

King County Metro Transit Developer Workshop Notes

Purpose

This document provides bullet point notes and a synthesized summary of the ideas, feedback, questions, and recommendations from the Developers Workshop hosted by King County Metro on October 22, 2009. This document was made available to workshop participants for comment, correction, and addition prior to finalization and distribution to workshop attendees and workshop registrants who were not able to attend.

Workshop Bullets

Below are bullet notes captured during the workshop. They are provided here as a reference and an aid to memory.

Initial bullets after Kevin Desmond's keynote and Panel questions...

- Analyze and publish reliability data by route and time (when do buses miss schedules and connections e.g. for work commutes)
- Consider re-writing consent form disclaimer (protection for data user as well as for KCM)
- Is there a coordinated way to get data from other local transit agencies (e.g. Sound Transit)
- Who is the decision-maker? Who decides what gets implemented? How do you pitch an idea to Metro and get it through the system?
- Do I have to talk to you if I'm developing an application? - If it benefits the public?
- Are you going to a model where you supply the data - and the rest is up to us...?
- Love what you're saying – want to ask about how to get data **to** KCM as well as from (e.g. inbound to KCM from applications or users)
- Providing the ability to acquire entire datasets is the way to go
- #1 thing you can do is let people talk back to you
- Have you thought about a more formal process for correcting errors in the data (when identified by application developers or users) – love to see a bug tracker used for feedback and follow-up
- Possibly useful site is getsatisfaction (<http://www.getsatisfaction.com/>)
- How will you handle traffic and load for real time information (capacity and scaling)?
- Had no idea you were required to get permission to use KCM data (currently acquiring via UW) (Permission is a liability waiver for schedule data, real time data from UW doesn't require permission)
- Encourage you to look at a streaming model (for data provisioning)

- Consider use of pubsubhubbub (<http://code.google.com/p/pubsubhubbub/>)
- Am trying to grab *all* GTFS feeds (from multiple jurisdictions). Would love it if you put it up online (examples given included on Google's repository(<http://code.google.com/p/googletransitdatafeed/wiki/PublicFeeds>), Microsoft site (<http://ogdisdk.cloudapp.net/>) or federal site (<http://www.data.gov/>)) and not require disclaimer (or at least not signed disclaimer)
- Consider alternative licensing for people who are non-commercial
- (for tri-met) legal learning curve – do you think that other agencies can leverage what you've done to shortcut the process – which is extremely time consuming to go through
- (for tri-met) – how do you define “partnership” – (for metro) how do I get metro to advertise my application
- Providing a template (for licensure) - suggest GNU public license as a model – (for tri-met) can you help develop a license that could be a template that other jurisdictions could adopt?
- It doesn't scale for small companies to address a different license for each agency
- When will Metro GTFS be available?
- GIS – the data is “turn left - turn right” is there a shape file that tells you where the bus goes – is there any easier way to pull data out (from KCM distributions).
- Rather have a data source I could reference
- New plan for snow - any info on re-routing that will be real-time?
- Request data for “as-planned” as well as “current” service
- Real-time service change data would be really important
- Resources – Transit Developer Group (Google) - and GTFS changes forum

Breakout Group One (Stephen/Gunnar)

- Feedback on errors – e.g. bugzilla (<http://www.bugzilla.org/>)
- How to access raw AVL feed – security issues? – expose current feed now! – do not wait for GPS
- Hold a walkthrough on current data
- Make some version of “hello world”
- WIKI
- Mailing list
- Announcement / notification mechanism
- How to transition from AVL-> GPS
- Want to know ETA – of bus(es) to destination (e.g. realtime tracker lateness/misses for transfer buses in trip planner) – (use current capability like Brian F)
- KC working on an internal offering
- Thru-routing info -> one bus away
- Short-term goals: alert mechanism to developers on data additions – creation of a discussion group

- Long term: complete listing of all transit data
- If you have it – release it
- Priorities desirable
 - How can developers help?
 - Let developers work with raw data
 - KC – consider github (<http://github.com/>as a resource
 - Use a distributed version control system
 - Bug-tracking
 - Code sharing
 - Overall strategy for the above
 - Resource page on KCM website
 - Short-term: make all data available
 - Info resource to find out what government (KCM) limitations are – be current
 - Larger stop ID numbers on some markers

Breakout Group Two (Tim/Trang)

- What's coming – when is ST data coming?
- Where to get it?
- Notification when GTFS is done or new table is added
- Blog or forum for developers to post info
- Use existing blog or forum or add on
- Something is already out there – point people to it
- It's important to put current AND all historical data out for developers
- Is APC data going to the public
- Developers love raw data
- Build APIs – make it easy to use/access
- Put all data out
- Problems w API – name of the bus going
- Make platform neutral and open
- Consider use of REST API
- Full data set
- Make a new API
- Duplicate an OLD API
- New York API
- Real-Time API standard
- Easy discovery API
- Short Term Goal: Build APIs for real-time data and route data – create forum for communications and documentation
- Long-Term: Fill holes in APIs – predict arrival times

Breakout Group Three (Dan/Bob)

- Some have used transit data – some have not
- Make shape data more accessible
- Real time info off street to help feed systems

- Exceptions – record of (archival) AVL data could be used as a test feed to develop apps against
- Making snow (geo) areas available to developers – some people don't even know what our data means
- Re-route – status changes dynamically?
- Make time-route exceptions available in machine readable format
- Email alerts for re-reroutes – standard communication process regardless of method
- What level of granularity could public subscribe to? Can it be down to route level?
- Wants: GTFS is #1 – then Real-Time AVL (coach # important)

Is it worth doing this again – what form should that take?

- “all of the above”
- Online – message board – developer website / connection
- Quarterly meetings
- Open houses (regular)
- Conference calls
- Webcam/video event
- Email distribution list
- Focus on a challenge metro would like to have addressed (e.g. a coderush)
- Have a place to share code and data
- Metro developer wiki

Summary

Provision of Data and Functionality

A high priority for the group was provisioning of King County Metro transit data. In general, participants would like KCM to make any and all existing organizational data sources available for access and re-use. Such data should be easily discoverable and easily acquired. There should be the minimum possible restrictions on the usage of data, and the minimum possible amount of bureaucratic process required in order to obtain data.

It was suggested that KCM should not make assumptions about what particular data might be of interest or utility to the developer community, and that doing so would unnecessarily stifle creativity and innovation. In addition to current schedule, route, and vehicle location data, participants were interested in performance data, financial data, historical data, passenger count data, and data relating to planned or alternative service, as well as actual service. Examples given included the possibility that historical data might be used in an application measuring transit service performance trending, or as a source of test data for use during application development - the general suggestion being that the potential utility of data cannot be assessed until it is made available.

In regards to the form or format of data, participants suggested that raw datasets were desirable. Static or periodically updated datasets would ideally be able to be accessed in their entirety or queried against to retrieve a desired subset. Real-time information might be provisioned in a streaming form or using a publish-and-subscribe method that would allow the retrieval of a set of base data and subscription to subsequent changes.

There appeared to be some preference for the availability of a KCM hosted data service that could be referenced from within an application, as opposed to downloadable datasets, although the latter was also seen as desirable.

In general, it was suggested that existing data should be made available as soon as possible in whatever form they currently exist, with follow on efforts to increase the frequency or the sophistication of the method by which such data is made available. The rough progression suggested appeared to be: existing datasets in current form, followed by datasets standardized to a common format or specification, online datasets which support query capabilities, raw real-time data streams, and defined data services and APIs.

Attendees expressed concern that any data service developed and hosted by KCM be architected with the physical infrastructure, expansibility, and performance characteristics sufficient to support the potential capacity and load demand imposed by calling applications. Suggestions included development of a request prioritization scheme that would prioritize resource allocation according to call type (prioritizing calls for real-time data higher than those for static data for example) or the possibility of hosting such a service on a robust infrastructure external to KCM.

It was suggested that, in developing externally-accessible data services or application programming interfaces, KCM should plan for the ability to accept incoming data as well. Some examples given included providing for inbound notification of errors in data, the ability for applications to report (or permit the applications' users to report) service performance or customer experience information.

In addition to planning for inbound data flows, a related suggestion appeared to be developing an open customer feedback system (analogous to the "bug-tracking" software commonly used in the software development process) that would allow registration and open review of customer reports on various issues, track patterns of reported issues, and record the follow-up actions taken by KCM. The concept appeared to be that this could facilitate improved citizen feedback and ownership, function as a public accountability mechanism for KCM, enable a structured approach to identifying emergent service issues, and aid in developing transit service priorities.

In addition to provisioning data, there was some suggestion that there may be useful pieces of functionality, common to many consuming applications, which could potentially be provided by KCM. One example given was for a geo-coding service.

It was identified that it would be desirable to obtain data from all regional transit organizations, ideally in a consistent format and with a single point of contact for acquisition, and that KCM could potentially play a role in facilitating any inter-agency cooperation required in order to make this possible. It was also suggested that other (non-transit) King County datasets, such as roads-related data may of relevance to the community.

Facilitating the Development Community

Another priority for attendees appeared to be facilitation of the development community. In general, attendees appeared to favor some measure of on-going communication and information sharing among the members of the development community themselves, and between the community and KCM. A number of ideas for achieving this were proposed.

For example, it was suggested that KCM could host, sponsor, or participate in a unique online resource for the community. Possibly useful components of such a resource were suggested and included message boards or online forums, notification groups, wiki's, content repositories, and code-sharing, bug-tracking, and version control facilities. It was suggested that such a resource could serve a broader community of transit-oriented developers across many jurisdictions. Alternatively, it was suggested that there are many such resources extant that could potentially be combined and used for the purpose of facilitating the community.

It was also suggested that KCM emplace a “developer resource” section on its website that would provide reader access to datasets, answers to frequently asked questions, and announcements of new, updated, or restructured datasets.

Although there appeared to a be clear emphasis on the desirability of using online resources for communication and information sharing, there was also some request for periodic in-person meeting or open-house events – with such events potentially being shared with online attendees through the use of conferencing tools.

Data Usage and Legal Requirements

Participants expressed concern about both the process of acquiring rights to re-use KCM data and the content of the data sharing agreement itself. The process was seen as cumbersome or unclear in that some participants were unaware that KCM provides some data currently, the currently required data sharing agreement is not readily found online, the requirement for a physical signature is perceived as onerous, and it may not be not clear if the agreement is required if KCM-generated data is sourced through the University of Washington. It was suggested that the content of the agreement may not be clear as to what, if any, legal rights are being asserted by KCM, whether there is a distinction in the rights being provided for those who make commercial versus non-commercial use of the data, and that the agreement may not be constructed to adequately identify and protect the rights of data recipients.

On a broader scale, developers have faced a challenge in dealing with a diversity of data licensing agreements and requirements when acquiring data from multiple transit agencies. The development, promotion, and broad adoption of some common template for such an agreement was seen as desirable.

Questions for KCM

Participants posed a number of questions for KCM's consideration. Many of these can be summarized as a request for a defined strategy, process, and goals for how KCM would propose to partner with or facilitate the work of developers. Examples include a desire to know if KCM intends to partner directly with developers to create applications or if its role would primarily be in the provision of data; how a developer could make an application development proposal to KCM, what that process might entail, and who would be determining which such proposals might be pursued; whether KCM would advertise or promote externally developed applications, and what resources might be made available to answer technical questions, provide data walkthroughs, or speak to KCM systems and data limitations and capabilities.