



FaceBase 2 Hub Progress and Plans

Carl Kesselman

University of Southern California

Facebase Hub, PI



Progress to date...

- Transition from U Pitt to ISI
- Clean up of existing data
- Gathering of project requirements via short-term teams (Imaging, Genome Browser and RNA-Seq)
- Initial new data model
- New request process for human data
- New wiki and mailing lists



Migration of FaceBase 2 Site

Transition of existing site completed in Oct.

- No interruption in service
- Downloads of both animal and human subject data ongoing
- No new data being added to current site
- Only essential updates being made
 - e.g. critical security patches to Drupal
- Great thanks to the U Pitt team for their assistance in a smooth and relatively painless transition



FaceBase 2 Hub

Significant progress made toward new hub

- **Design goals**
 - Simplified navigation/ease of use
 - Integration of richer tools for interacting with data
 - Streamlined upload and curation processes
 - Set stage for future functions such as personalization and more analysis
- **Completely new implementation**
 - No longer using Drupal
 - Better security, simpler software stack
 - Proper database service backend
 - Makes it easier to organize and find data
 - Web 2.0 application platform for data manipulation

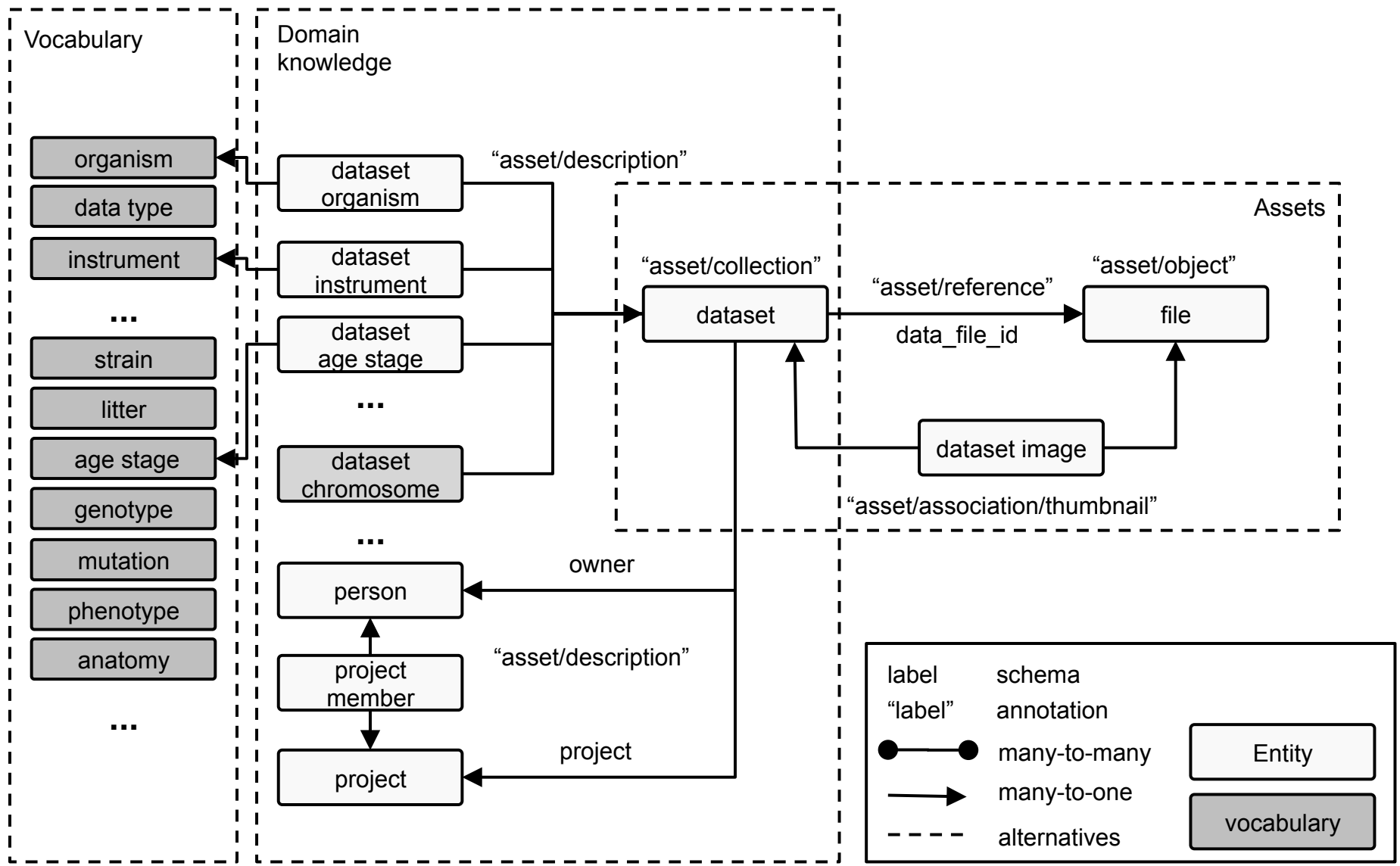


FaceBase Data Model

Significant work on rationalizing the structure of the data

- **Legacy Data**
 - Restructured existing data to be easier to search
 - Cleaned up a number of inconsistent values
 - Augmented the model with additional data types
 - Still a ways to go
- **Defined and are using more standardized terminology**
 - OCDM, standardized gene names, etc.
- **Richer, more consistent model for new data**
 - Will make it easier to find what you want, and download only what you want.
- **Initial progress with Imaging, RNA-Seq and Genome Browser**
short term groups on defining standard terms and values

FaceBase “Legacy” Schema



Leveraging the Gene Expression Omnibus for Craniofacial Research



- **Original grant proposal involved using Stanford Microarray Database software for Gene Expression analysis in FaceBase**
- **SMD development and maintenance is no longer supported and the code base would be difficult to maintain and adapt to FaceBase**
- **Functionality available in GEO is reaching that available at SMD, how can we leverage this for Craniofacial Research instead?**

Leveraging the Gene Expression Omnibus for Craniofacial Research



- **Several FaceBase datasets are already available in GEO**
 - These can be processed and analyzed using GEO2R
- **Two FaceBase datasets have been curated by GEO staff**
 - Data have been processed, grouped into biologically relevant differentials and associated with gene identifiers
 - Enables users to analyze the data without having to process it themselves
 - Enables users to view differentials for specific genes
 - E.g. Tgfbr3 at <http://facebasetest.usc.edu>



What we've done so far

- Enabled browsing and queries of *relevant* GEO datasets, directly from FaceBase
 - Merged views of curated and uncurated entries for the same data set
 - Queries by gene for curated datasets
 - Currently identifies relevant datasets using author-based queries: not optimal
- Associated each GEO dataset with the corresponding FaceBase dataset when possible
- Integrated into FaceBase II Data Browser



Questions for next steps

- **Can we develop a standard for GEO submissions from FaceBase projects?**
 - One-to-one relationships to data sets available at FaceBase
 - Putting “FaceBase” in the project field, similar to what ENCODE does
- **How can we get more FaceBase data sets curated by GEO? Should FaceBase researchers be involved in the curation process?**



FaceBase 2 Data Browser

- Faceted browsing
- More uniform display of data
- Integrated 3D data visualization
- New data types for OPT, etc.
- Integration with GEO-derived data products
- Auto-configuring to changes in data and data model



FaceBase 2 Preview

Caveats:

- This is all a work in progress, some things don't work yet, some things still missing, some things will change
- Data cleanup for legacy data still underway
- Very limited performance tuning
- “Mini-site” functionality not addressed yet
- UI will undergo UX improvements at some point in future



We want your feedback!!

- **Feedback form available from web site**
- **Tell us anything, no matter how small**
 - **We will respond to all submissions**
- **Again, please consider this a preview for a limited audience**
- **Form is linked from home page and on the Contact page.**



Information Sciences Institute

Demo.....

Dashboard

This is the name of a saved query

Bookmark Query

Collection **Mouse** ▾

click ☰ on the left to apply filters

☰ ☰ ⚙ +

displaying 20 of 648 results

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background
258	transgenic
accession	owner
FB00000007	steve.potter
description	
Gene expression microarray using tissue obtained by laser capture microdissection (LCM) from the Rathke pouch of mouse embryos at E10.5. The results are also presented on the Facebase Genome Browser Mirror local track <a href="http://	

Dashboard

This is the name of a saved query

Collection Mouse

click on the left to apply filters

displaying 20 of 648 results

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background	description
258	transgenic	Gene expression microarray using capture microdissection (LCM) from embryos at E10.5. The results are available in the Facebase Genome Browser Mirror: genomebrowser.facebase.org/cgi-bin/genomebrowser.facebase.org/lookup?org=mouse&db=mm9&accession=FB00000007 FaceBase_24Samp_Types_Averaged. This track represents samples of different regions of de
accession	owner	
FB00000007	steve.potter	
age-stages	microarray-platform	
{E10.5}	Affymetrix	
somite-count		
~36		

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background	description
258	transgenic	Gene expression microarray using capture microdissection (LCM) from embryos at E10.5. The results are available in the Facebase Genome Browser Mirror: genomebrowser.facebase.org/cgi-bin/genomebrowser.facebase.org/lookup?org=mouse&db=mm9&accession=FB00000007 FaceBase_24Samp_Types_Averaged. This track represents samples of different regions of de
accession	owner	
FB00000007	steve.potter	
age-stages	microarray-platform	
{E10.5}	Affymetrix	
somite-count		
~36		

Gene expression microarray - mouse E10.5 Rathke pouch

id	genetic-background	description
258	transgenic	Gene expression microarray using capture microdissection (LCM) from embryos at E10.5. The results are available in the Facebase Genome Browser Mirror: genomebrowser.facebase.org/cgi-bin/genomebrowser.facebase.org/lookup?org=mouse&db=mm9&accession=FB00000007 FaceBase_24Samp_Types_Averaged. This track represents samples of different regions of de
accession	owner	

Columns/Fields

DONE

enabled 18 of 23

Search

- All Fields 648
- id 648
- pubmed-id 80
- accession 648
- prior-accession 104
- owner 648
- title 648
- genetic-background 583
- gene 87
- genotype 404
- mutation 364
- age-stages 641
- microarray-platform 47
- affymetrix-genchip 41
- illumina-sequencer 17
- mp-new 18
- somite-count 27
- chromosome 73
- chromosome-start-position 73
- chromosome-end-position 73
- enhancer 73
- desc-html 648