- The amount of publicly available data in mass spectrometry (MS) proteomics repositories like PRIDE, PeptideAtlas, GPMDB, MaxQB, ProteomicsDB, HPM and others, as well as data from third-party global PRIDE reprocessing, is growing at a fast rate.

- The Universal Protein Resource (UniProt, http://www.uniprot.org) is integrating proteomics data from all available public sources to provide comprehensive protein data sets with high quality experimentally proven evidences.

### UniProt protein existence (PE) values

http://www.uniprot.org/manual/protein_existence

#### pipeline to update PE values

#### MS-proteomics repositories filtered lists of peptides

#### Data availability

Protomics mapping files are publicly available since UniProt release 2015_03 (http://www.uniprot.org/help/2015_03/ftp/releases/) from a dedicated section on the Downloads page of the UniProt website (http://www.uniprot.org/downloads). Which points to the UniProt ftp where statistics and a readme file are also provided. Mappings are currently available for ten species coming from two MS proteomics repositories. More species will be added as soon as additional MS proteomics repositories will be integrated.

- Develop a second iteration of the pipeline addressing the UniProtKB sequences which are not belonging to the reference proteomes.
- Incorporate and release to the public (either in our mapping results and into a standard format suitable for MS proteomics search engines) the increasing amount of variation data (ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/variation/)

### Next developments

- All filtered repository peptides, not only the ones matching to the in-silico digested UniProt sequences.
- Share with gene build developers/providers/curators to improve build algorithms.