Pennsieve A Collaborative Platform for Translational Neuroscience and Beyond

Zack Goldblum[1,2], Zhongchuan Xu[1,2], Haoer Shi[1,2], Patryk Orzechowski[3,4], Jamaal Spence[1,5], Kathryn A Davis[1,6], Brian Litt[1,2,6], Nishant Sinha[1,2,6], Joost Wagenaar[3*]

1 **Center for Neuroengineering & Therapeutics**, *University of Pennsylvania*
2 **Department of Bioengineering**, *University of Pennsylvania*
3 **Department of Biostatistics, Epidemiology and Informatics**, *University of Pennsylvania*
4 **Department of Automatics and Robotics**, *AGH University of Krakow*
5 **Department of Neuroscience**, Perelman School of Medicine, *University of Pennsylvania*
6 **Department of Neurology**, Perelman School of Medicine, *University of Pennsylvania*
* *For correspondence, please email joostw@seas.upenn.edu*

Background

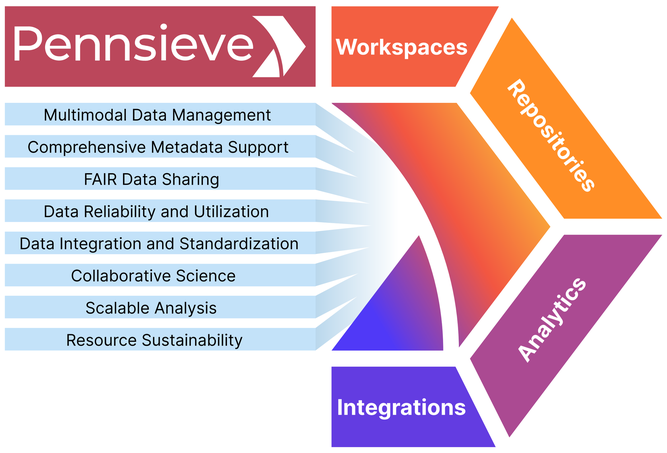
- The current neuroscience data landscape is highly fragmented and generally organized into proprietary formats and modality-specific archives[1].
- This fragmentation inadvertently creates data silos that hinder large-scale research efforts and diminish the potential for scientific discovery[2].
- As neuroscience transitions from primarily closed to open science, the adoption of FAIR (Findable, Accessible, Interoperable, and Reusable) principles has become crucial for reproducible research[3].
- **We introduce Pennsieve, an open-source, cloud-based scientific data management platform focused on integrating complex datasets, fostering collaboration, and FAIR scientific data publishing.**

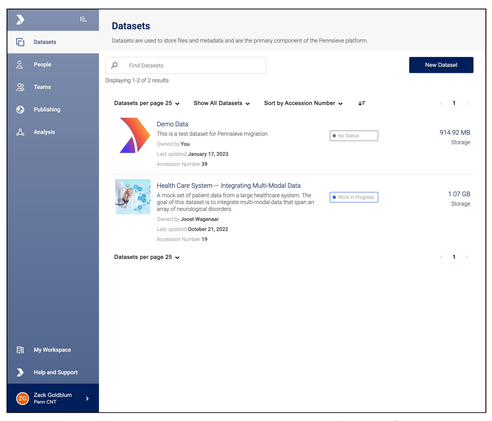
Check out Pennsieve!

Find our pre-print here: https://arxiv.org/abs/2409.10509

Methods

- Pennsieve serves as a data management infrastructure for the global scientific community.
- It is built around shared workspaces, public data repositories, scalable analytics, and integrations that facilitate collaborative research at scale.



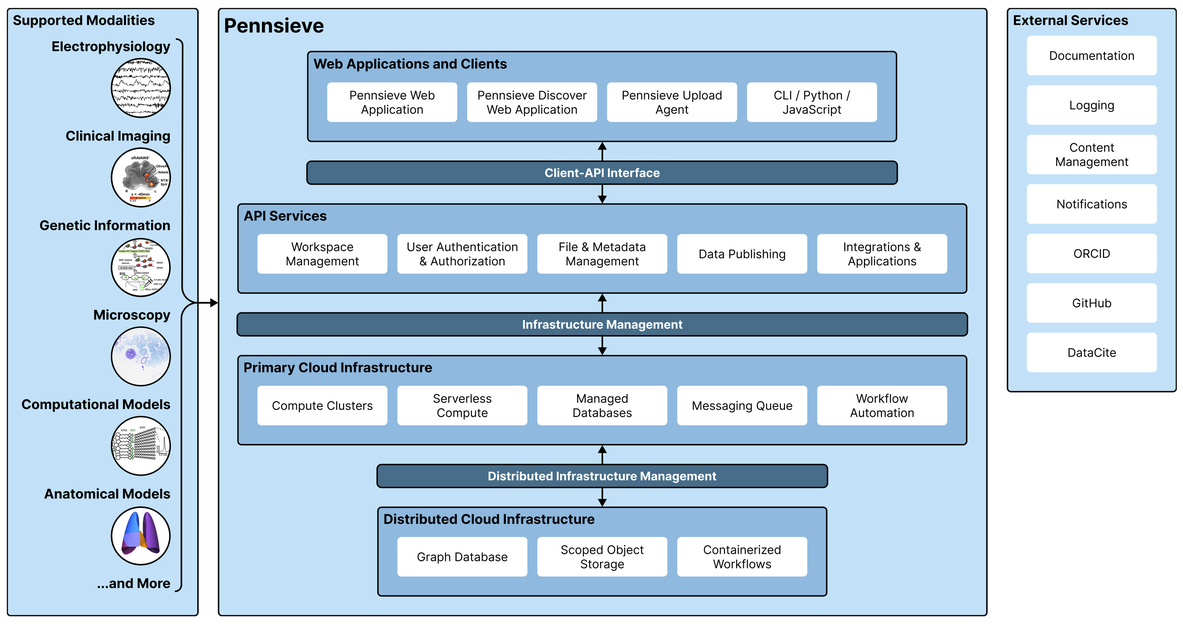


Pennsieve's web application interface.

Pennsieve Discover

Pennsieve Discover is the the platform's dedicated public data repository for FAIR scientific data publishing.

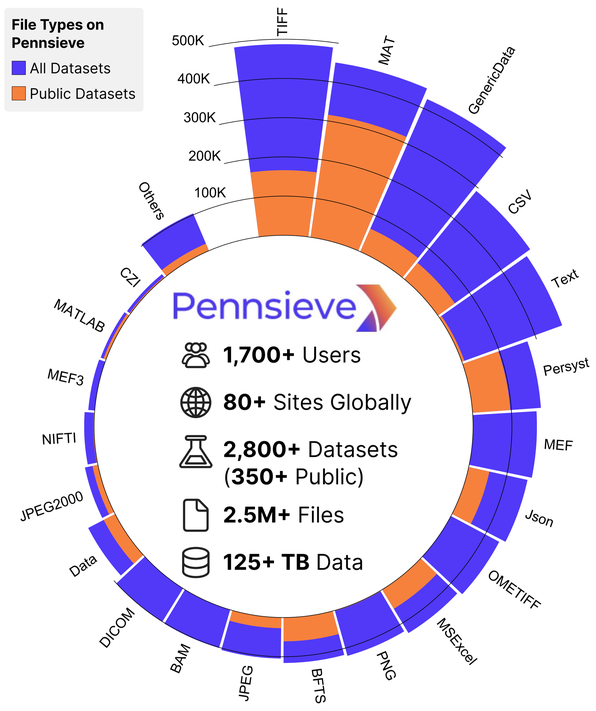
- Researchers use Pennsieve to create and manage datasets within collaborative workspaces while maintaining data privacy.
- Multimodal data management allows users to integrate diverse data types (clinical, imaging, timeseries, multiomics, and others) alongside custom metadata schemas.
- It features tools for data annotation, visualization, and analysis.
- Pennsieve supports the full data lifecycle: from initial data upload and curation to public sharing via the Pennsieve Discover portal.



- Pennsieve's multi-tenant structure supports multiple consortia through dedicated web applications and programmatically via APIs.
- Its cloud architecture ensures scalability, availability, and security.

Results & Impact

- Pennsieve stores over **125 TB** of scientific data, with **35 TB** publicly available across **350+** high-impact datasets.
- Leveraged by **100+** daily users that download over **80 GB** of data each week from Pennsieve's public services.



Pennsieve is utilized by more than **1,700** users across **80+** research sites worldwide.

- Pennsieve forms the core for several major neuroscience research programs:
  - NIH SPARC Initiative
  - NIH HEAL RE-JOIN Initiative
  - NIH HEAL Initiative's PRECISION Human Pain Network
- Supports large-scale, inter-institutional projects at clinical sites through the University of Pennsylvania.
- Also empowers individual research labs and specialized projects (e.g., Epilepsy.Science, PedQuEST, SEED Project).

Conclusion

- Pennsieve is one of the largest, fully-maintained data resources for the global neuroscience community.
- Its flexible infrastructure supports research across several scales, from individual labs to large consortia.
- Pennsieve's core focus on long-term platform and data sustainability makes it an enduring platform that can address evolving challenges in the data lifecycle across multiple scientific domains.
- Future developments will focus on supporting sustainable, distributed analytics and integrations with other platforms to expand Pennsieve's impact within the larger data ecosystem.

References & Acknowledgments

1. Martone, M. E. The past, present and future of neuroscience data sharing: a perspective on the state of practices and infrastructure for fair. Frontiers in Neuroinformatics 17, 1276407 (2024).
2. Ferguson, A. R., Nielson, J. L., Cragin, M., Bandrowski, A. & Martone, M. Big data from small data: data-sharing in the 'long tail' of neuroscience. Nature Neuroscience 17, 1442–1447 (2014).
3. Wilkinson, M. D. et al. The fair guiding principles for scientific data management and stewardship. Scientific Data 3 (2016).

Pennsieve, formerly known as Blackfynn, was built over several years, and received funding from multiple sources, including foundations, Michael J. Fox Foundation, Epilepsy Foundation, CURE Epilepsy; institutions, Penn Center for Neuroengineering and Therapeutics, Penn Institute for Immunology & Immune Health, Penn Institute for Biomedical Informatics, and Children's Hospital of Philadelphia; as well as federal funding from National Institutes of Health, grants no. 1U24NS135547, 1U24NS134536, OT3OD025347, OT3OD025347, UC4DK112217, R44DA044929, K01ES025436, and U24NS063930; and Defense Advanced Research Projects Agency, grants no. DARPA-D2-1801, and DARPA-BAA-15-35. N.S. received funding from the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health, awards no. K99NS138680, R01NS116504, and R01NS125137, as well as the Department of Defense, grant no. W81XWH2210593. In particular, Pennsieve would like to acknowledge leadership at NINDS and NIH Common Fund: Walter Korshetz, Vicki Whittemore, Randall Stewart, Andrew Weitz, Felicia Qashu, and others, who seeded the vision for Pennsieve and its predecessor, IEEG.org, through their funding, guidance, and encouragement