

# Present and future of proteomics data curation at the PRIDE database

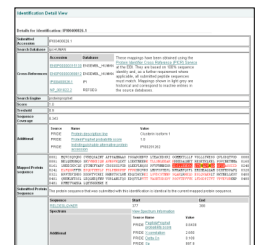
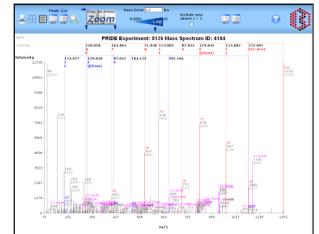
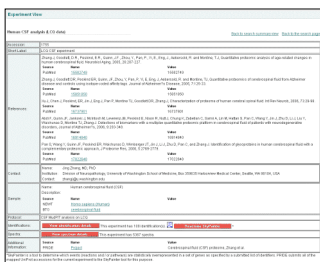
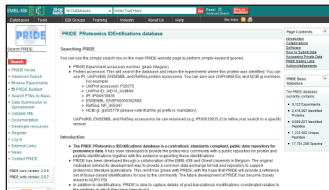
*Juan A. Vizcaíno, Henning Hermjakob and Lennart Martens*



The **PRIDE database**<sup>1-3</sup> provides a publicly accessible, centralized data repository for **mass spectrometry (MS)** based proteomics data. PRIDE stores protein and peptide identifications, spectra and valuable metadata.

PRIDE allows data to remain private during peer review, while allowing optional access to collaborators, journal editors, and peer reviewers.

PRIDE is a recommended platform for data sharing and dissemination by several journals<sup>4,5</sup>.

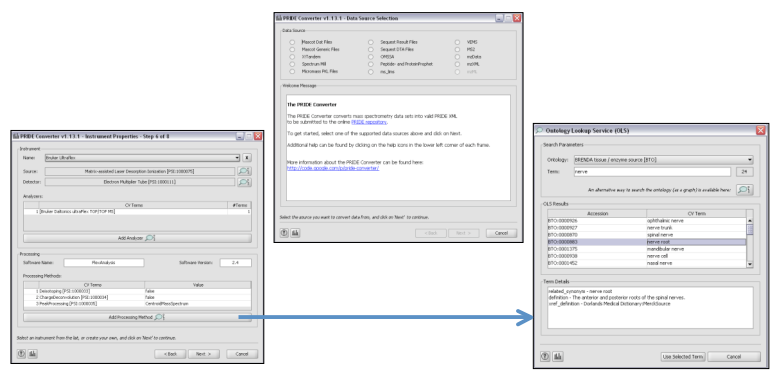


PRIDE relies on an XML-based data format for submissions (**PRIDE XML**), which is built around the HUPO-PSI mzData standard for MS

Due to the heterogeneity of MS derived data, conversion to PRIDE XML is far from trivial

The free and platform independent **PRIDE Converter** tool allows the conversion of a variety of data formats into PRIDE XML in 8 simple steps.

<http://code.google.com/p/pride-converter/>



PRIDE Converter uses The Ontology Lookup Service (**OLS**) to annotate and enrich the metadata

Data curation in PRIDE will be significantly extended. **High-quality data** will be included in a new repository called **PRIDE-Q**



...but, what is **MS high-quality data** ?

We will have to find out...,  
(some research is needed)  
(minimal requirement rules, ...)  
..., but we will do it in a completely open way for the scientific community

If you want to take part: Open position. W/09/035/EBI. Deadline: 26/04/2009.

## References:

1. Martens L et al. Proteomics 2005; 5: 3537-3545.
2. Jones P et al. Nucleic Acids Res. 2006; 34: D659-63.
3. Jones P et al. Nucleic Acids Res. 2008; 36: D878-83.
4. Editors. Nat. Biotechnol. 2007; 25: 262.
5. Editors. Nature Methods 2008; 5: 209-209

EMBL- EBI  
Wellcome Trust Genome Campus  
Hinxton ,Cambridge  
CB10 1SD  
UK

