open govemment minded citizens to develop this plan and provide input on our Open Data initiative.

A key component to the Open Data platform that is chosen by the City of Oakland will be its ability to provide Dataset level performance metrics and analytics. In addition, we will facilitate regular online engagement with users and citizens to gather feedback on usefulness of the posted datasets, and input around what datasets should be launched next.

In addition to staff trainings and Open Data information sessions that are currentiy scheduled throughout the month of July, we will be participating in the 2nd amual Code for Oakland event where we will be asking citizens throughout Oakland for feedback about how technology could improve civic engagement, improve digital education and literacy in our residents, and provide tools to attract and sustain local businesses in Oakland. By gathering this information and seeing what tools developers wish to create, we will be able to better prioritize datasets that we release.

COORDINATION

This report was prepared in coordination with the Office of the City Administrator's Communications Director and Online Engagement Manager and the Department of Information Technology's Interim Director, as well as the Budget and City Attomey's office.

FISCAL IMPACT

While staff is still in the process of evaluating options, below are the estimated costs of building in house using Open Source solutions, or contracting with a proprietary vendor. An additional option would be to contract out the development of the platform and maintain it in house, the cost would be subject to an RFP process and staff hours and hardware costs would still apply.

Build in House		Build using Proprietary Software	
Upgrade Intemal Systems	$12k	Upgrade Internal Systems	$15k
Additional Staff Needed		Upfront Development Costs	$10k
System Programmer II – ½ FTE	$50k	Yearly Subscription	$30k to $40k
Database Administrator – ½ FTE	$63k	Total	**$55k** to $**65k**
Total	**$125k**		

SUSTAINABLE OPPORTUNITIES

Economic:
The Open Data initiative will save the City money in the long run by reducing the number of Public Records Requests and time staff spends tracking Public Records Requests. It will reduce duplication of efforts and result in more efficient City operations.

Environmental:
The initiative will reduce the use of paper record tracking.

Social Equity:
Technology enhancements will improve accessibility to public records for the disabled, improve transparency and increase civic engagement.

For questions regarding this report, please contact Nicole Neditch, Online Engagement, Office of the City Administrator, at (510) 435-9429.

Respectfully submitted,

SCOTT P. JOHNSON
Assistant City Administrator

Reviewed by:
Karen Boyd, Citywide Communications Director
Office of the City Administrator

Ken Gordon, Interim Director
Department of Information Technology

Prepared by:
Nicole Neditch, Online Engagement Manager
Office of the City Administrator