Breakout Session 1: Track A

Hybrid- and Multi-Cloud Storage Strategies for Cost-effective Deployment of Data Resources

Dr. Robert Schuler
Lead Scientist, USC Information Sciences Institute

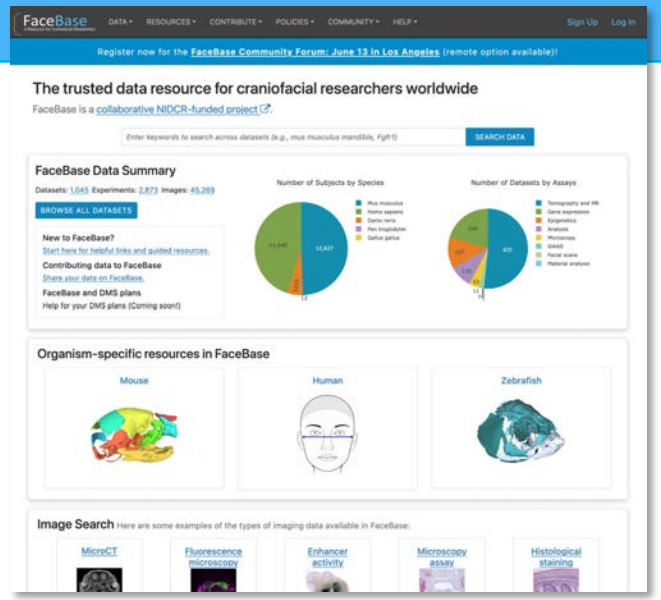
Hybrid- and Multi-Cloud Storage Strategies for Cost-effective Deployment of Data Resources

Rob Schuler, FaceBase (Co-I)

USC Information Sciences Institute

2024 NIH/ODSS Cloud Supplement Program PI Meeting

FaceBase Data Integration and Analysis Hub



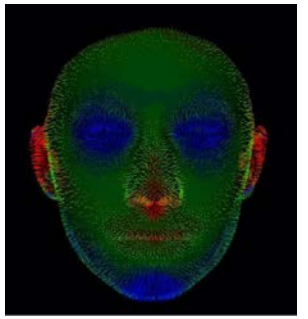
The trusted data resource for craniofacial researchers worldwide

Resources spanning dental, oral, and craniofacial (DOC) research

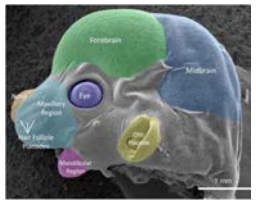
Established in 2009 in collaboration with the NIDCR

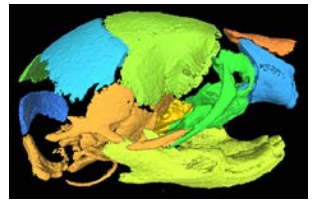
Goal: To serve as the trusted online data resource for craniofacial and dental researchers worldwide

1. Provide a comprehensive data resource in support of advancing research in craniofacial development and malformations
2. Promote multidisciplinary collaboration and research in the dental and craniofacial fields
3. Integrate genomic and phenotype data from multiple species
4. Utilize diverse resources and strategies to enhance data reproducibility

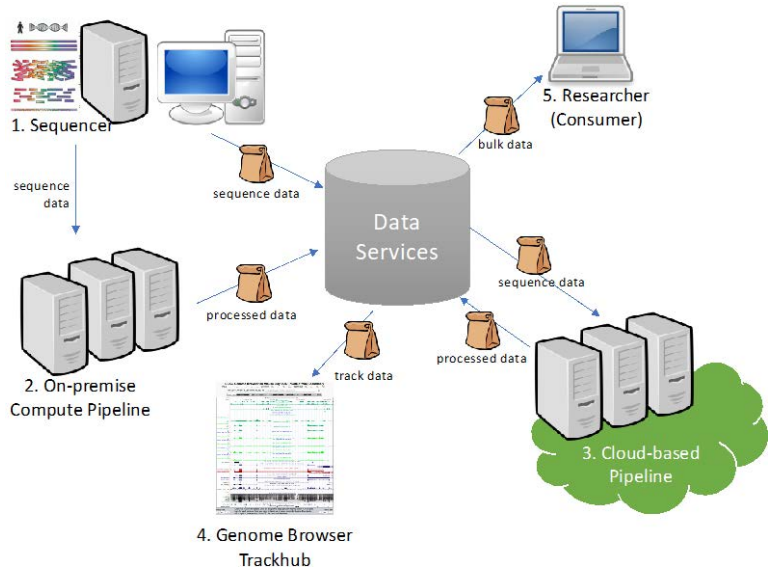


1,000s of *images* and *genomics* on human subjects and animal models





FaceBase Data Sciences Platform

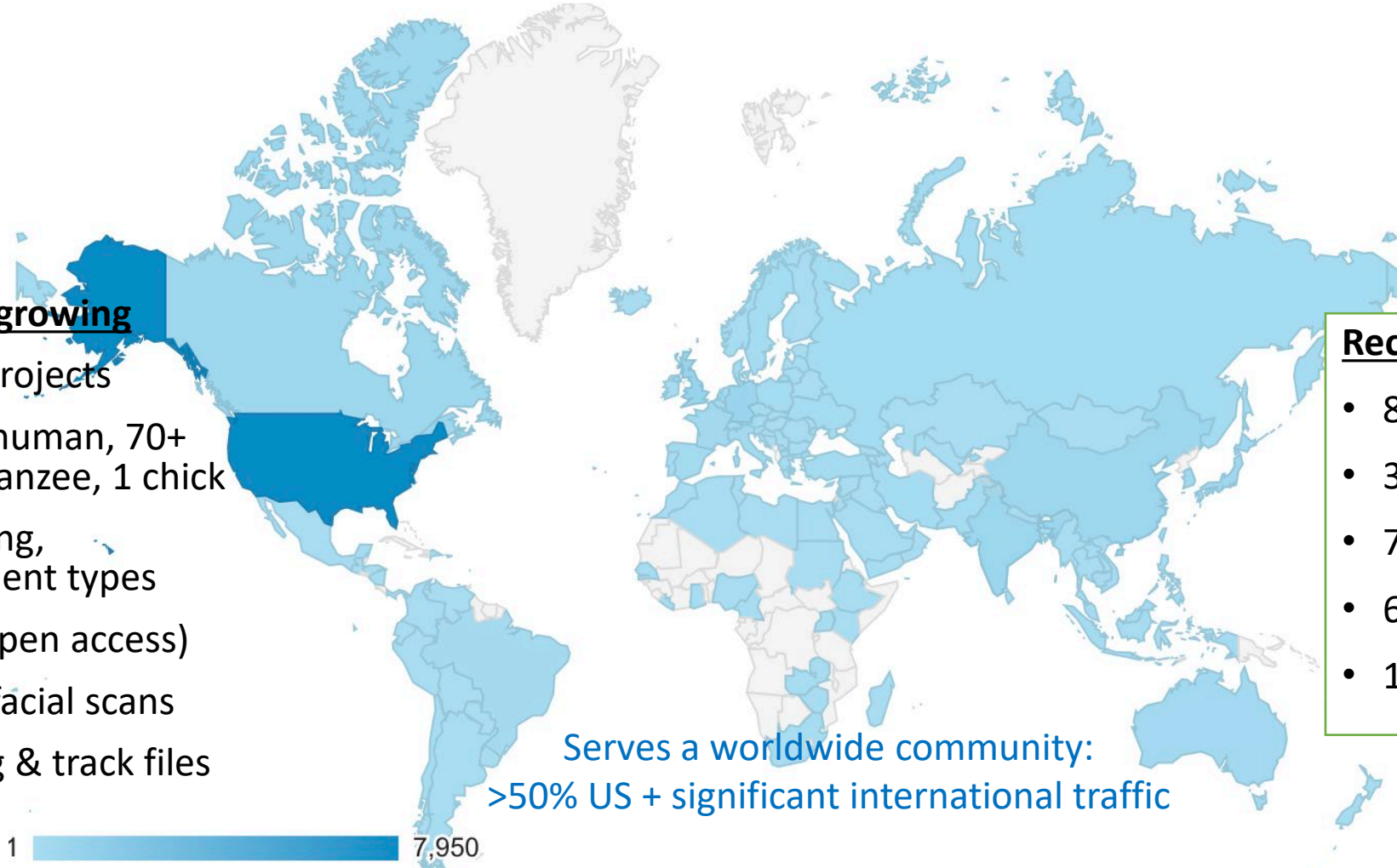


FaceBase
A Resource For Craniofacial Researchers

FaceBase Serves Global Research Community

1050+ Datasets and growing

- 50+ contributing projects
- 890+ mouse, 80+ human, 70+ zebrafish, 2 chimpanzee, 1 chick
- Imaging, sequencing, and other experiment types
- 45,000+ images (open access)
- 10,000+ (human) facial scans
- 7,300+ sequencing & track files



Serves a worldwide community:
>50% US + significant international traffic

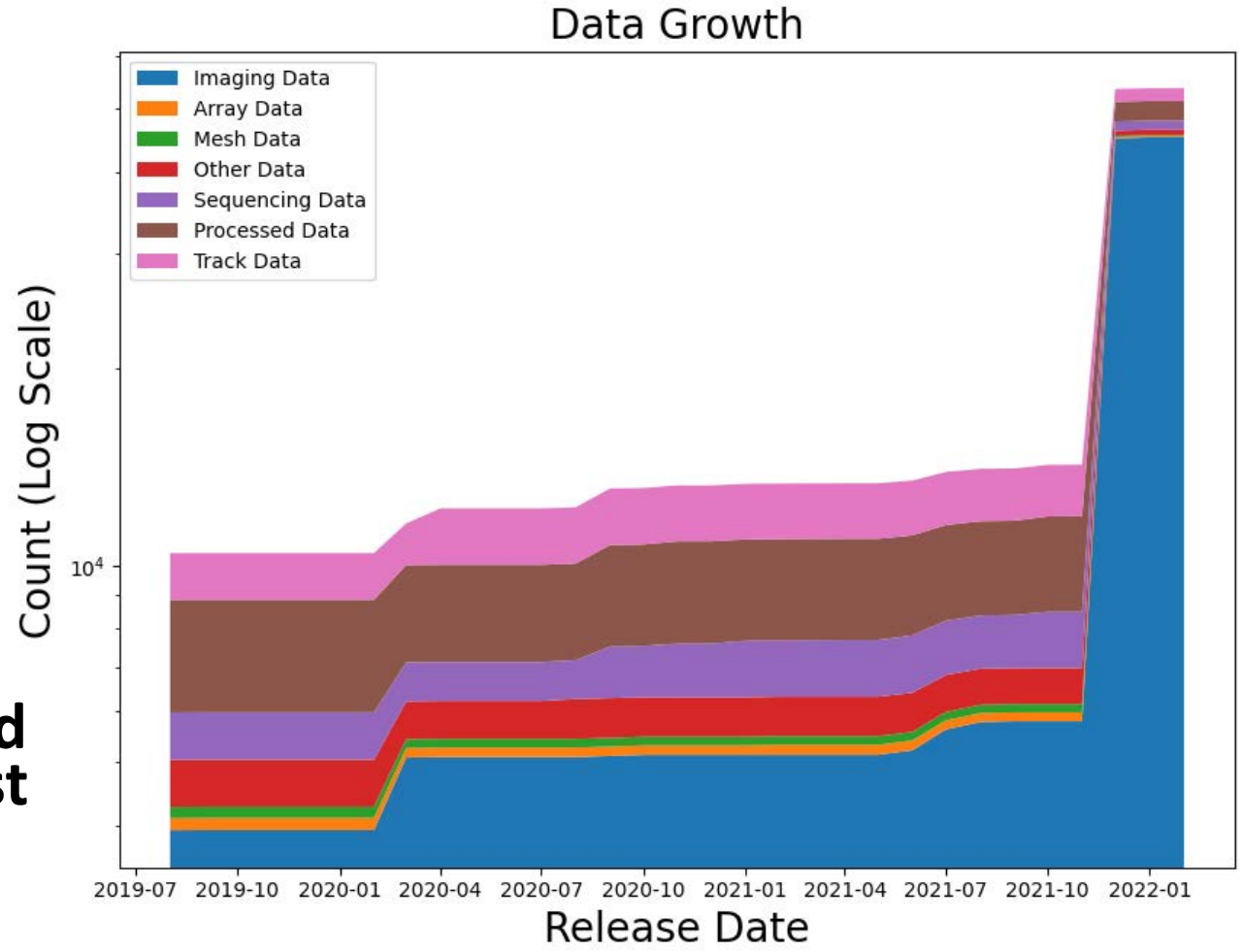
Recent 6 mo. Usage

- 8,500+ visitors
- 31,000+ page views
- 7,200+ downloads
- 6,500+ image views
- 165,000+ track views

Total number of publications: 200+ (regarding FaceBase data and resources)

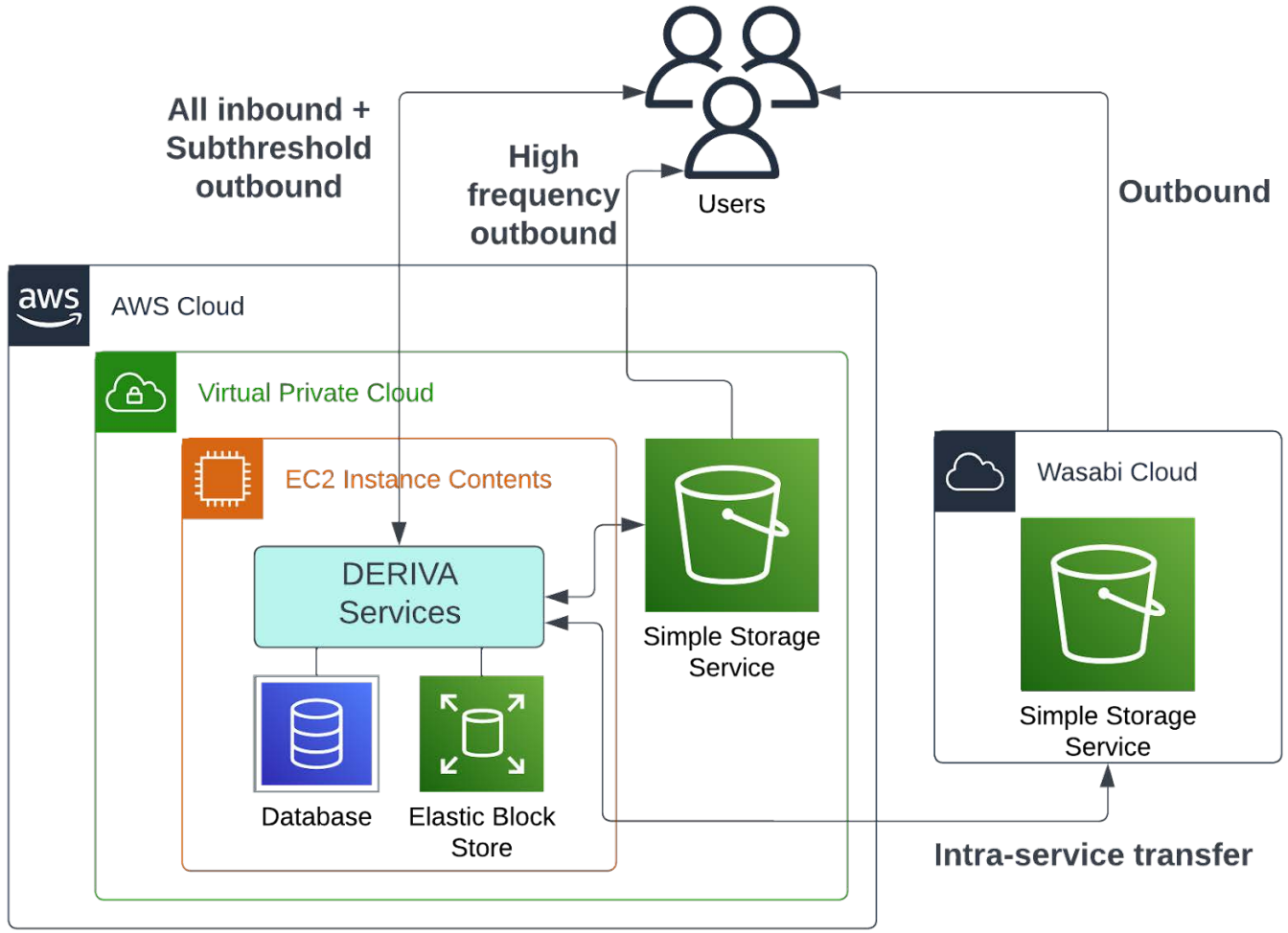
Challenged to Scale with Data Volume Growth

- On the Cloud…
  - Increased data volume = +$
  - Increased data access = +$
  - Difficult to predict too
- AWS Simple Storage Service (S3)
  - AWS = Amazon Cloud Storage Provider (CSP)
  - $0.023/GB-month (ex. 1TB = $23/mo)
  - “Intelligent Tiering” to drop cost with caveats based on frequency of access, cost to move between tiers, latency to access from tiers
- **<u>Objective</u>: explore hybrid- and multi-cloud storage strategies to optimize for the most cost-effective storage providers and schemes available**



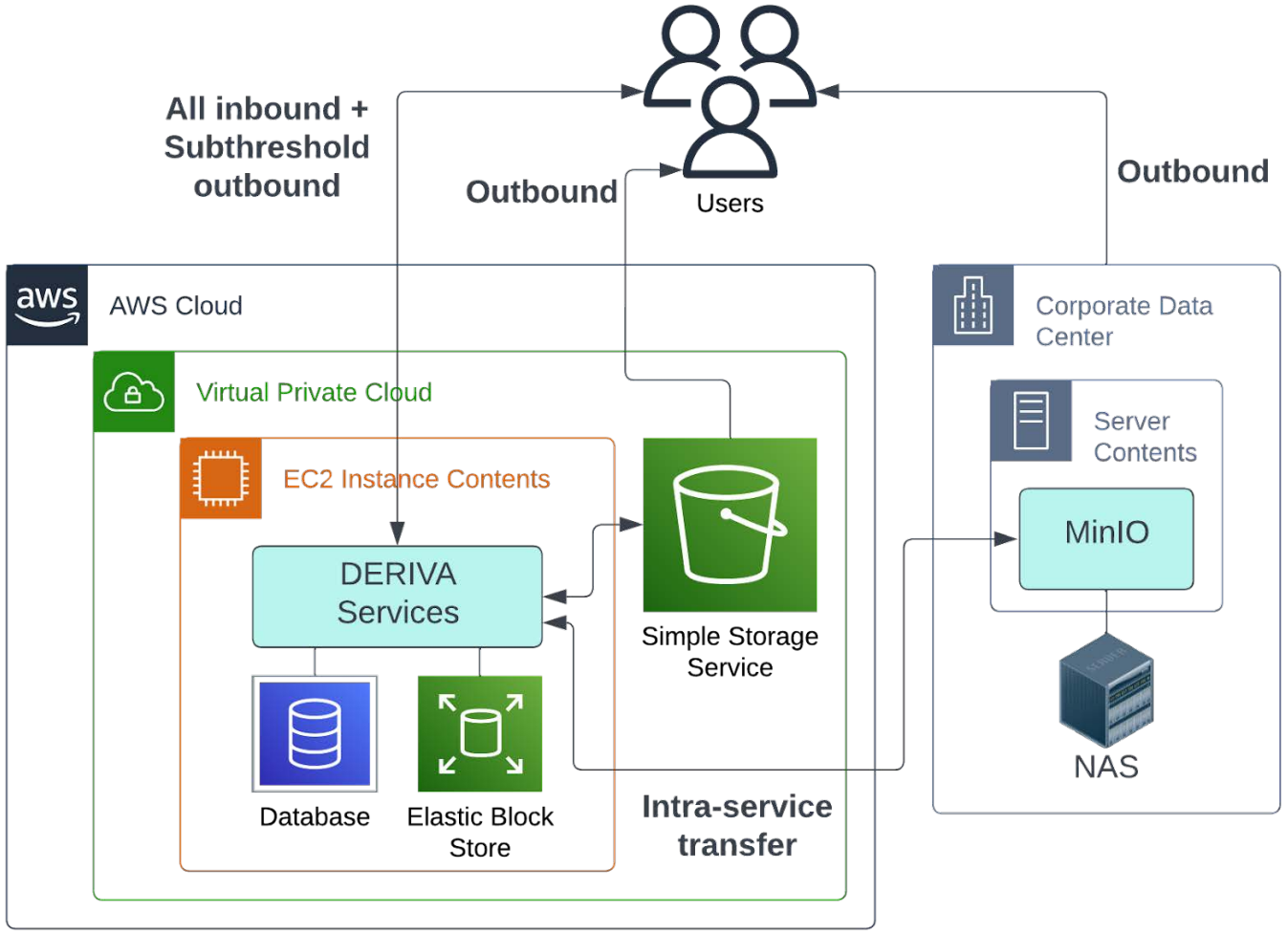
Leverage competitive storage cloud service pricing through a *multi-cloud storage strategy* (Aim 1)

- Route some data from AWS to Wasabi (low $) storage resources
- Wasabi offers low fixed monthly fee for storage based only on data volume (i.e., no egress charges) *but* egress expected to match storage volume
- Lower frequency accessed objects are better suited to Wasabi
- PoC: enhancement to DERIVA Object Storage Server to integrate multiple public/private CSPs



Utilize on-premise resources with fixed, low or no-cost through a *hybrid-cloud storage strategy* (Aim 2)

- Identify and “offload” some data from public CSPs to institute storage resources (<<$)
- Institutional resources provide “Cloud-like” storage but are unlikely to match the level of performance and availability of major CSPs
- May be suitable for any retrieval frequency but where lower availability and performance okay



Current Progress

- PoC developed: enhancements to the DERIVA object storage micro-service ("hatrac") to integrate with multiple S3-compatible back-end storage CSPs
- Test bed configuration in-progress
  - case 0 (control): all AWS
  - case 1 (multi-cloud): Wasabi storage
  - case 2 (hybrid-cloud): MinIO + institutional storage
- Experimental evaluation planned over the next months

Ability to mix-and-match storage CSPs and even on-premise resources may provide scalability characteristics to support high growth and unpredictable data retrieval patterns

Thank You

FaceBase Hub Team

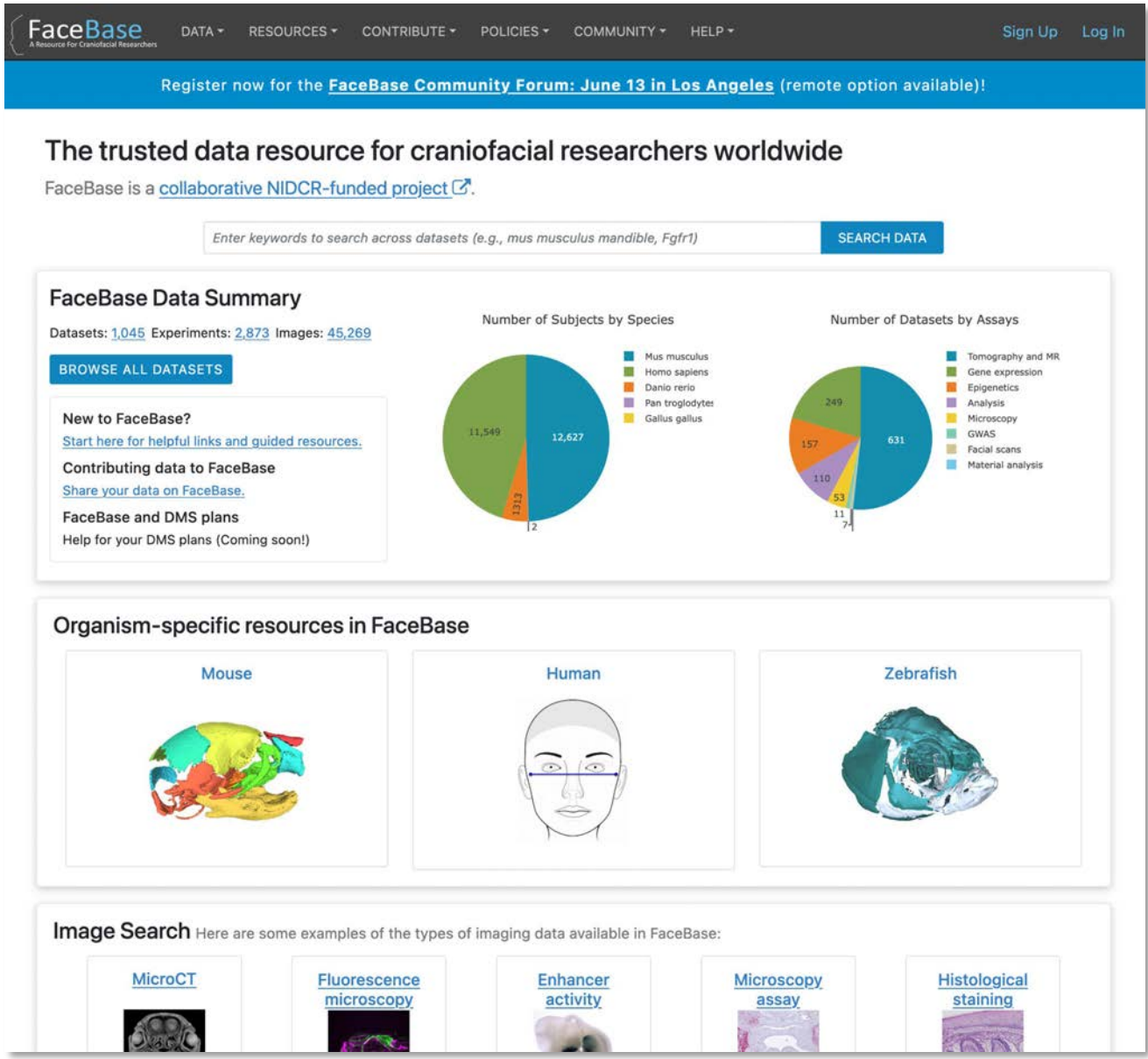
- Alejandro Bugacov
- Yang Chai (co-PI)
- Jifan Feng
- Joe Hacia
- Thach Vu Ho
- Carl Kesselman (co-PI)
- VyVy Nguyen
- Laura Pearlman
- Rob Schuler
- Cris Williams

Website: www.facebase.org

Twitter: @FaceBaseDataHub

Email: help@facebase.org

Sponsor: NIH / NIDCR (U01DE028729)



USC Viterbi Information Sciences Institute

Center for Craniofacial Molecular Biology