Content

red marked - need discussion
blue marked - need discussion especially with John
green marked - important

A Agenda

0 Which code base for publicdata.eu? version 2.0alpha

5 data views for datasets as a mean to understand what is in the dataset << not as a personalization feature but as a part of the dataset itself created by the owner
5.1 How to re-use them for different datasets (relevant?, I think so: "Budget of Berlin 2010" & "Budget of Berlin 2011" might have the very same format)?
5.2 Could users set their favorite view for a dataset?

6 Feeds (already there?) + Emails

0 Definitions

DEF1 Entity
 - user
 - dataset
 - organization

DEF2 Data View
 - user-defined view of a resource

DEF3 Subscription
DEF3.1 Single-Entity-Subscription (only a single entity is of interest and is set by the user)
 - follow user
 - follow dataset
 - follow organizations
DEF3.2 Multi-Entity-Subscription (many entities are of interest and it can change dynamically without user intervention)
 - subscribed to a search (set of keywords, organizations, tags, formats etc.)

DEF4 Subscribed Item
 - an entity that I have subscribed to by the corresponding subscription

DEF5 Notification Mean
 - bubbles with numbers in it (internal mean)
 - emails (external mean)

1 User Stories

Also see CKAN's previous activity and notification user stories (I've gone through these and made sure they're covered in this document too).

Also see CKAN's Trac tickets with keyword "activity_streams"

1.1 Subscription Service (i.e. "following")
S01
As a user I want to see a list of all my subscriptions (follow user, follow dataset, search subscriptions)
seanh: *We agreed (meeting Thu 20th Sep) that his might not be necessary if we can put "Unfollow" buttons on the activities in the activity stream themselves. Better to avoid needing a list of subscriptions if possible.*
sven*: I think this will depend on the final UX. So it's John's turn.*

S02
As a user I want to be able to follow and unfollow other users.
seanh: *This is in CKAN 1.8, but the follow/unfollow buttons are missing in CKAN 2.0. They need to be added back into the templates in CKAN 2.0.*

S03
As a user I want to be able to follow and unfollow datasets.
seanh: *This is in CKAN 1.8, but the follow/unfollow buttons are missing in CKAN 2.0. They need to be added back into the templates in CKAN 2.0.*

S04
As a user I want to be able to follow and unfollow groups.

S05
As a user I want to be able to follow and unfollow organizations.
seanh: *Organizations for CKAN 2.0 are currently under development, the optimistic hope is that it will be finished early October. The plan is for CKAN 2.0 to have both organizations and groups, two different things, see the organizations and groups spec.*

seanh: *For unfollowing the plan is to have "follow" buttons turn into "unfollow" buttons on user, dataset, group and organization pages when you are following them (as already happens with users and datasets in CKAN 1.8), and also to have "unfollow" buttons on the activities in your activity stream themselves. John needs to think about the UI design for this.*
sven: *Need to see whether a additional list (not necessarily a new page) is needed.*

S06

As a user I want to see a list of my single-entity-subscriptions where every single-entity-subscription is represented by the entity itself and the last activities of this entity.
seanh: *This sounds the same as S01 to me? I think avoid this unless it is really necessary*.
sven: *Yes, somehow, but we'll see how the final UI will look like. This is not totally clear right now because we need a visual feedback from John.*

S07
As a user I want to create a subscription by subscribing to a search. **Available in Sven's branch**
seanh: *Just to clarify, "subscribing" to a search is the same as "following" a user, dataset, group or organization, i.e. it means you will get activity from that search in your activity stream.*
sven: *And being notified in some way.*

S08a
As a user I want to see new datasets that matches my subscribed searches. **Available in Sven's branch**

S08b
As a user I want to see changed datasets that matches my subscribed searches. **Available in Sven's branch**

S09
As a user I want to have a page where I can see all the datasets that matched my search subscription. **Available in Sven's branch**
seanh: *Is this necessary? How is this different from the original search page itself? Would a link back to the original search page suffice?*

S10
As a user I want to unsubscribe from a search subscription. **Available in Sven's branch**
seanh: *In meeting Thu 20th Sep we discussed putting "unsubscribe" buttons on the activity stream items themselves to enable this. Perhaps if the user returns to the exact same search page again there should also be an "unsubscribe" button shown, and perhaps the activity should have a link back to the search page. Search page could then double as a page for editing/updating search subscriptions (see my comment on S12 above).*

S11
As a user I want to create SPARQL subscription that uses the semantic.ckan.net triple store to specify which datasets I am interested in. **Experimental, Sven's branch**
sven: *This is the research part of my thesis. I'll integrate LODStats into CKAN and then provide that data to the Virtuoso triple store. After that the user can query that data by using SPARQL and so he can retrieve the datasets that match his interests.*

The publisher-defined tags (value-name pairs) are not always entirely compete and maybe wrong (typos etc.). LODStats extracts statistical data from RDF-datasets and stores them into a triplestore. Then, the user can find datasets based upon a SPARQL query and should be able to subscribe to them as they can do with the regular search.
So, this feature is for experts and requires a certain knowledge of SPARQL. However, it provides a tailored and semantic-ful filter view of the available datasets in CKAN

S12
As a user I want to do the same stuff (watching activities and entities, unsubscribe etc.) with SPARQL subscription I can do with search subscriptions. **Experimental, Sven's branch**

S13 Showing numbers of followers
When a user, dataset, group or organization is shown (e.g. on its own page or in a list such as search results) show next to it the number of followers that it has. Clicking on the number should link through to the followers page for the entity (which shows a list of all the users that are following it).
seanh: *The backend for getting the number of followers or a list of the followers of a user or dataset is already implemented (will need to be implemented for groups and organizations as well).*
The frontend needs to be reimplemented in CKAN 2.0. Numbers of followers linking through to follower pages was implemented for users and datasets in CKAN 1.8, but it is missing from the CKAN 2.0 templates. Even in CKAN 1.8 number of followers was only shown on the user and dataset pages, not next to users and datasets in lists e.g. in search results.

1.2 Notification Service

N01
User: when I am logged in to the site I want to be notified if my activity stream has new activities.
- *Notification will be in the form of a "bubble" that appears in the top-right of the site when you have new activities, contains the number of new activities you have.*
- *"New" activities are defined as activities that occurred since the last time the user loaded her dashboard page.*
- *Presumably there is some cut-off point to prevent the number getting too long e.g. if you have more than 99 activities it just says "99+"*
- *When the user clicks on the button they are taken to their dashboard page (where their activity stream is) and loading this page automatically resets the number to 0.*
- *This just tells you how many new activities you have, it doesn't tell you anything about the contents of the activities, until you click on it and go to your dashboard.*

Note: there was some discussion about having multiple numbers, e.g. showing how many new activities you have from each user, dataset, group etc. that you follow. We decided this can come under filtering of the dashboard activity stream, which has its own ticket, we will not work on it as part of this ticket.
https://github.com/okfn/ckanext-pdeu/issues/13
Note: Sean thinks it would be cool if when you click on the bubble your activity stream is shown in a "popover" on top of the page you're on, rather than taking you away to your dashboard page. But lets maybe ticket that for later, if we want to do it at all.

N02
As a user I want to receive notifications by emails when there are unseen changes on CKAN in the form of a digest.

???
1. Subject
2. Body
3. Grouping

N03
As a user I want to turn on/off notifications by external means.
seanh: *Does "notifications by external means" mean email notifications? i.e. I want to be able to toggle email notifications on or off?*

Dashboard

DB01
As a user I want to see a consolidated activity stream of me and the entities I subscribed to.
seanh: ***This is available in CKAN 1.8.*** *It's still there in CKAN 2.0 but there are no links to the page, needs a bit of work to integrate it with the CKAN 2.0 UI.*

DB02
As a user I want to filter my dashboard activity stream to see only activities from my search subscriptions.

DB03
As a user I want to filter my dashboard activity stream to see only the activities from my followed datasets.

DB04
As a user I want to filter my dashboard activity stream to see only the activities of my organizations.

DB05
As a user I want to see a list of datasets that are about my region. **Experimental, Sven's branch**
seanh: *Sven's idea here is that CKAN will get the user's location (e.g. from their browser, FOAF file, user's input on profile) and show them a list of "Nearby datasets" that they might be interested in.*
sven: *Really great would it be, when I can integrate semantic data here.*

DB06
As a user I want to see a list of suggested datasets that match my interests. **Experimental, Sven's branch**
sven: *The same here: integration of semantic data would be great. I am working on this.*

DB07
As a user I want to see popularity statistics for my datasets within the last month.
seanh: *CKAN already has internal tracking of page views, it can track how many times a dataset has been viewed, how many times each of a dataset's resources has been viewed (visits to the resource preview page) and how many times the resource has been downloaded (clicks on the resource download button), so I think we can use this CKAN feature here.*
sven: *Yes, John need to think of visualization.*

DB08
As a user I want to see how many datasets of mine have been followed within the month.

seanh: *Hmm, showing the total number of followers a dataset has is easy (already supported in CKAN) but how many times followed recently will need some implementation work (it may be possible to do this by counting activity stream items in the db! Might be quite easy...)*
sven: *So, low priority. I think the already implemented stuff should be visualized first.*

DB09
As a user I want to see how many datasets of mine have been commented on within the month.
sven: *As for DB08. Evaluating the activity stream in order to get the data.*

Activity Streams

AS02
When I visit a dataset's page (whether I am logged in or not) I should see that dataset's activity
stream, which contains all updates to that dataset.
seanh*: Currently available in CKAN 1.8 but **not in CKAN 2.0**, needs to be integrated into the
new CKAN 2.0 templates.*

AS03
When I visit a group's page (whether I am logged in or not) I should see that group's activity
stream.
seanh: *Do we want this or not? I'm not sure. And what would be in the group's activity stream?*
sven: *John is currently sorting out the right activities for that. So, we've to wait for that.*

AS04
When I visit an organization's page (whether I am logged in or not) I should see that
organization's activity stream.
seanh: *Do we want this or not? I'm not sure. And what would be in the organization's activity
stream?*
sven: *John is currently sorting out the right activities for that. So, we've to wait for that.*

AS05 Site-wide activity stream
seanh: *In the backend a site-wide activity stream of all activity of the site is already
implemented and you can fetch it via the API. Do we want to show this somewhere on the site?*
sven: *Idea: Maybe somewhere on the CKAN startpage?*

AS06
As a user I want to see the activities in activity streams grouped together when the subjects
does multiple actions of the same type within 24 h.
seanh: *This is about collapsing multiple consecutive activity stream items into one item, to avoid
flooding of the activity stream. I think John had an idea for how to do this.*
sven: *AFAIK John needs to sort out the right activities for that.*

AS07 Activity streams pagination
Currently user, package and group activity streams only return the most recent 15 activities,
even though all activities are kept in the db. Do we want to add pagination - to both the API and

the HTML pages - to support retrieving older activities? http://trac.ckan.org/ticket/1746
seanh: *I'm not saying that we should implement this now (I think it's low priority) just putting here for completeness.*
sven: *I personally find pagination very annoying ever time I encounter it. I like the "load more" button as John proposed.*

AS08 See when a dataset was last updated

See the date of the last activity to a dataset on the dataset's page, linking through to the dataset's activity stream.
seanh: *This is another one from the old activity streams user stories, just putting here for completeness. If a dataset's activity stream is going to be shown on the dataset's main page anyway then this item is unnecessary, but if we want to put the activity stream on a different page then date of last activity might be a good idea for a link through to that page.*
sven*: could be one of the suggestions/recommendations: "newly added datasets"*

AS09 Atom and/or RSS feeds for activity streams

seanh: *It should be just a bit of templating to publish each activity stream as a feed also, but I don't think this is high priority.*

AS10 Activity shown next to users

As a Visitor I want to see what status people have by seeing small bits of info next to their name, e.g. a sign to indicate being a superuser/sysadmin and/or the number of datasets they have so that I know the approximate activity and authority of users I come across
seanh: *This is another one from the old activity spec. I think it's low priority and also not very well specified currently.*
sven: *Not very specific indeed. I can't understand it.*

LODStats Integration

L01 LODStats-XML/RDF integration
As a user I want to see an XML/RDF representation of a dataset containing semantic statistical data from LODStats. **Experimental, Sven's branch**

L02 LODStats-N3 integration
As a user I want to see an N3 representation of a dataset containing semantic statistical data from LODStats.**Experimental, Sven's branch**

L03 LODStats-TripleStore integration
As a user I want to use SPARQL interface on CKAN to query the corresponding triplestore which contains semantic statistical data from CKAN and LODStats. **Experimental, Sven's branch**

L04 LODStats-Subscription service 1 integration
As a user I want to subscribe to a SPARQL request when the response from the corresponding triplestore only includes only links to datasets on CKAN. **Experimental, Sven's branch**

L05 LODStats-Subscription service 2 integration
As a user I want to subscribe to a SPARQL request when the response from the corresponding triplestore includes other things than only links to datasets on CKAN. **Experimental, Sven's branch**

Idea:
- add some new tables to CKAN model, let LODStats do its work and save the data into these tables.
- integrate the data into read.n3 and read.rdf
- let a script export from CKAN and import into the triplestore (Virtuoso)
- find a way to sparql the triplestore on CKAN
- use the responses for: suggestions/recommendations, subscriptions / or use the normal database instead

Data Explorer

DE01 As a user I want to see data in the generally most appropriate form (table for spreadsheets, map for location data)

DE02 As a user I want to decide whether to create a new data view or choose from the list of all my data views when viewing a resource.

DE03 As a user I want to create a data view when viewing a resource.

DE04 As a user I want to choose a data view from the list of all my data views when viewing a resource.

DE05 As a user I want to see a list of all my data views.

DE06 As a user I want to see a list of the (datasets, resource) for which I had created a data view when I select a data view out of the list of all my data views.

DE07 As a user I want to remove a data view by removing it from the list of all my data views or by removing it from the resource's page.

- Include a **metric** to decide for when one data set is similar to another; improve preview functionality - default preview to be set to map view

Save view functionality (Recline technology)

http://publicdata.eu/ - has been integrated with the recline technology. However, some of the data sets (ie csv) are not easily viewed with in the Preview (grid, map, graph). What is the right format?

Currently, added the new http://reclinejs.com/ to publicdata.eu as part of upgrading CKAN (it allows you to preview tabular data (csv, xls) by storing as structured data (user can query & filter the data)

-----

2 Technical Stuff

Subscription Service
==subscription table = id, data type, definition type, definition, owner id, is activated, last evaluated, last modified==
id
data type = dataset, user, organization,
definition type = search
definition = data to define what items is this subscription about; adhering to definition type
owner id = id of the user that created the subscription
is activated
last evaluated
last modified

What are changes for ...?
User: - his activities
Dataset: - its activities
Group: - its activities[and the activities of its datasets and its users]
Search: - changes in the list of subscribed items [added, removed, changed etc.]

==subscription item table = id, subscription id, data, status==
id
subscription = foreign key to corresponding subscription
data = data describing the entity (id, name, description, images etc.) in JSON << presentation purposes
status = has the entity described in data changed since the last time the user has seen them?

Additionally integration of LODStats data to enhance metadata of datasets.

The system should not evaluate the changes immediately. It is maybe too much load for the servers. So, instead of online evaluation, batch evaluation should be preferred. (Or asynchronous instead synchronous evaluation)

So, there is function 'subscription_item_list_update' which updates the item list (and of their states) of a specific subscription.

subscription_item_list_update will be called:
- when the user comes to a page of a subscription
- for every subscription when the user is active on CKAN
- once in a while when the system has to send emails and the last evaluation of the subscription is

The overall goal is
       - to minimize the effort for the system
       - to maximize the up-to-date-ness of the user

Notification Service

provide some sort of notification list that will be flushed (i.e. the users get their emails and balloons pop-up) once in a while

maybe again on a statistical basis or just on predetermined times (every 35 minutes or so) - configurable

notification list might not be a real database table, notifications can be constructed on the fly

When should a user be informed about?

- when there are unseen activities to which the user has subscribed to
- after 24 h the last activity on CKAN
- at 6am, 12am, 6pm, 12pm

Dashboard

Let user input his location or his FOAF file or get it from his browser.

Compare the location with metadata from LODStats and display they as a list of suggestions.

Activity Streams

Option 1 (sorted by time Activity1 > Activity3 > Activity4 > ...)
Entity1 (last change: 2012-01-23)
	Activity1
	Activity2
Entity2
	Activity3 <<< special case: would not be displayed as a group
Entity3 (last change: …)
	Activity4
	Activity5
Entity4 (last change: …)
	Activity6
	Activity7
	Activity8
Entity5 (last change: …)
	Activity9
	Activity10

Option 2 (activities are grouped in the order they actually have)
Activity1
Activity3
Activity2
Entity3 (last change: …)
	Activity4
	Activity5
Activity6
Activity9
Entity4 (last change: …)
	Activity7
	Activity8
Activity10

Data Explorer

LODStats important to reduce processing work.

When having a special RDF is should be recognizable and display the data in a special way. Example:
unemployment data as RDF from the ministry of labor -> presentation as graph is preferred by some user << How to identify similar datasets?
- Have more or less the same structure (URIs are basically the same) and temporal/spatial granularity but differ in literal values (numbers and strings).
- Use (exactly) the same vocabularies. (LODStats)
Idea: count the number of equal URIs, when 90% equal => assumption: the same type of dataset