PROTEOBASE: a free, sustainable public proteomics repository enabling flexible metadata capture and data reuse

PROTEOBASE
PUBLIC PROTEOMICS DATA. ACCESSIBLE. REUSABLE. SUSTAINABLE.

Kyla Teplitz[1]; Yevhen Panko[1]; Dmitry Sergeev[1]; Nikolay Galanov[1]; Alex Serebryian[1], Brian Connolly[2]; Michael J. MacCoss[2]; Peter Askovich[1]

[1]Europa Dx, Redmond, WA; [2]University of Washington, Seattle, WA

UNIVERSITY of WASHINGTON

EuropaDX

1 THE CHALLENGE

Why proteomics data reuse is difficult

Incomplete and inconsistent metadata and uncertain sustainability limit interpretability and reuse of mass spectrometry datasets.

Missing or inconsistent file-level metadata
Raw files are often uploaded without clear association to samples, methods, or experimental context.

Details buried in manuscripts
Experimental information is scattered across text and supplementary materials.

Uncertain accessibility
When repositories lose funding datasets and literature links can become inaccessible.

Better metadata + sustainable stewardship = reproducible proteomics research

2 PROTEOBASE WORKFLOW

From raw data to reusable public resource

UPLOAD
- Raw vendor files & .mzML
- Processed files
- Metadata tables
- Any study assets

ANNOTATE
- File-level metadata entry
- Attach metadata files
- AI-assisted metadata suggestions

REVIEW
- Researcher validation & editing
- Private collaboration
- Anonymous reviewer links

PUBLISH
- Public release under **CC0 or CC BY 4.0**
- Persistent DOI
- Public link sharing

REUSE
- Searchable datasets & metadata
- Facilitate data reproducibility and reuse

Advanced cloud reanalysis available through EuropaXp OmicsCloud.

EuropaXp OMICS CLOUD

3 AI-ASSISTED ANNOTATION

Reduce manual entry while maintaining author control

1 Select Annotation Assistant
From dataset object, open file annotation manager and select "Import with AI".

2 Upload Supporting Files
Upload full-text manuscript, .csv files, or other documents with experiment information.

3 Automated Extraction
Proprietary AI extracts potential experimental annotations from uploaded files.

File	Sample	Organism	Method	...
File1.raw	Sample A	Human	DIA	...
File2.raw	Sample B	Human	DDA	...
File3.raw	Sample C	Yeast	DIA	...
...	...	...	...	...

4 Suggested Metadata
Extracted annotations are mapped to files and presented for review.

5 Researcher Review and Edit
Researchers review, edit, and approve all metadata prior to publication.

AI-Assisted, human-reviewed metadata improves completeness and interpretability of datasets.

4 SUSTAINABLE STEWARDSHIP

Preserving community resources for the long term

Chorus
The Chorus Project
(formerly chorusproject.org)
Public datasets and literature links established over many years.

PROTEOBASE
New home for legacy datasets with continued public access and preserved links.

Public Accessibility
Datasets remain freely accessible to the community.

Continuity
Literature citations and dataset links remain valid.

Long-term Stewardship
EuropaDX is committed to long-term funding and maintenance of PROTEOBASE.

Sustainable stewardship ensures public proteomics data remains available for the future.

5 PRELIMINARY OUTCOMES & CONCLUSIONS

Preliminary deployment outcomes

- ✓ End-to-end submission, annotation, and publication workflows operational.
- ✓ Improved explicit association of metadata with individual files.
- ✓ Reduced manual annotation burden through AI-assisted workflow.
- ✓ Preservation and continued accessibility of legacy Chorus datasets.
- ✓ Flexible metadata-attachment workflows supporting data reuse.

Conclusions

PROTEOBASE improves metadata quality and ensures long-term accessibility of public proteomics datasets.

By combining flexible metadata workflows, manuscript-assisted annotation, and sustainable stewardship of legacy resources, PROTEOBASE strengthens reproducible proteomics research.

Access PROTEOBASE
proteobase.org

METADATA SUPPORT

PROTEOBASE supports association of structured metadata files alongside datasets. Users may attach metadata in community formats such as SDRF and other file-based standards.

.SDRF

Examples of supported metadata file types:
- SDRF (.txt, .tsv, .csv)
- Sample tables
- Experiment designs
- Instrument methods

PROTEOBASE USER INTERFACE (Examples)

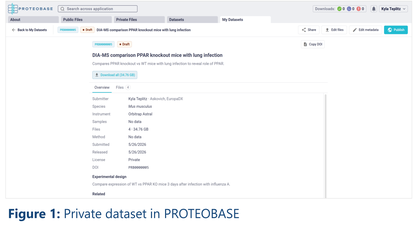


Figure 1: Private dataset in PROTEOBASE

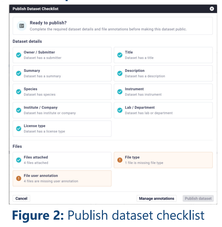


Figure 2: Publish dataset checklist

ABOUT PROTEOBASE

PROTEOBASE is a free public repository for proteomics data. It is built and maintained by EuropaDX, Inc. to support open science, reproducibility, and long-term access to proteomics datasets.

proteobase.org

CONFLICT OF INTEREST DISCLOSURE

The MacCoss lab receives support from Agilent, Sciex, Shimadzu, ThermoFisher, and Water to support the development and maintenance of Proteowizard and Skyline for the mass spectrometry community.

Michael J. MacCoss is on the scientific advisory board for EuropaDX, Inc.

EuropaDX, Inc. (booth 501) stewards and finances PROTEOBASE.

Contact: info@proteobase.org