



How do resource managers use metrics to articulate the size, impact and scope of their resource, and the stakeholders of the resource?

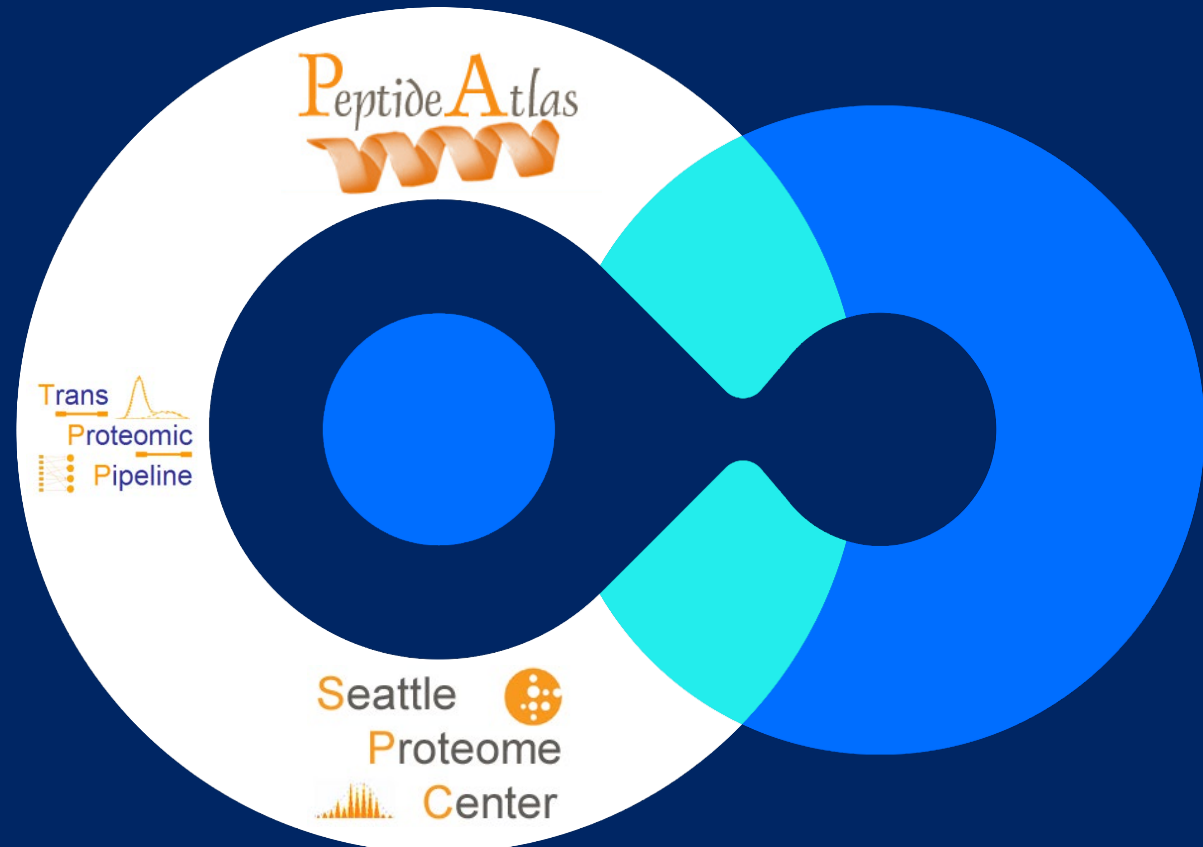
Professor Robert L. Moritz,
Institute for Systems Biology, Seattle, WA, USA
Faculty and Head of Proteomics Research



rmoritz@systemsbiology.org



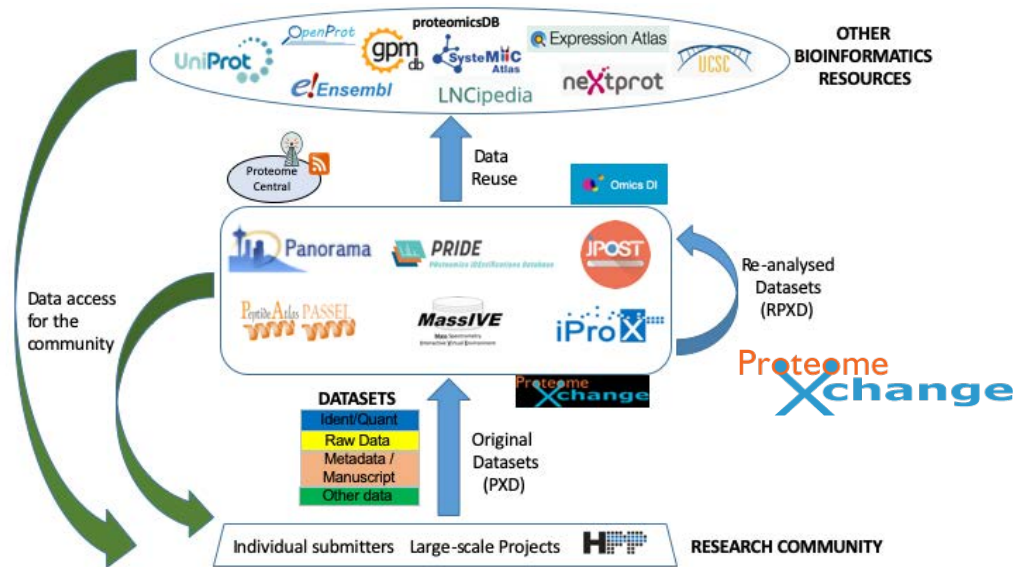
[@r_l_moritz](https://twitter.com/r_l_moritz)



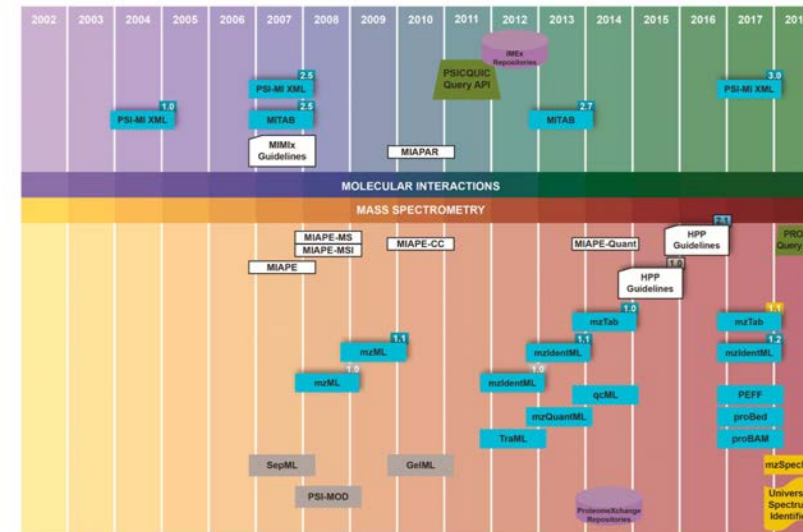
About me

- Professor, Faculty member Institute for Systems Biology since 2008
- Career started in protein analysis at the Ludwig Inst Cancer Research in 1983
- Moritz lab at ISB is a dual Wet chemistry/Computational biology lab
- Developed the Australian Proteomics Society in 1995
- Past world Human Proteome Organization (HUPO) Vice president
- Current Co-chair Human Proteome Project (HPP-HUPO)
- Developed the worlds first web accessible free proteomics computational facility to serve Proteomics data analysis (Australasian Proteomics Computational Facility)
- Moritz Lab develops high-quality well accessed resources in mass spectrometry-based protein analysis from the Trans-Proteomic Pipeline to the PeptideAtlas suite of online resources endorsed by the Human Proteome Organization as the Protein primary resource

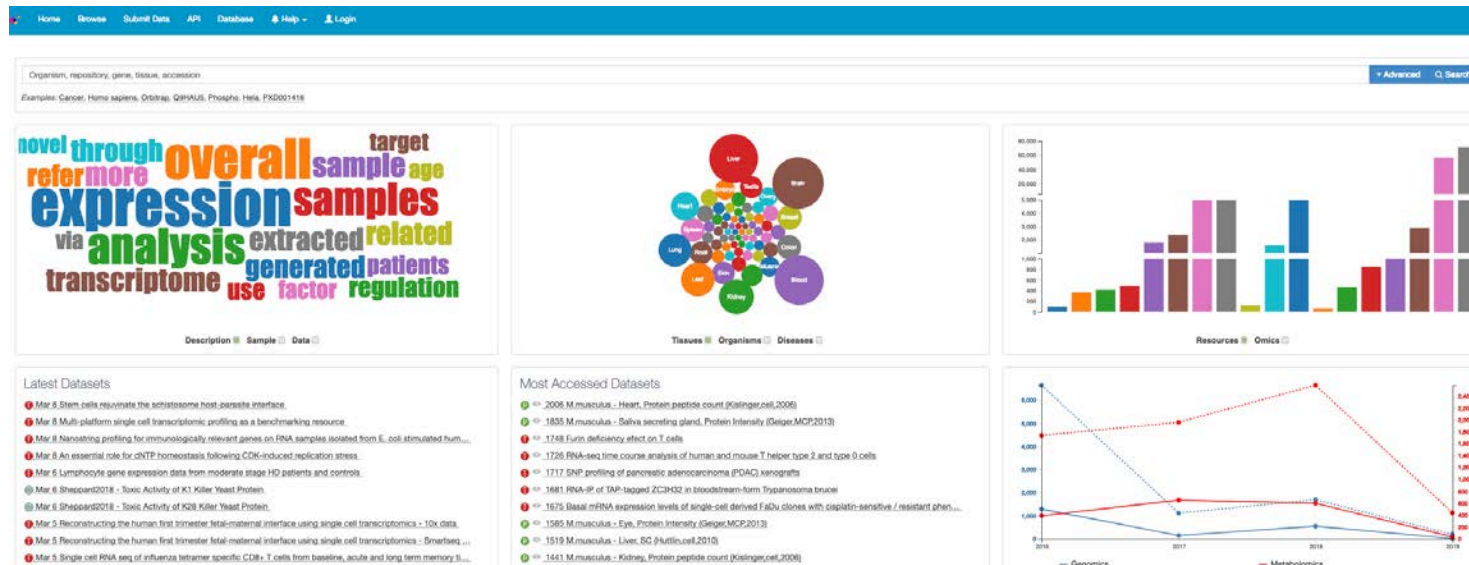
As part of a greater community



ProteomeXchange consortium of proteomics data repositories links data producers and resources that reuse the data



HUPO Proteomics Standards Initiative – developing data formats and data standards since 2002



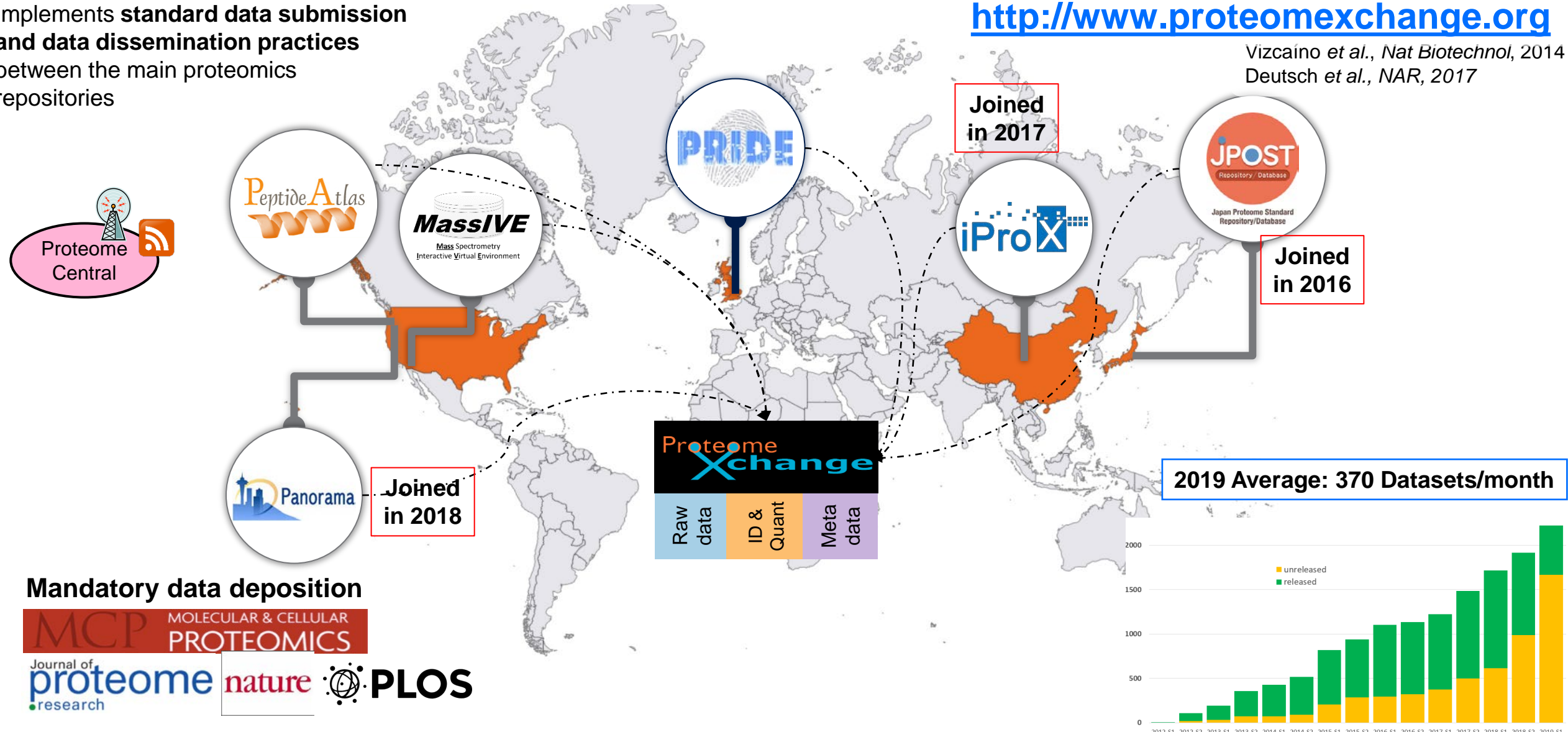
OmicsDI.org aggregates metadata from proteomics, metabolomics, transcriptomics datasets all at one place

ProteomeXchange: A global consortium of proteomics repositories

Implements **standard data submission**
and data dissemination practices
between the main proteomics
repositories

<http://www.proteomexchange.org>

Vizcaino et al., *Nat Biotechnol*, 2014
Deutsch et al., *NAR*, 2017



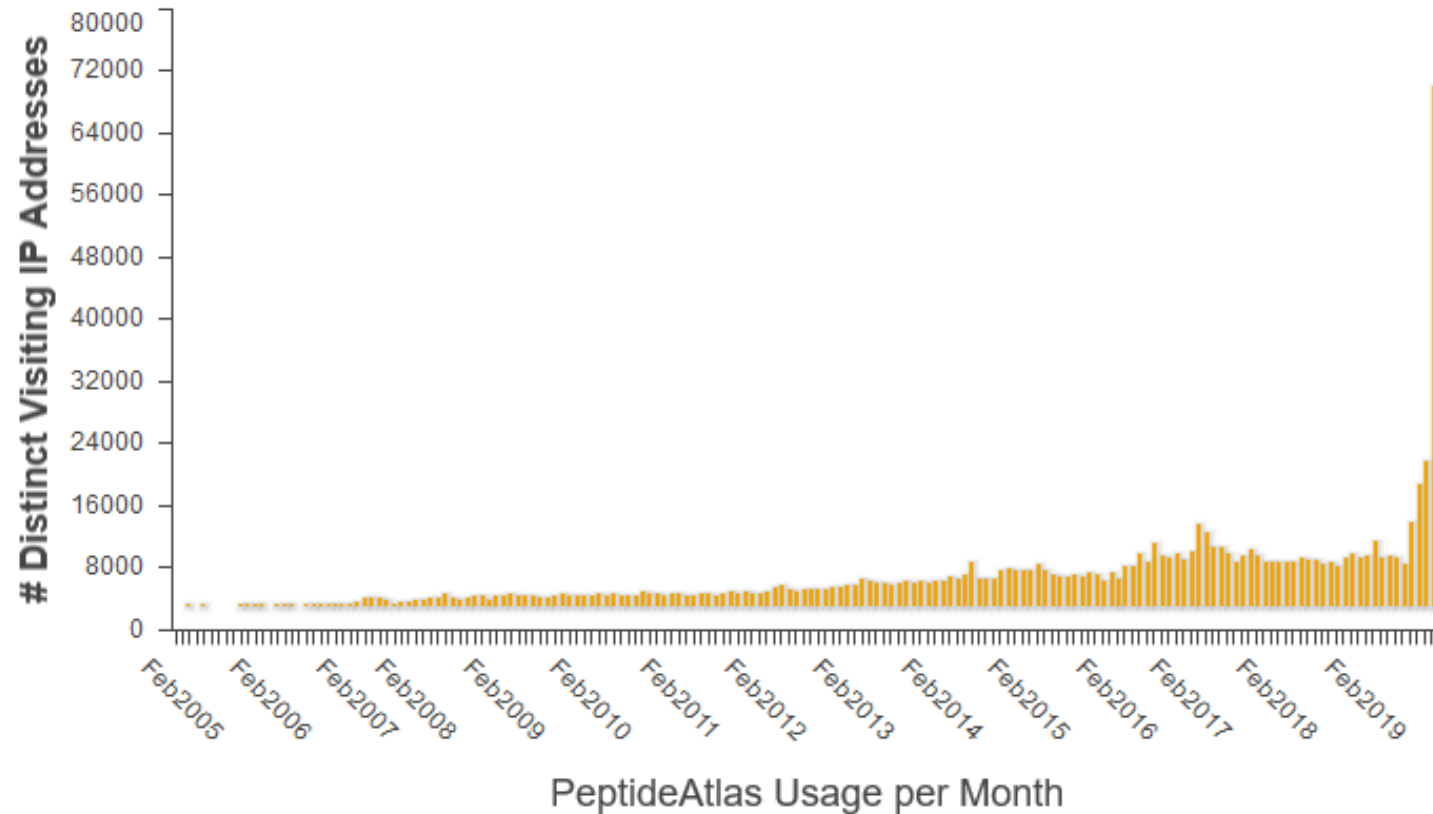
Metrics – Number of users

- Probably the best metric if we could really know it
- This is fine for data submitters (often undercounts because of a lab designated submitter)
- But requiring registration and login for all access hinders quick casual use, may hinder search engine indexing
- Users are becoming more sensitive to and stay away from user-tracking tactics. They don't want to be tracked!



- Using IP addresses as a proxy for number of users is perilous:
 - All users at an institution behind a firewall appear as one very busy user
 - One user can launch hundreds of cloud instances that access a resource
 - API keys can help, but users don't want to be tracked!

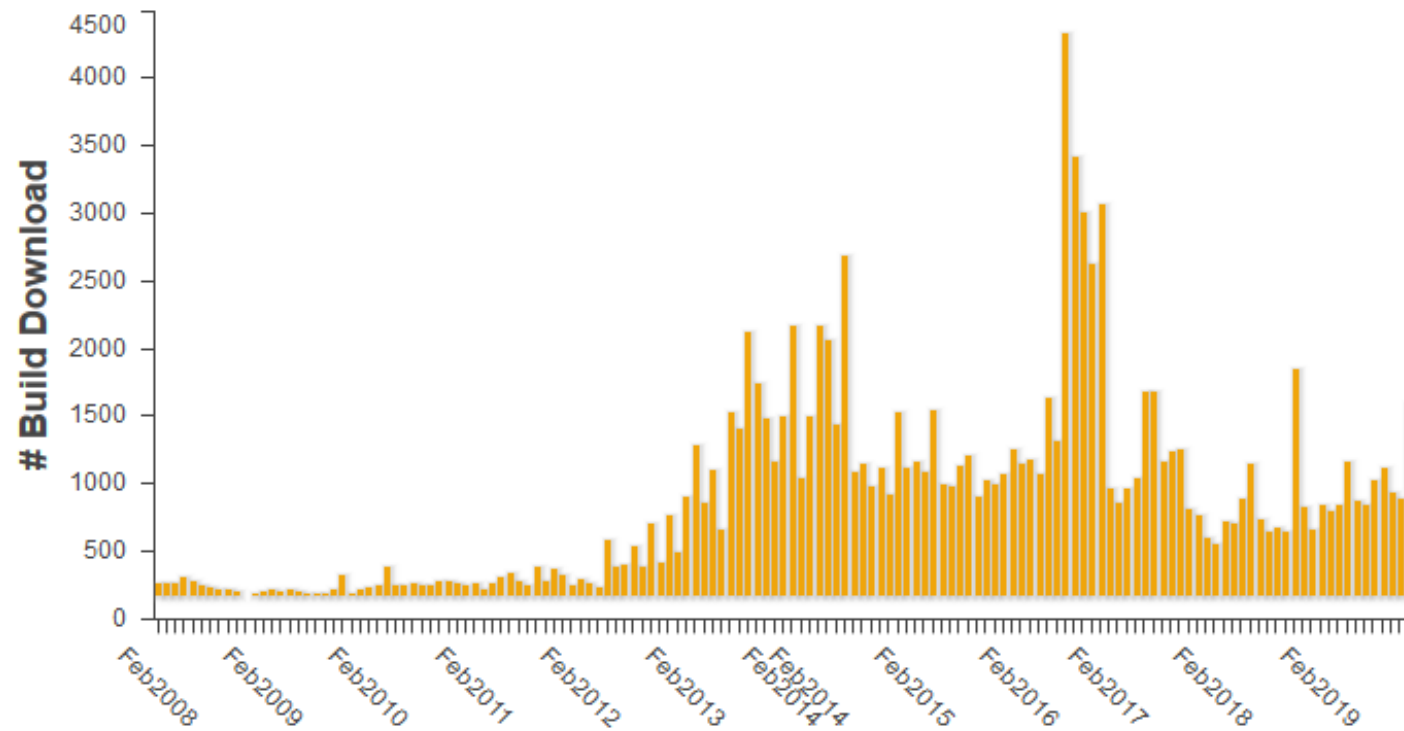
Metrics – Number of visiting IP addresses



Number of distinct visiting IP addresses per month (known web crawlers removed)

Dramatic rise in the number of distinct IP addresses hitting PeptideAtlas web pages and web services starting second half of 2019. Mostly unresolved IP numbers. Scripts running on cloud computing instances?

Metrics – Number of downloads



Number of build downloads
per month

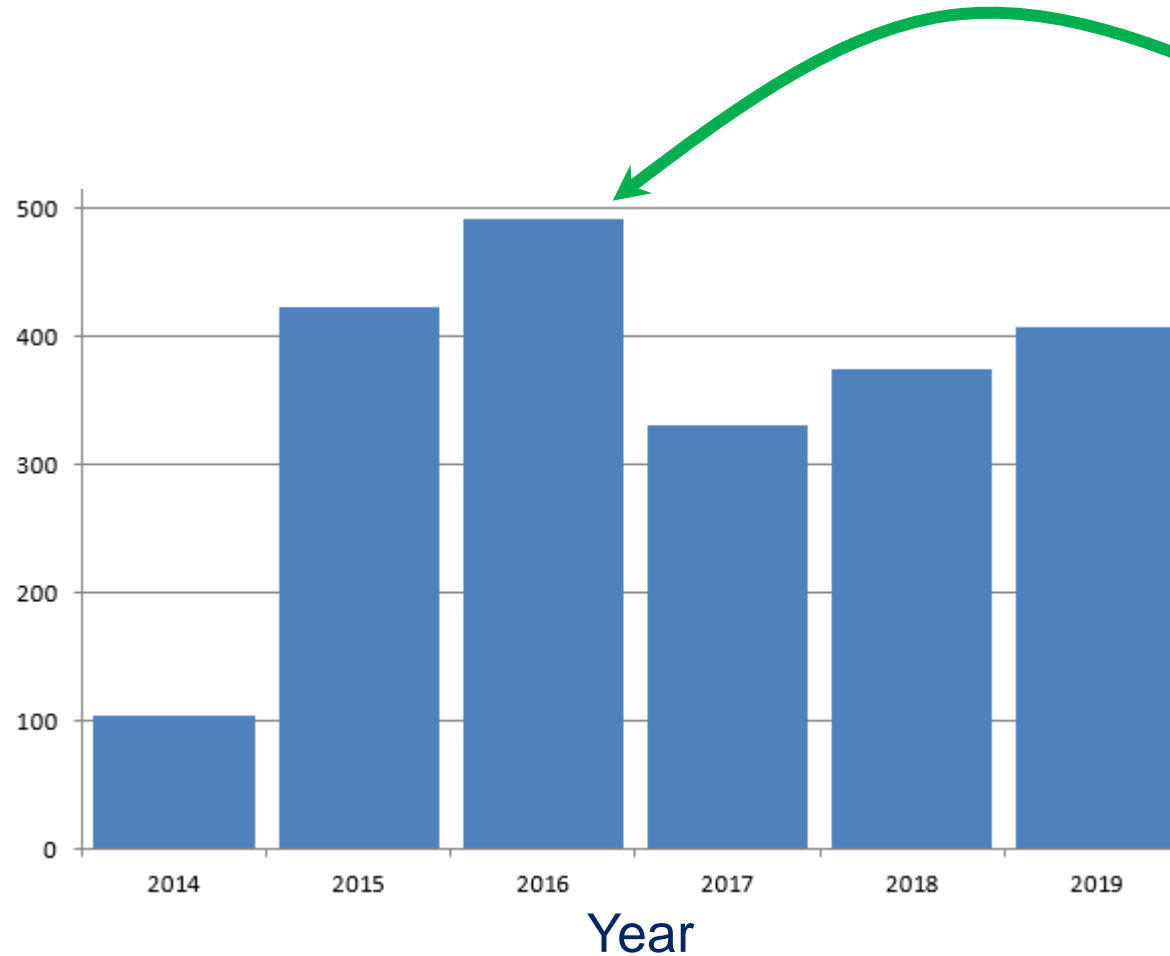
PeptideAtlas Builds Downloads per Month

Number of downloads of data products provided a clearer picture of who is using and reusing the data we produce. But the availability and convenience of web services means less downloading entire builds, and more fetching just the information you want.

Metrics – Number of citations

- Often authors forget to cite all the resources they used, even if mentioned

ProteomeXchange
citations per year



Novelty of a new resource gets lots of citations at first, but then soon only the very diligent still cite a common resource

Metrics – Multiple strategies

OmicsDI tries to give recognition to the dataset producers, not the repositories. This fosters more public data dissemination.

Yasset Perez-Riverol



Email: yperez@ebi.ac.uk
Affiliation: EMBL-EBI
ORCID: [0000-0001-6579-6941](https://orcid.org/0000-0001-6579-6941)
Public profile: <https://www.omicsdi.org/profile/xQuOBTAW>
Home page: <https://orcid.org/0000-0001-6579-6941>

I'm a Project Leader of Multiomics at the EMBL-European Bioinformatics Institute (Hinxton, Cambridge, UK). I earned undergraduate degrees in Software Engineer (2006) and a doctoral degree in Biochemistry (2013) from the University of Havana. After finishing my PhD in Havana he joined the PRIDE team in 2014. I have lead several development projects such as PRIDE Inspector Toolsuite, and Omics Discovery Index a major resource to find, discovery and link omics datasets.

Datasets

Filter yper@ebi.ac.uk's datasets ...

In-depth Analysis of Protein Inference Algorithms using a Workflow Framework and Well-Defined Metrics - yeast Swiss-Prot (equal to proteome) analysis

This dataset is no actual new study but the results of the benchmark of the Yeast Gold Dataset used in the manuscript "In-depth Analysis of Protein Inference Algorithms using a Workflow Framework and Well-Defined Metrics".

2016-08-10 | [PXD003066](#) | [Pride](#)

[reanalysis](#) [yeast](#) [Technical](#)

[Cite](#)

PIA - iPRG2008 Benchmark Dataset

This dataset is no actual new study but the iPRG2008 benchmark dataset used in the PIA manuscript.

2015-05-08 | [PXD000793](#) | [Pride](#)

[mouse](#) [iPRG2008](#) [benchmark](#) [Technical](#) [Reference](#)

[Cite](#)

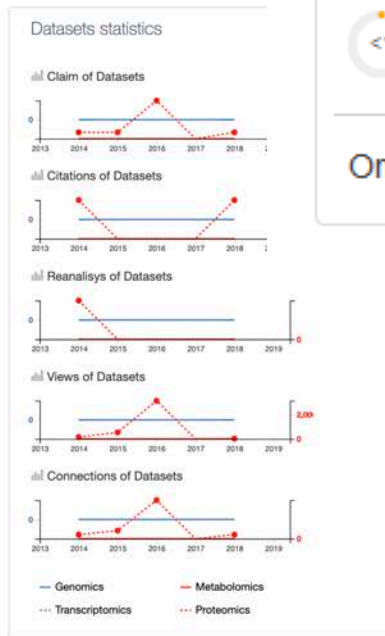
In-depth Analysis of Protein Inference Algorithms using a Workflow Framework and Well-Defined Metrics - PXD000603 Swiss-Prot analysis

This dataset is no actual new study but the results of the benchmark against the PXD000603 used in the manuscript "In-depth Analysis of Protein Inference Algorithms using a Workflow Framework and Well-Defined Metrics".

2016-08-10 | [PXD003068](#) | [Pride](#)

[reanalysis](#) [human](#) [Technical](#)

[Screenshot](#) [Cite](#)



Feedbacks

Comments Community Login

Recommend Sort by Best

Sc 1 51 Views size 10

53 9028 Connections


4 20 Citations

1 1 Reanalyses

<1 2 Downloads

Omics score: **599**

[Cite](#)



Conclusions

- Number of users is probably the best metric
- But is very problematic to obtain an accurate number unless one puts up a very elaborate system to track users. And users do not like that.
- Tracking multiple other metrics is nice, but is it really useful?
- Dare we say: NIH/reviewers don't seem to care to provide funding for keeping a well-used resource going; they want to fund new resources and ideas
- Continued funding of resources appears to hinge on perceived novelty and value of NEW features (with some minimal evidence of past use helpful)
- Research community expects to have high quality online resources accessible without cost
- Sustained development and maintenance costs of online resources are not compatible with community expectations