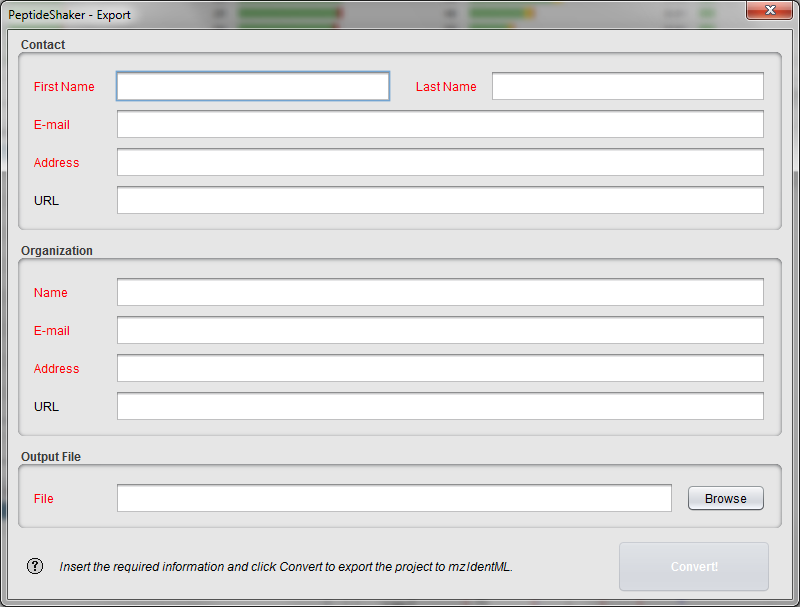
Submitting to Online Repositories

In this section, we will generate an mzIdentML1 file from the previously used HeLa dataset. Load the example project into PeptideShaker2.

Click on ‘Save As...’ in the ‘File’ menu and select ‘Export to PRIDE’. The following dialog appears:



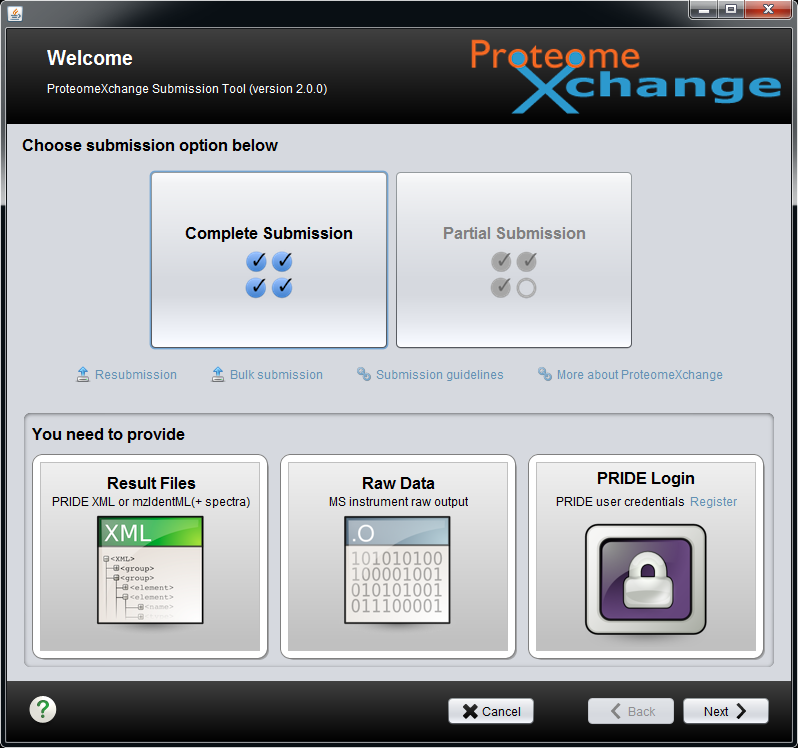
The information inserted here will be used to annotate your dataset in PRIDE3.

*Why is it important to annotate our data? How to find the proper annotation?* *[3.1a]*

Selecting an output file and clicking on the 'Convert!' button will start the creation of the mzIdentML file. In order to save time, the corresponding file has already been generated for you and is located in the resources folder.

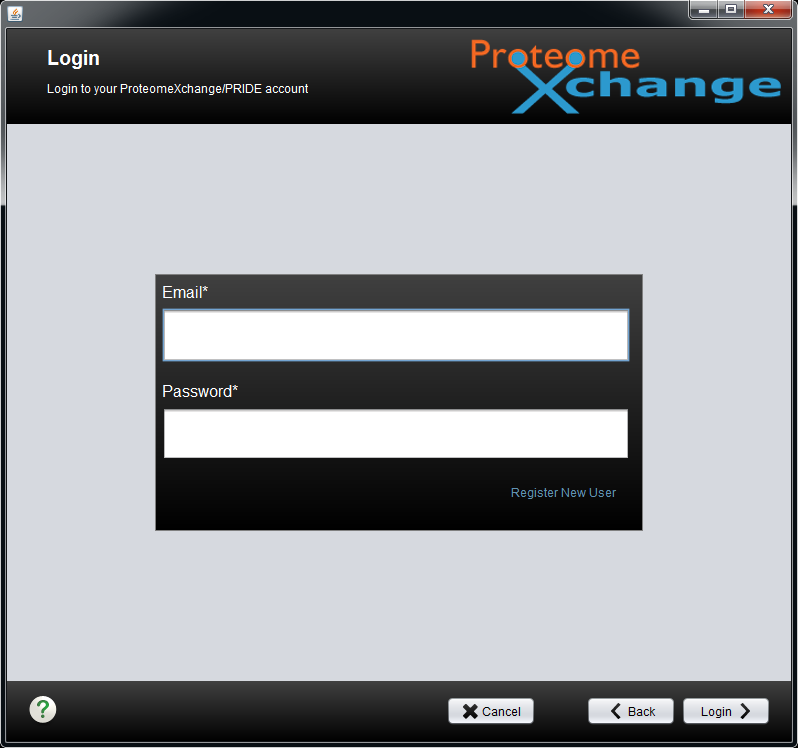
Start the ProteomeXchange4 submission tool located in the software folder (PX\_Submission). The tool can also be downloaded or launched directly from <http://www.proteomexchange.org> after clicking on “Submit Data”.

You should now see the following screen:



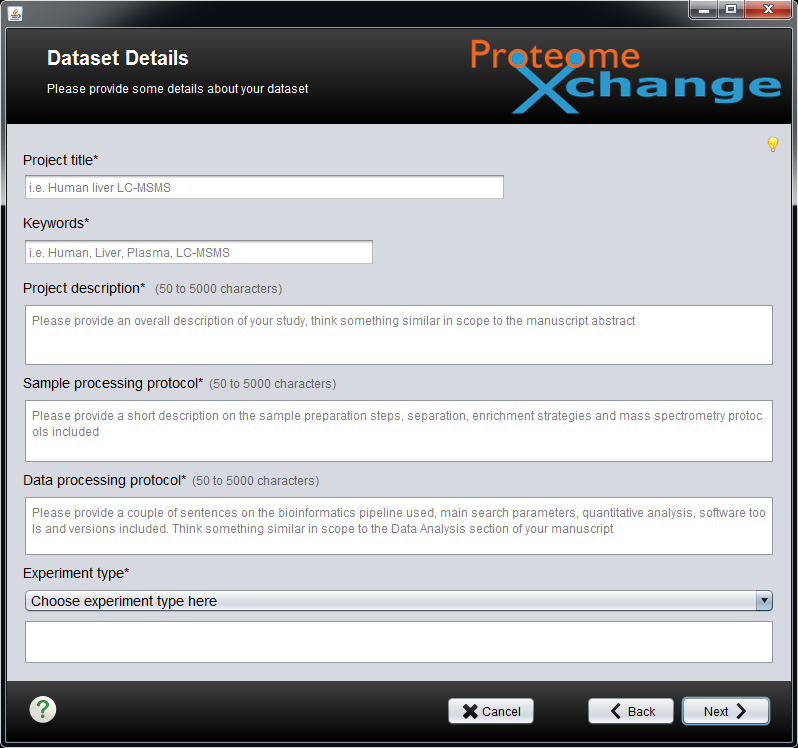
Note that a complete tutorial of the PX submission tool is available in the document folder downloaded with the submission tool. The following is an example on the data we generated.

Click on ‘Complete Submission’ then ‘Next’. You will be asked to register using your PRIDE credentials:



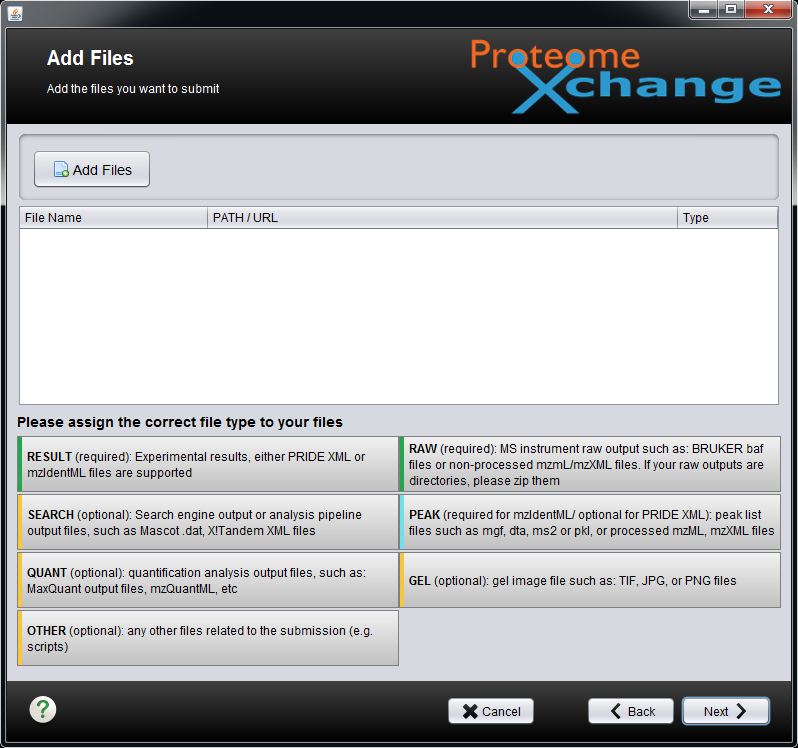
As a test account you can use: "pridestudent@ebi.ac.uk" and "student01". *But remember to get your own account when submitting your own data!*

You should now see the following dialog:



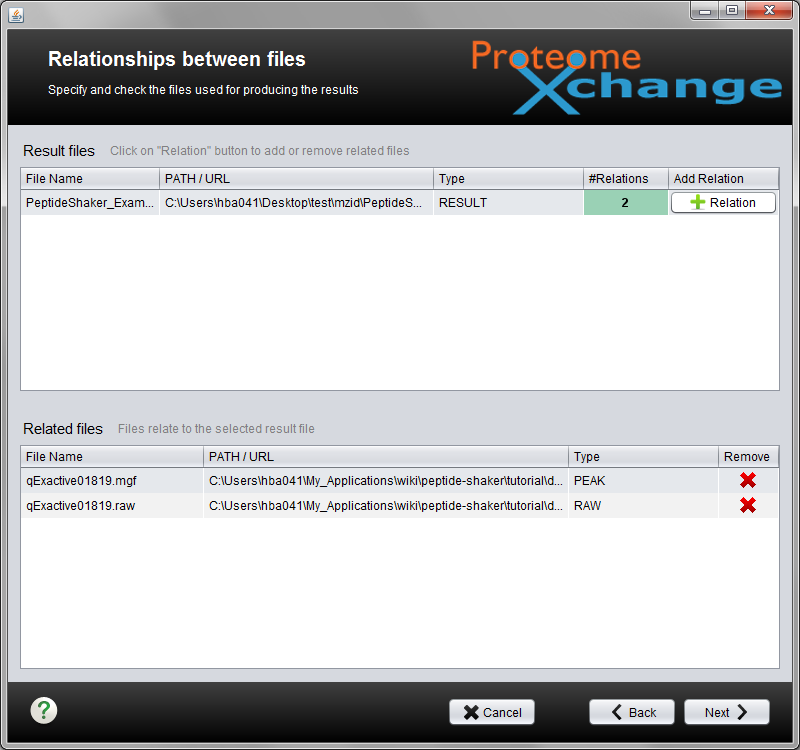
Here you are asked to fill the description of your project. While the annotation conducted during the conversion was standardized for automatic recognition, here the input is user oriented. You can thus simply reuse parts of your manuscript.

After clicking on “Next”, you will be able to load your files:



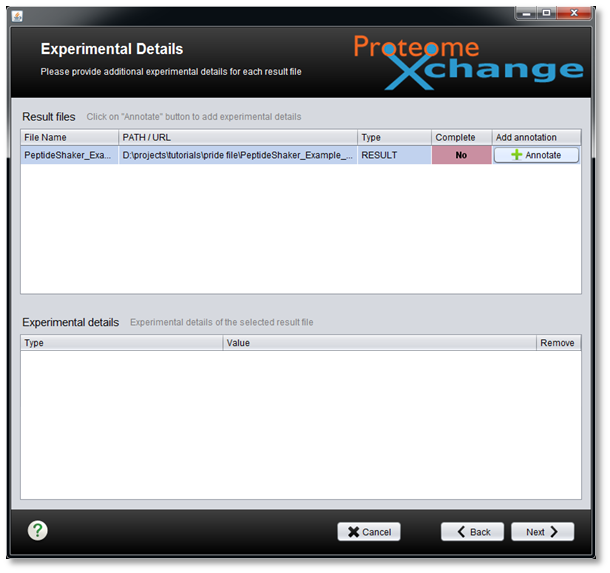
Here, load 1) the raw file (available in the resources folder), 2) the mgf file and 3) the mzIdentML file we just created. Note that the submission tool recognizes the different file formats.

In the next step, we will indicate that the mzIdentML file results from the processing of the raw file. Add a relation between the raw file and the mzIdentML file. You should see the following:

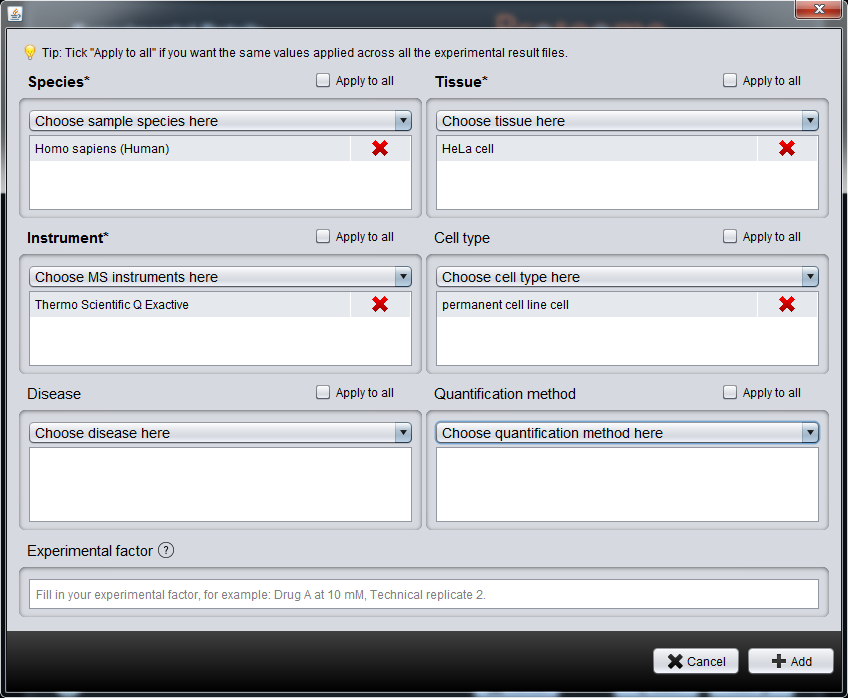


*In which case will you have different mzIdentML files with different relations? [3.1b]*

After clicking on “Next”, you will be asked to annotate your result files:

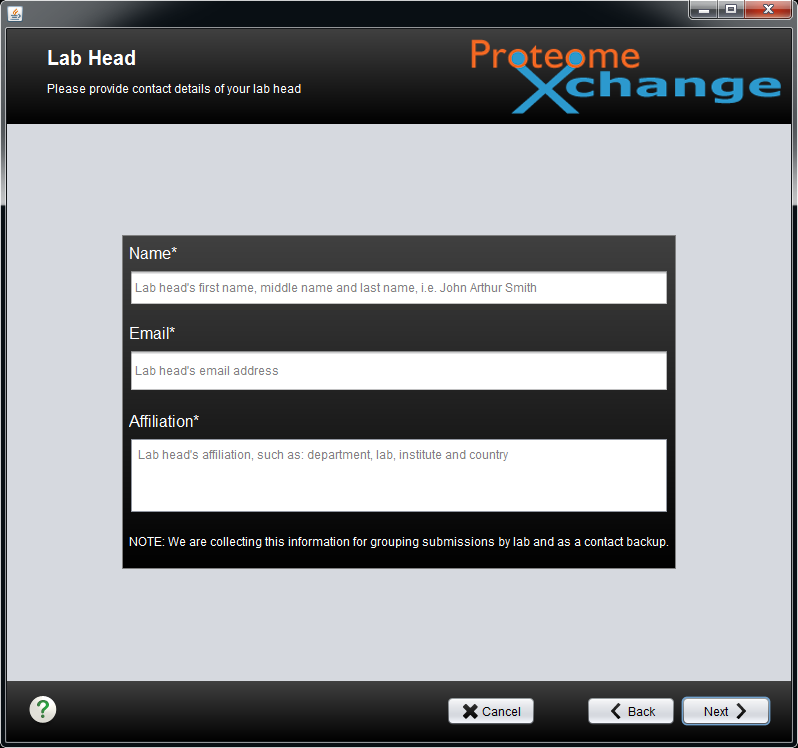


Clicking on “Annotate” will allow you to add more information. Note that annotating species, tissue and instrument is mandatory. Select ”homo sapiens” as the species, ”HeLa cell” as the tissue type, ”Thermo Scientific Q Exactive” as the instrument and ”permanent cell line cell” as the cell type:

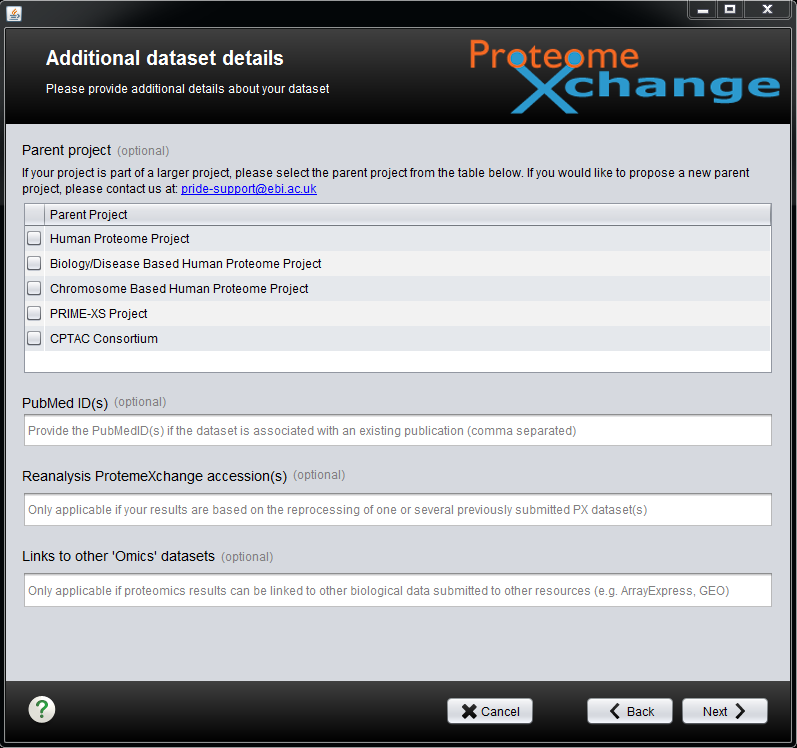


*What else can we add? [3.1c]*

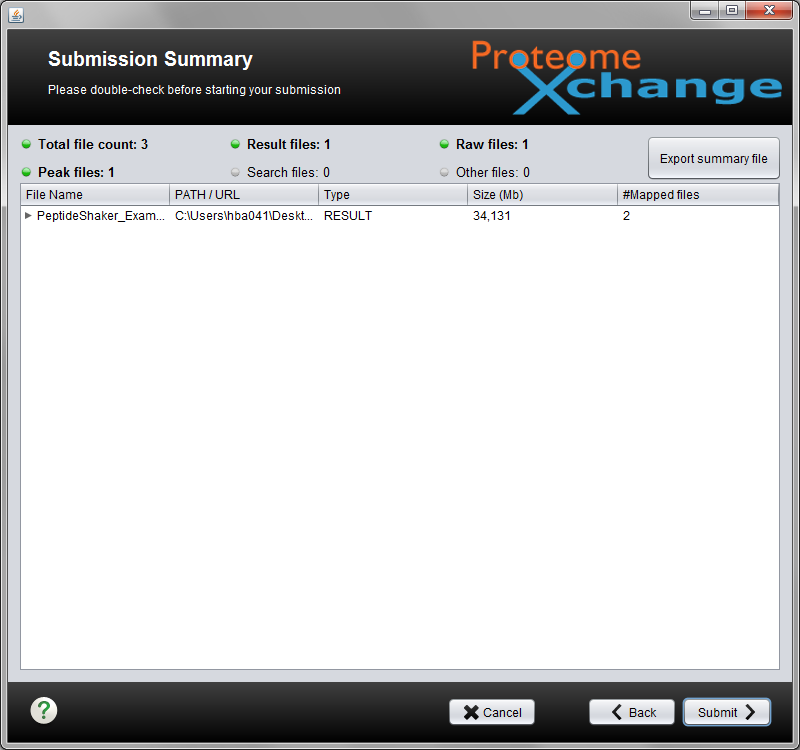
Finally, you will be asked to provide the lab head contact details:



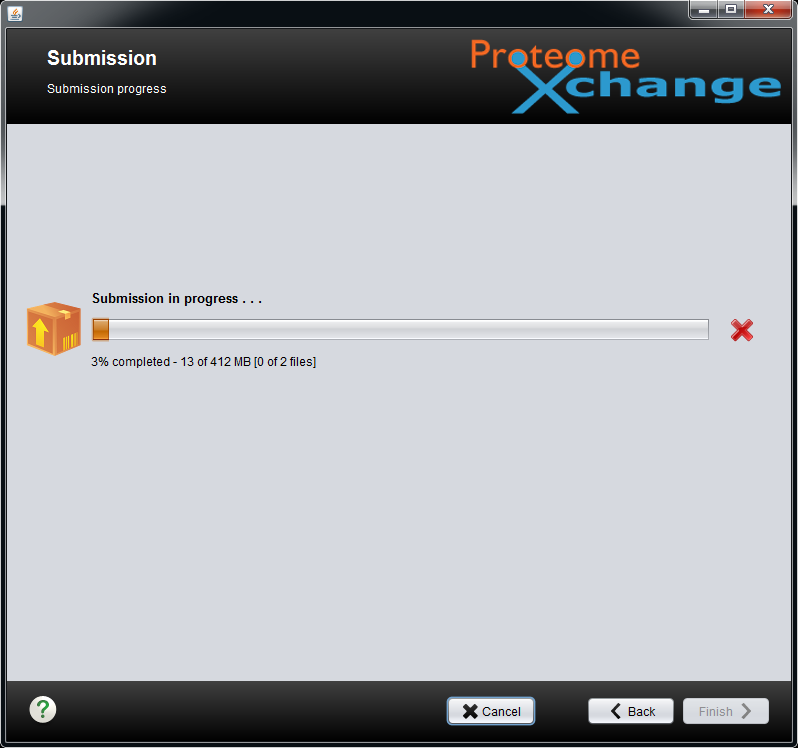
Next, you can link your identification results to parent projects, other datasets or scientific publications:



When all these steps are completed, a summary of your submission is displayed:



The tutorial data is already online so we will **not** upload it again, but clicking on “Submit” would have started the upload automatically.



Once the upload is finished, the PRIDE team will contact you at the address used to log in and will provide you with the accession details of your dataset. Note that your dataset will be kept private until you notify the PRIDE team that it should be made publicly available.

*What is the difference between a public and a private dataset? Can I give a reviewer access to my data? [3.1d]*

References

1. Jones, A.R. et al. The mzIdentML data standard for mass spectrometry-based proteomics results. *Mol Cell Proteomics* **11**, M111 014381 (2012).

2. Vaudel, M. et al. PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat Biotech* **33**, 22-24 (2015).

3. Martens, L. et al. PRIDE: the proteomics identifications database. *Proteomics* **5**, 3537-3545 (2005).

4. Juan A Vizcaíno et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol* **32**, 223–226 (2014).