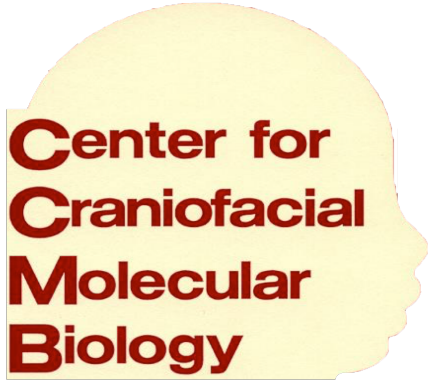
FaceBase 3: A Comprehensive Resource on Dental and Craniofacial Research

Rob Schuler, Carl Kesselman, Yang Chai

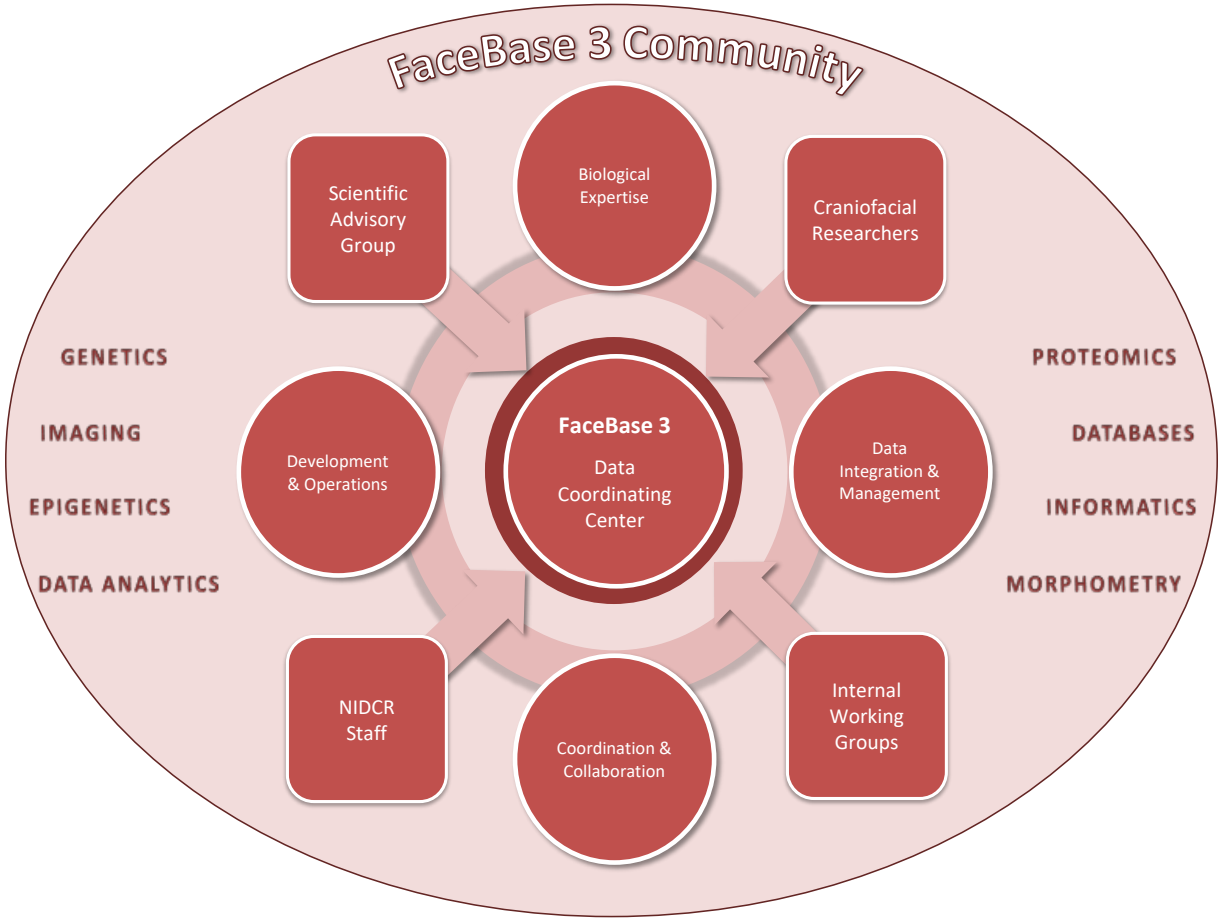
FaceBase Data Hub



Outline

- FaceBase 3 Overview
- Updates on resources available on FaceBase
- Sharing data through FaceBase

Organization and Mission

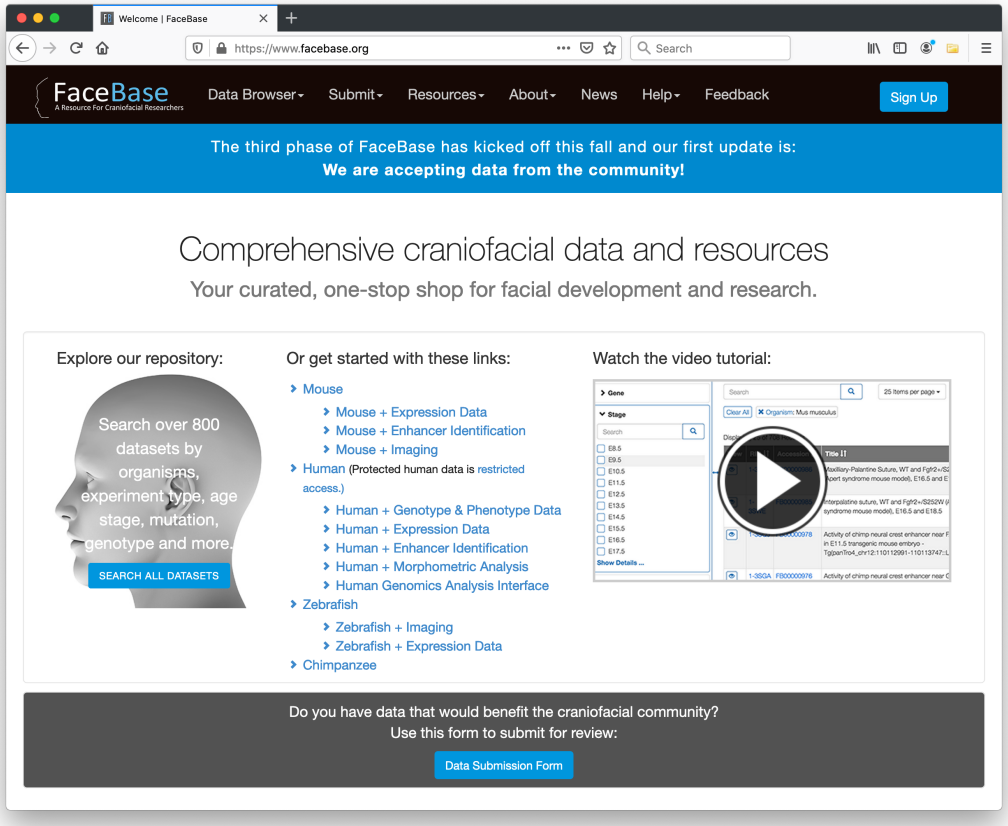


- To be a comprehensive resource on dental and craniofacial research -- fostering data sharing and curated resources.
- New objective to expand data to new experiment types, model organisms, and processes by engaging the whole community.
- FaceBase 3 is no longer a “hub and spoke” consortium but open to all researchers to participate.
- Comprised of a core team of craniofacial and computer science expertise w/ advisory and experts groups for oversight.

Status

- 850+ Datasets and growing
  - 746 mouse, 70 human, 39 zebrafish, 2 other
  - 630+ imaging, 120 sequencing, and other experiment types
  - 700+ detailed experiment records
  - 3,100+ detailed biosample records
  - 3,900+ images, 3,700+ sequencing, 1,650+ tracks, and more (~10 TB)
- Usage statistics (last 6 months)
  - 7,700+ visitors, 21,000~ views
  - 695 downloads
  - 3,400+ image views
  - 84,000+ track reads

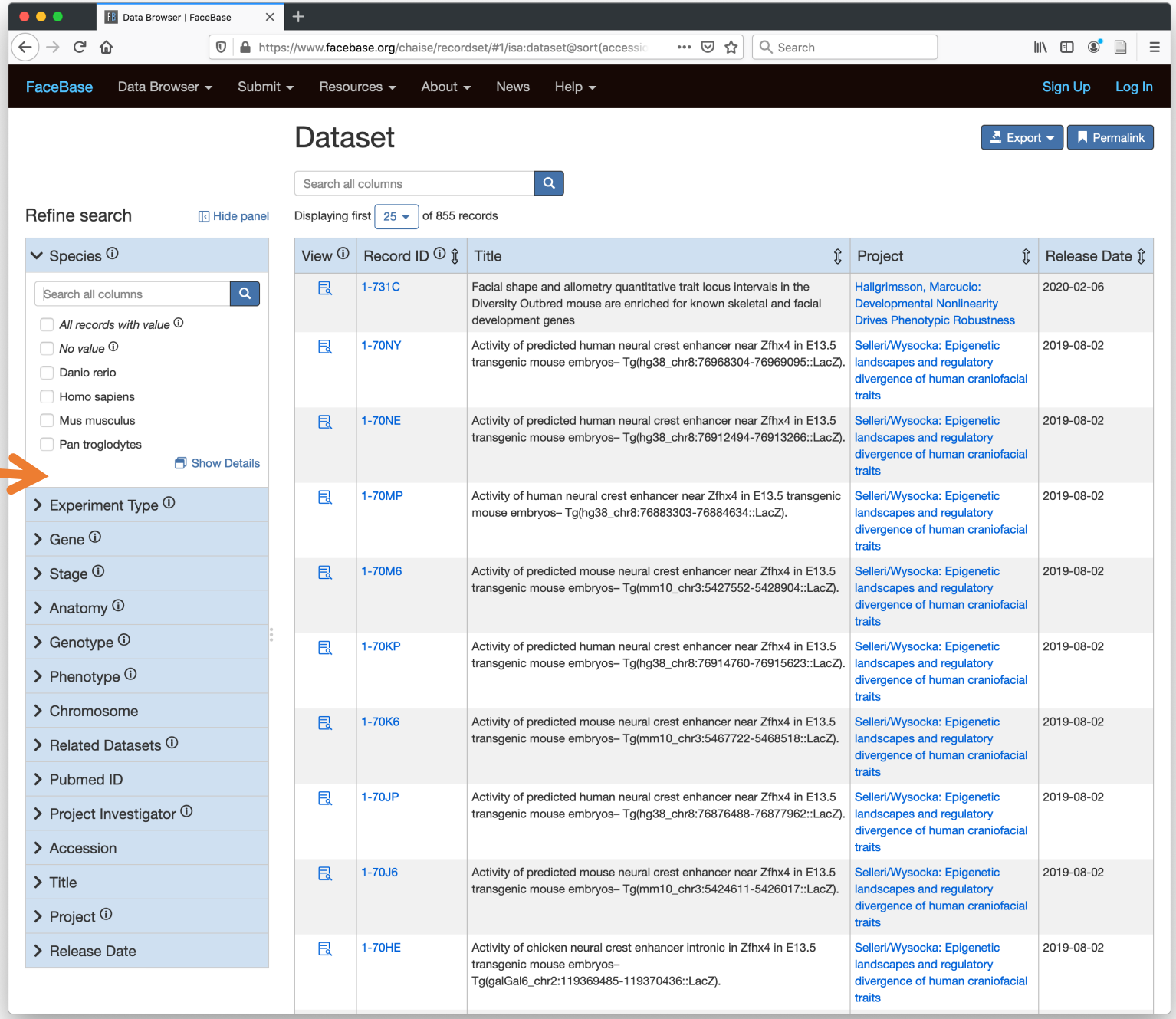
www.facebase.org



Data browser

- “Faceted Search” helps you find datasets using specific filters on species, experiment type, and other key attributes.
- Filters dynamically adjust to show you what is available – no “dead ends”

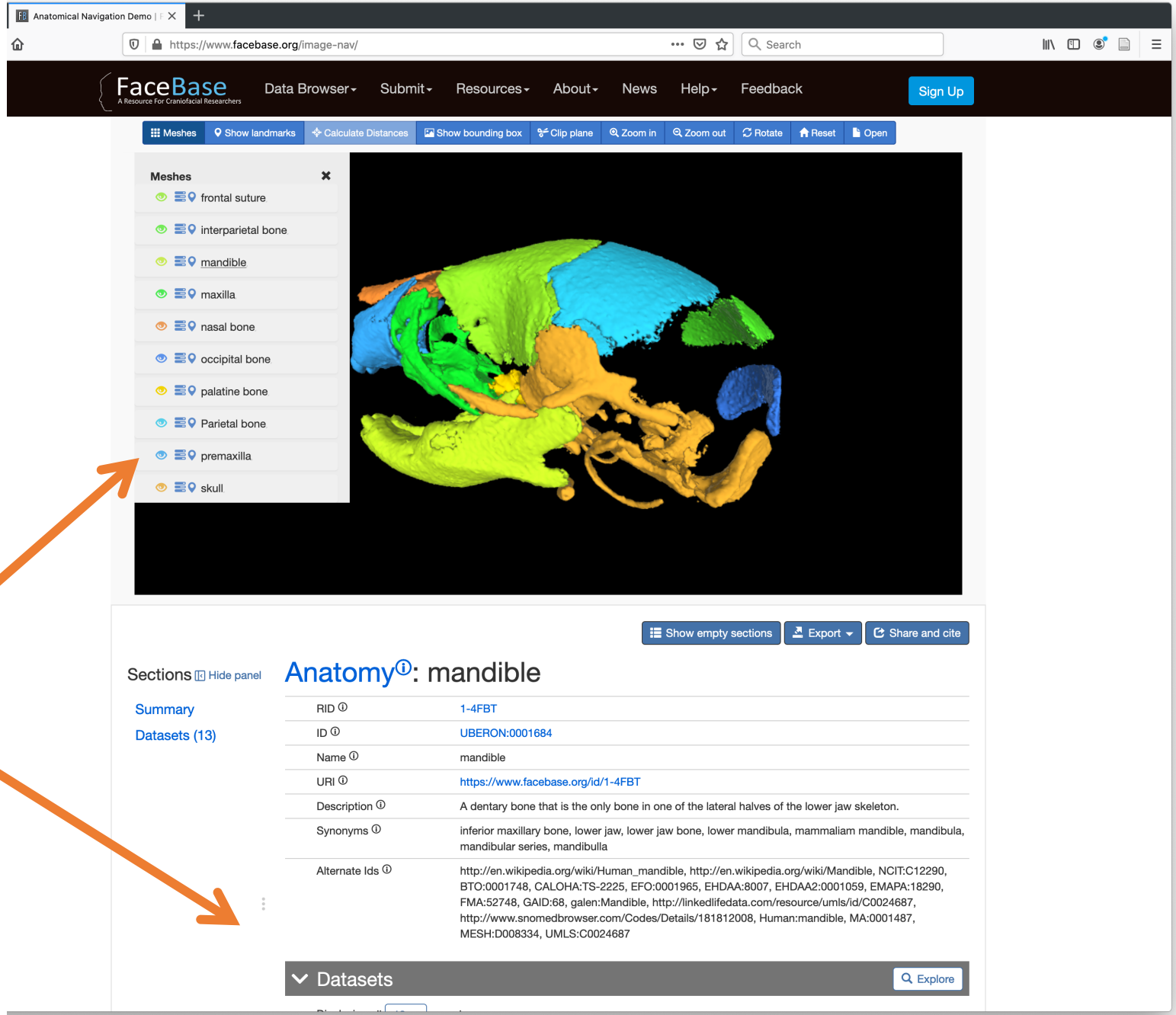
- Data Browser Link



Anatomical search

- Increasing focus on visually-guided anatomical search.
- Currently based on E18.5 wild type mouse model, but working toward other species and stages.
- **Linked to anatomical terms** and all datasets that are annotated with the term.

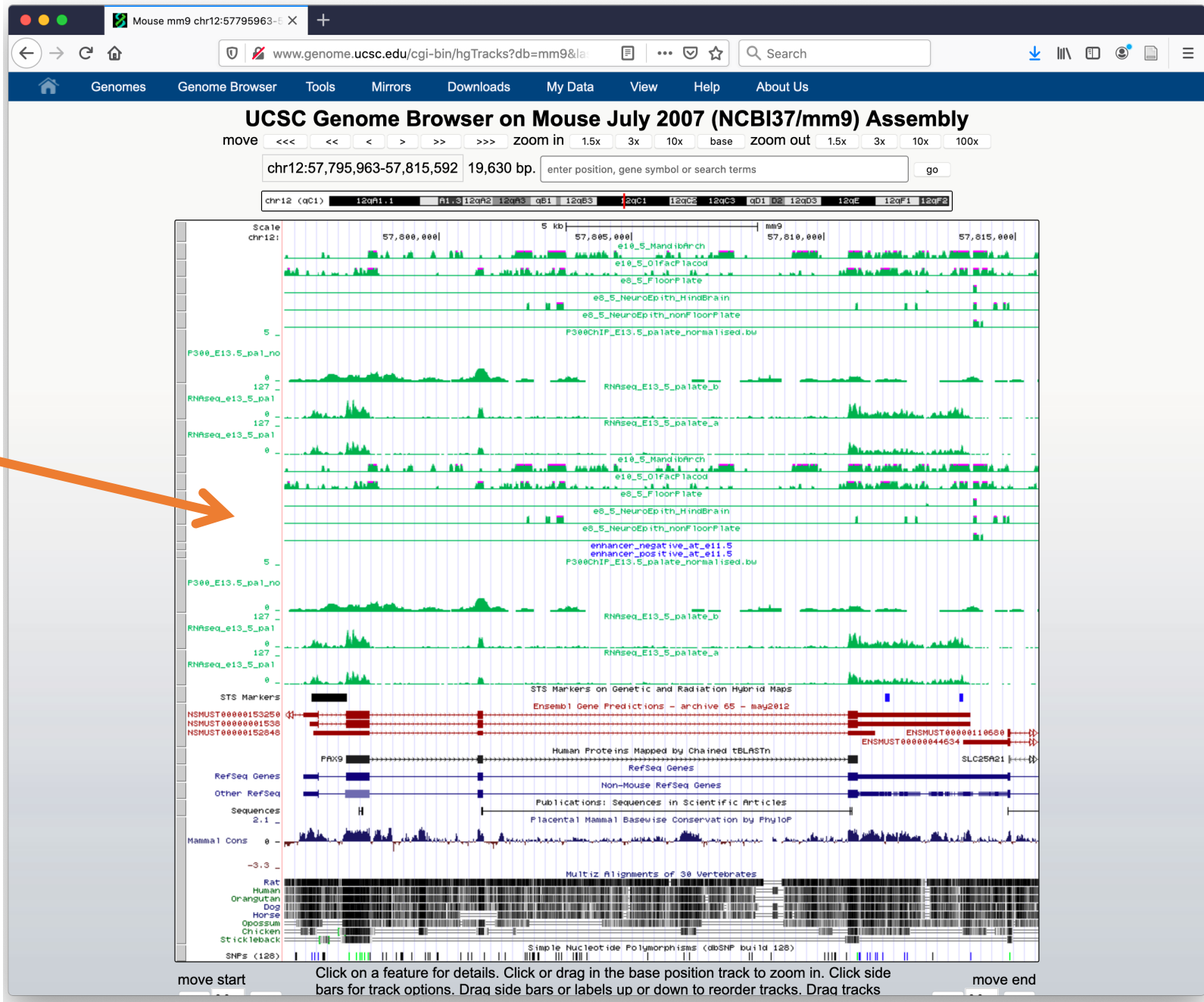
- [Image Navigation Link]
- [Example Zebrafish Dataset Link]



Genome browser

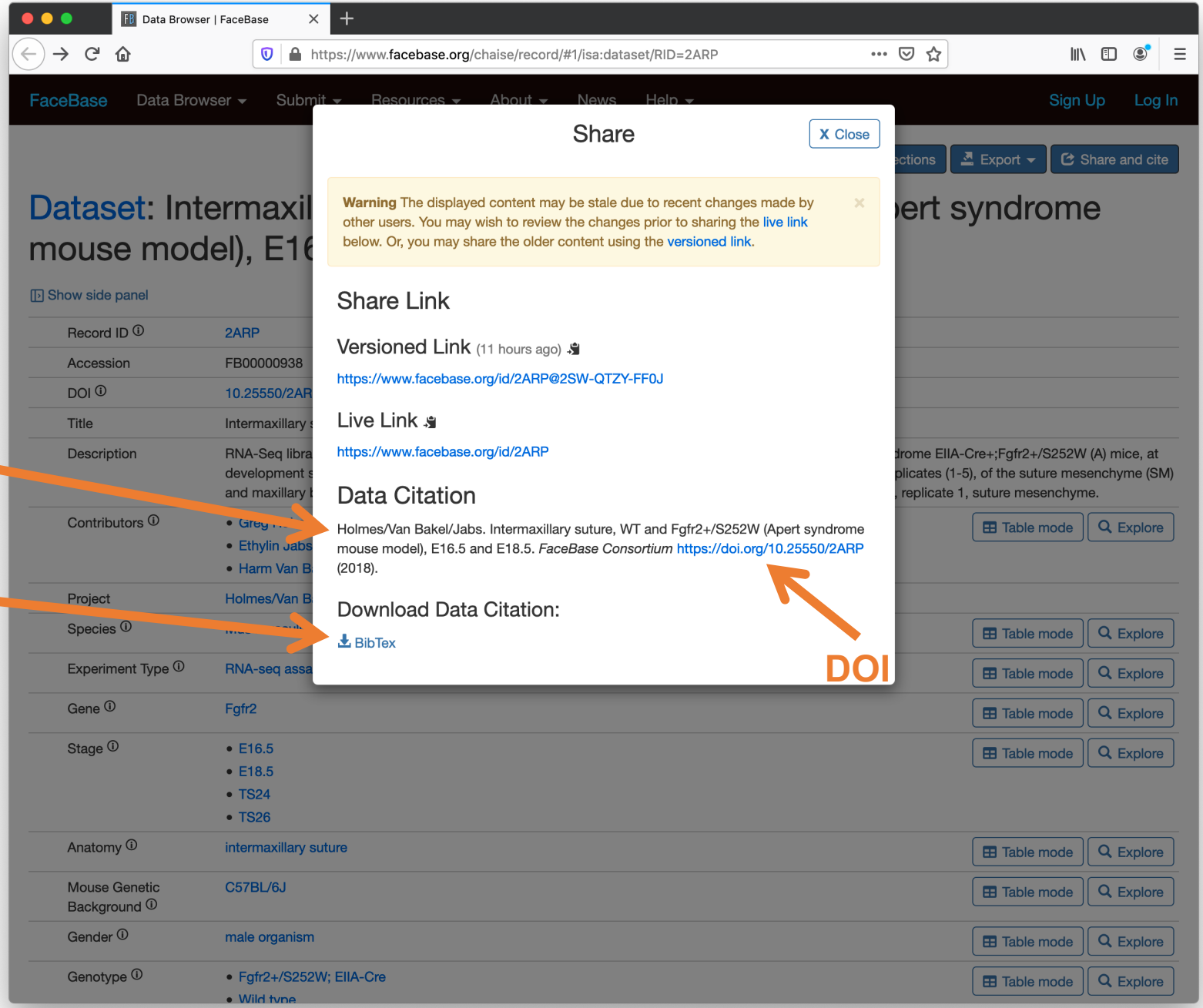
- FaceBase Track Hub viewed on UCSC Genome Browser
- Each FaceBase dataset w/ tracks has embedded Genome Browser
- Metadata links users back to FaceBase from UCSC Browser

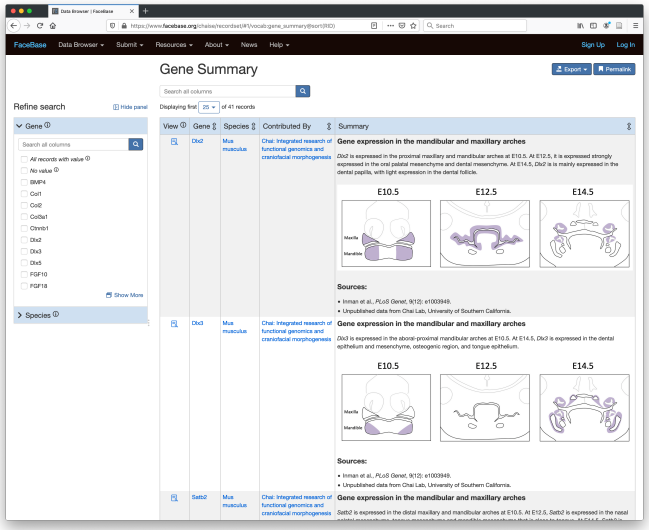
- [Example Dataset Link]



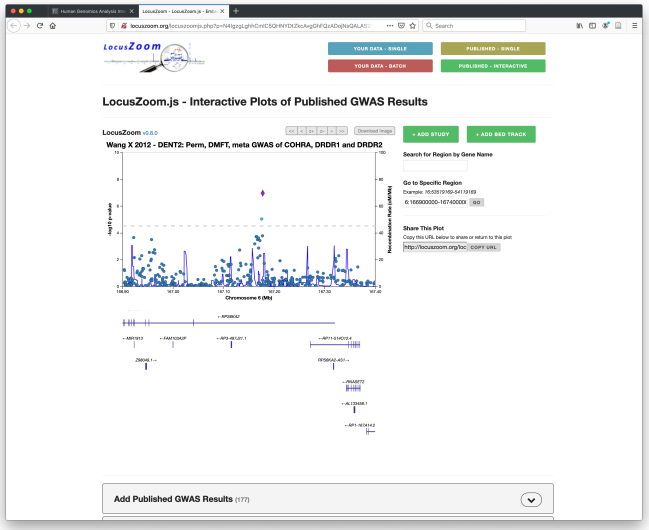
Data citation

- **Digital Object Identifier (DOI)** as stable, persistent, citable references to all FaceBase datasets
- **Data citations** formatted per leading publishers' standards and downloadable to reference managers (BibTex format)
- Proper citation improves the **dissemination** of FaceBase data
- Structured metadata records improve the **reproducibility** and **reuse** of FaceBase data.
- [Dataset Link]

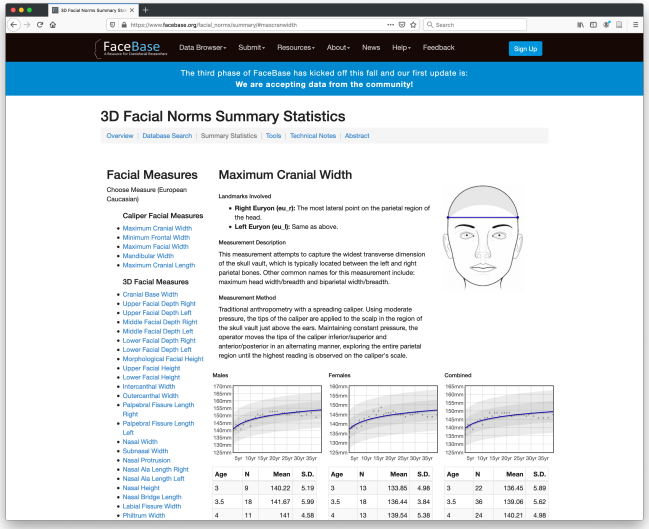




Detailed gene summaries



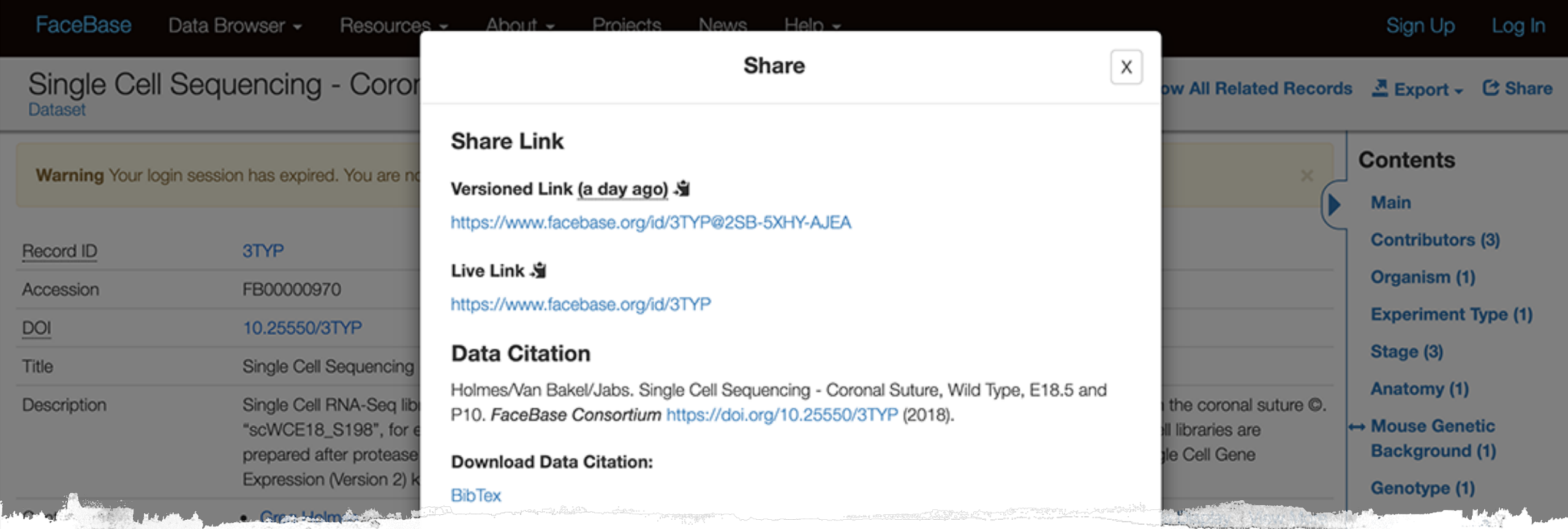
Interactive plots of GWAS data



Facial norms measures

More at the Resources Hub…

https://www.facebase.org/resources/



Why share your data?

- **Dedicated focus on craniofacial development**
- **Increase the visibility and impact of your research**
- **Cross-reference with publication**
- **Satisfy data sharing requirements for grants and publications**
- **Data are “published” like first-class academic works**

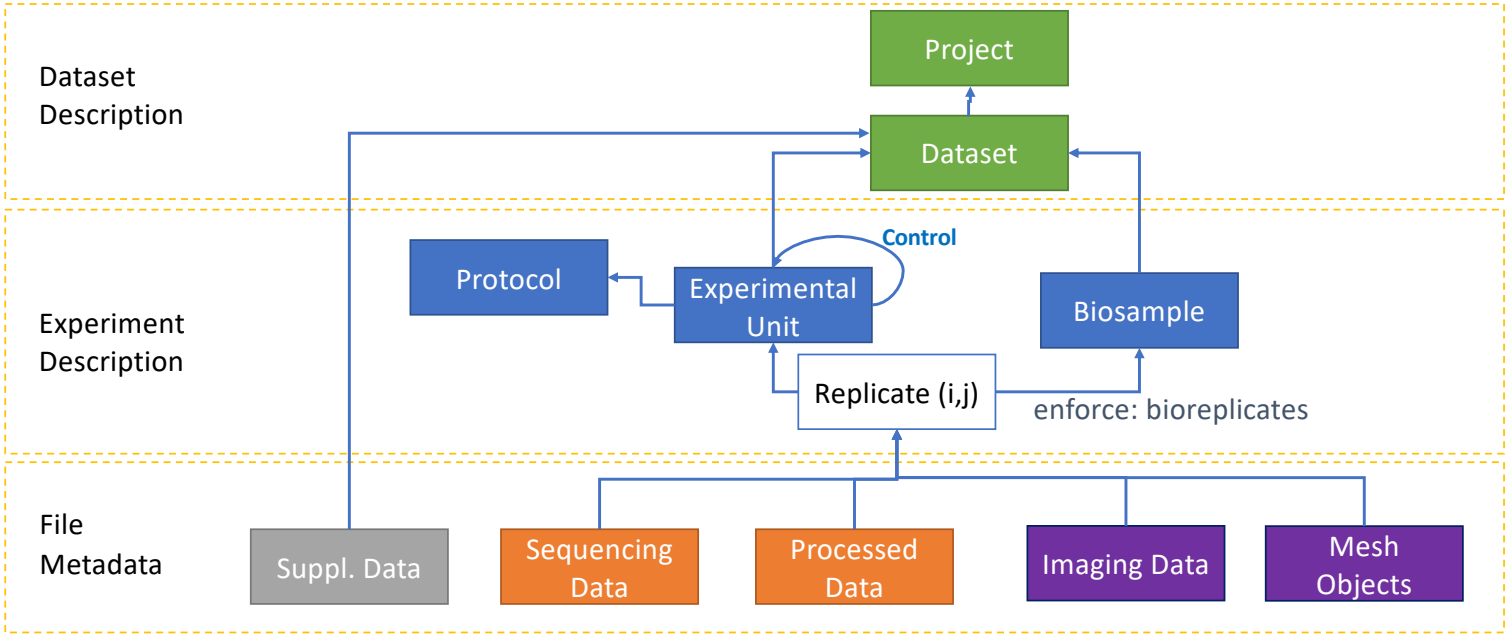
Example Dataset Link

What data are we prioritizing?

- Human, mouse, zebrafish, additionally chick and xenopus
- Gaps in existing data
  - More data on cell populations
  - Facial GWAS
  - Mouse histology, gene expression of e8.5-e10.5
  - Detailed stage transcriptional and epigenetic analysis of zebrafish and mice
- New data types
  - Single-cell RNA seq
  - Human and murine dental development
  - Spatial transcriptomics
  - Curated resource lists – antibodies, recombinant proteins, small molecules that have been authenticated for their utility in craniofacial research

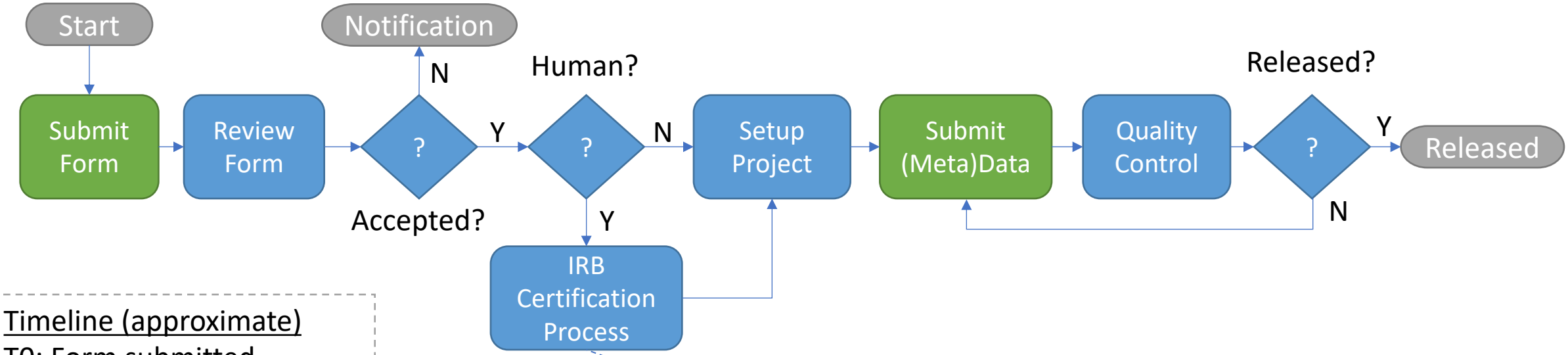
https://www.facebase.org/submit/data-priorities/

How are data organized?



- Dataset: the primary collection of data
- Experiment: generic experiment type (-omic, imaging, etc.)
- Biosample: details of specimen used
- Various Data: files and file metadata linked into the model

How to submit data to FaceBase?



Timeline (approximate)

T0: Form submitted

T+2 weeks: Review decision

T+3 weeks: Project setup

T+5 weeks: Submit data*

T+6 weeks: QC review

* Based on user averages

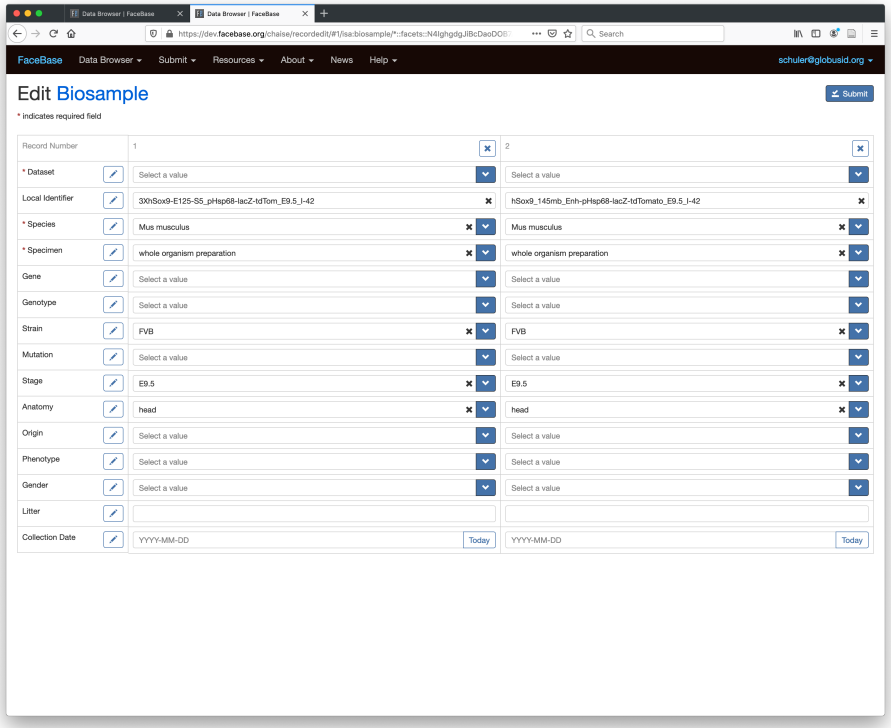
IRB Certification Process

- Individual level data classified as human subjects
- Requires USC certification of your IRB decision
- Tracks are not considered restricted data
- Timeline TBD

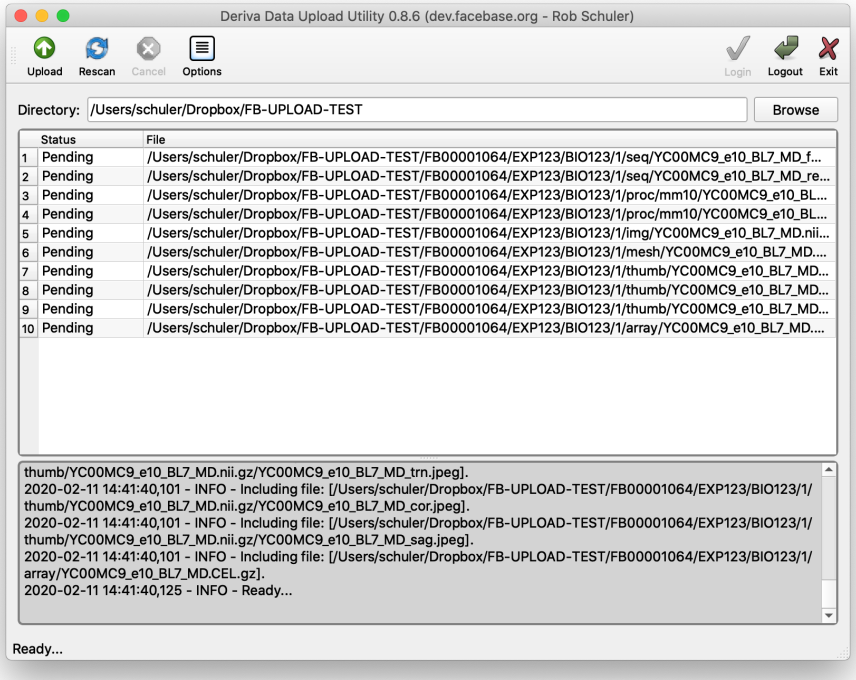
User Activity

Hub Activity

What tools are available for curation?



(1) Online data entry and file submission forms



(2) Graphical desktop applications for bulk data upload (mac, windows, linux)

(3) Command-line clients (mac, windows, linux) and **(4)** Python APIs (not pictured)

Researchers have submitted their own datasets with 100s to 1000s of files, usually in a few days

More information…

- Website: www.facebase.org
- Data submissions: www.facebase.org/submit/submitting-data
- Feedback: Link from website menubar
- Contact us: help@facebase.org
- Sponsor: NIH / NIDCR (U01DE028729)

- Leadership (Co-PIs): Carl Kesselman; Yang Chai
- Team: Rob Schuler (CS & Technical); Bridget Samuels (Biocuration); Alejandro Bugacov (Data); Cris Williams (Communications); Joe Hacia (Bioinformatics); Thach Vu Ho (Data curation)