ProteomeXchange submissions via PRIDE

Juan Antonio Vizcaino [1]
Proteins
Beginner
1 hour

How to submit MS/MS data to ProteomeXchange via PRIDE. This course is aimed at proteomics researchers with data that they want to make available through public-domain databases of proteomics experiments.

Learning objectives:

- Have a basic knowledge of the resources PRIDE and PeptideAtlas (including the PASSEL component)
- Be able to submit mass spectrometry data to the ProteomeXchange consortium via PRIDE
- Be able to confidently use the ProteomeXchange submission tool
- Be aware of the complete and partial submission pipelines

What is ProteomeXchange?

ProteomeXchange (PX) provides a standard way for submitting mass spectrometry [2]-based proteomics [3] data to public-domain repositories. Once you have submitted your data to the PX entry point, it will be automatically disseminated to all other repositories in the consortium [1].

ProteomeXchange simplifies the submission and distribution of proteomics data.

It has a centralised infrastructure that allows:

- Data submission to a mass spectrometry [4]-based proteomics data repository (PRIDE [5] for tandem mass spectrometry [6] (MS/MS) data, PASSEL for selected reaction monitoring [7] (SRM) data);
- Private access to data in the repository during the review process (for journal editors and reviewers);
- Data dissemination to the other members of the consortium once the corresponding manuscript is published;
- A straightforward way for all proteomics researchers to keep up to date with new available datasets via the ProteomeXchange RSS feed [8].

Some scientific journals (including Proteomics [9], Molecular and Cellular Proteomics [10] and several journals from the Nature Publishing Group) recommend submission to proteomics repositories such as PRIDE.

Data resources in ProteomeXchange
In the ProteomeXchange (PX) consortium there are currently two kinds of proteomics [3] data resources:

**Archival resources** contain processed data [11] as published by the authors. PRIDE [12] is representative of this type of data in the consortium for tandem mass spectrometry [2] (MS/MS) datasets. The PASSEL component of PeptideAtlas [13] is representative for selected reaction monitoring [14] (SRM) datasets (Figure 1).

**Secondary data resources** build on the primary data provided by submitters and are stored in archival resources. PeptideAtlas [13] and UniProt [15] are examples of secondary data resources that are members of the PX consortium (Figure 1).

After submission to an archival resource such as PRIDE, once they are made publicly available, files are propagated to the secondary data resources, from where they are made available to users via an RSS feed (Figure 1).

![Diagram of ProteomeXchange workflow](image)

**Figure 1.** Basic workflow for MS/MS data in ProteomeXchange. *Raw data* [16] means mass spectrometer output files [17]. Figure taken from reference 1.

In addition to the databases, ProteomeCentral [18] is the resource that generates a unique identifier [19] for each ProteomeXchange dataset and, also, it constitutes a registry for all ProteomeXchange submissions (irrespective of the receiving repository). This queryable archive provides the users with an efficient way to identify datasets of interest. For instance, it is a way to monitor the re-use of particular datasets, and give an efficient way to monitor the volume and impact of the ProteomeXchange data exchange.

Next we will explore all the resources individually.
What is PRIDE?

The PRIDE [20] (PRoteomics IDEntifications) database at the European Bioinformatics Institute (EBI) is a centralised, standards-compliant public database containing MS-based proteomics data. It was originally developed to provide the proteomics community with a public repository for peptide and protein identifications, together with the evidence supporting these identifications [2, 21, 22].

PRIDE [5] is the initial submission point in the ProteomeXchange consortium for MS/MS data (Figure 2).
Figure 2 The PRIDE Archive web interface as of Feb 2014.

What does PRIDE store?

PRIDE [12] aims to represent the submitter's original view of the data. PRIDE stores:

- **Processed identification results** [22]: peptide and protein identifications, including post-translational modifications (PTMs);
- **Mass spectrometer output files** [17]: both raw data [16] and peak list [23] files;
- Quantification values: peptide and protein expression values (if available);
- **Metadata** [24]: technical and biological metadata;
- Other files related to one submission: gel images, scripts, etc.

Why do we need PRIDE?

The PRIDE [12] database is one of the main public repositories for proteomics [3] data that have been generated by MS approaches. In addition, PRIDE is leading the ProteomeXchange consortium.

Various journals in the field strongly support, and some even mandate, deposition of MS-based proteomics data into a public proteomics repository such as PRIDE (or others from the ProteomeXchange consortium). This enables researchers to understand, and potentially reproduce,
experiments described in a particular publication. It also allows you to reanalyse mass spectra data as protein sequence databases and search-engine toolkits improve.

In addition, PRIDE provides proteomics data to other data resources such as UniProt [25].

What are PeptideAtlas and PASSEL?

The PeptideAtlas [26] project, at the Institute of Systems Biology (ISB, Seattle, USA), annotates genome sequences of different organisms with peptides and proteins derived mainly from MS/MS data.

All data in PeptideAtlas are reprocessed using the Trans Proteomics Pipeline [27] (TPP). PeptideAtlas is organised into species-specific or sample-specific ‘builds’, which represent all peptides mapped to a single reference Ensembl [28] genome.

What is PASSEL?

The Peptide Atlas SRM Experiment Library (PASSEL [29]) is a component of the PeptideAtlas project that is designed to allow submission, dissemination, and reuse of selected reaction monitoring [14] (SRM) experimental results from analysis of biological samples.

Figure 3. The PASSEL web interface.

The raw data [16] are automatically processed in a uniform manner and the results are stored in a database. They can be downloaded or browsed via a web interface that includes a chromatogram viewer (Figure 3).

PASSEL allows the cross-analysis of SRM data, supports optimisation of SRM data collection, and
facilitates the review of SRM data.

PASSEL is the initial submission point in the ProteomeXchange consortium for SRM data.

In this course, we focus on submission of MS/MS data to PRIDE. For help submitting SRM data to the ProteomeXchange consortium via PASSEL, we recommend that you visit the PeptideAtlas website [30].

An extra resource: ProteomeCentral

ProteomeCentral [18] (Figure 4) is an extra component of the consortium pipeline. It generates a unique identifier [19] for each ProteomeXchange data submission and acts as a registry for all ProteomeXchange submissions. This queryable archive provides an efficient way to identify and monitor public datasets of interest (for example, you can monitor the re-use of particular datasets). ProteomeCentral provides an efficient way to monitor the volume of data submitted.

![ProteomeCentral](image)

**Figure 4.** The ProteomeCentral website as of February 2014.

ProteomeCentral allows you to:

- Search original PX submissions (PXD identifiers) in participating repositories by experimental and biological metadata [24];
- Search for reprocessed datasets (RPXD identifiers);
- Easily query the archive and identify datasets of interest;
ProteomeXchange submissions via PRIDE

- Have a centralised resource to disseminate published data from the archival repository (PRIDE for MS/MS data and PASSEL for SRM data) to the community via an [RSS message](#) linked to a PX XML file.

- The PX XML file contains essential metadata information about the datasets and how to retrieve files that are part of a PX submission.

- The PX XML messages available in ProteomeCentral allow secondary data resources to evaluate and integrate data.

**How to submit data to ProteomeXchange via PRIDE**

In order to submit your MS/MS data to ProteomeXchange via [PRIDE](#), you should use the [ProteomeXchange submission tool](#).

There are two pipelines for submitting to PX depending on the supporting information available for each submission:

- Complete submission
- Partial submission

The PX submission tool can be used in both cases. For larger datasets, it is possible to use a command line alternative which is based in a faster file transfer system called [Aspera](#) (see details [here](#)).

Each MS/MS data submission needs to include the following mandatory data types:

- **Metadata** [24] - Data that add context to the submitted experiment, such as what species the sample came from, are required as part of your submission to PRIDE/PX because they enable users to search for your experiments on the basis of the metadata;
- Peptide/protein identifications (processed results);
- **Mass spectrometer output files** [17] ([raw data](#)).

Once you have submitted your data, they can be held privately, allowing reviewers and journal editors access if desired, or until you choose to release the data to the public.

Next we will explore both pipelines.

**Proteomics data formats**

Proteomics data is available in a variety of formats, the ones used by Pride and ProteomeXchange are defined here:

<table>
<thead>
<tr>
<th>File name</th>
<th>File content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mass spectrometry output files ('Raw' data)</td>
<td>This is the data and metadata generated by mass spectrometers. The data may be the original profile mode or some basic processing, such as centroiding.</td>
</tr>
</tbody>
</table>
They may be available as mass spectrometer binary output files, or as peak list spectra in a standardised format (see below).

It is important that all the scans generated contain applicable metadata.

<table>
<thead>
<tr>
<th>Standardised MS data formats</th>
<th>Three MS data formats used in proteomics:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mzXML - developed at the Institute of Systems Biology (ISB)</td>
</tr>
<tr>
<td></td>
<td>mzData - (now obsolete) originally developed by the HUPO Proteomics Standards Initiative (PSI)</td>
</tr>
<tr>
<td></td>
<td>mzML - successor to the others (developed by ISB and PSI)</td>
</tr>
</tbody>
</table>

These data formats can be used to represent processed peak lists, as well as raw data. In addition to the mass spectra, they contain detailed metadata that gives context to the information.

| Processed peak lists | Heavily processed form of mass spectrometry data, usually derived from raw data files via various (semi-) automatic steps, e.g.: centroiding, deisotoping and charge deconvolution. These files are formatted in plain text, with typical formats like dta, pkl, ms2 or mgf. |

| Protein/peptide identifications | Proteomics mass spectra can be matched to peptides or proteins, resulting in identifications for those spectra. A protein match qualifies against an *a priori* or *a posteriori* defined threshold. In the case of fragmentation spectra, the initial identification will consist of a peptide sequence; subsequent steps will derive a list of proteins from the identified peptides. This is a discernible process with its own input and output files, or it can be implicit in the overall identification software. |

| Search engine output files | These files contain the data and metadata generated by the software (called search engines) used for performing the identification and quantification of peptides and proteins. The output is either plain text or XML. |

| mzIdentML [36] | provides a common format for the export of identification results from any search engine. |

To allow a full representation of the processed data in the PRIDE database and in the PX tool, the search engine output files need to be converted to PRIDE XML or mzIdentML, which can now be exported from a variety of tools (see updated list here [37]).

| PRIDE Converter 2 [38] | is the tool developed by the PRIDE team to make conversion to PRIDE XML possible, but there are also other tools available. |

| Protein/peptide quantification | Protein/peptide expression values can be obtained from a MS-based proteomics experiment and the data and metadata is used for performing the quantification analysis. |
ProteomeXchange submissions via PRIDE

Quantification software output files

The data and metadata generated by exclusively the quantification analysis:

- **mzQuantML** [39] - provides a common format for the export of quantification results from any search engine.
- **mzTab** [40] - represents both identification and basic quantification results.

Metadata

A term used to describe data that provides additional information about a particular data set. This information includes how, when and where the data set was generated and what standards were used. In the proteomics context the addition of metadata such as peptide and protein identifications and quantification of their expression values gives meaning to a simple collection of mass spectra output files.

Complete submissions

Use the [PX submission tool][41] to submit your datasets to the complete submissions pipeline. You will need to provide all the raw data [16], and all applicable related metadata [24] to support your submission. All processed identification result files will need to be converted to PRIDE [12] XML [42] or mzIdentML [43] (version 1.1) format. In addition, if the submission is performed with mzIdentML files, the corresponding peak list [23] files must be provided as well.

There are different tools to convert or export files to mzIdentML or PRIDE XML [44] (Check the previous section). To allow PRIDE to keep your submission secure you will need a PRIDE login. To get a username and password you will need to register [32].
Once the submission process is complete, you will be issued with a ProteomeXchange identifier [19] (accession [45] number) and a permanent DOI (Digital Object Identifier [46]) that uniquely identifies your submission.

The submission is also manually checked by one of PRIDE's curators. We consider this to be essential to ensure quality control of complete submissions. Please check the PRIDE/ProteomeXchange tutorial [41] document for more information.

The PX submission tool creates the appropriate relationships between the different types of files included in your submission (the raw data and the results are mandatory; the others are optional). It will allow you to add extra metadata for your dataset before submitting all the files to PRIDE, thereby producing more contextual datasets (Figure 5).

The complete list of data formats supported by PRIDE can be found in Appendix 1 [47].

**Partial submissions**

Although the option for partial submissions is available, it isn't the recommended first option as it significantly reduces the reusability of your dataset. You should only use this option if your search results cannot be converted/exported to the PRIDE XML [44] or mzIdentML [43] formats.

As a result, you will be issued with a PX identifier [19] but not with a DOI. In addition your dataset
will not be fully integrated in PRIDE, but will be searchable based on metadata [24], and the corresponding files will be made available on the PRIDE FTP [48] server to download.

The partial submissions pipeline requires you to have the raw data [16] (mass spectrometer output files [17]), search engine output files [49] (output from search engine or analysis pipeline) and a PRIDE login, so you will need to register [50] (Figure 6). Other files types can be included optionally (quantification output files, peak list [23] files, scripts, gel images, etc). Your username and password will allow PRIDE to securely process your submission.

Figure 6. ProteomeXchange MS/MS data workflow for partial submissions.

The PX submission tool creates appropriate relationships between the different file types used in a partial submission, i.e. the raw data and search engine output files. For bigger submissions, a faster file transfer command-line alternative is available (see here [34]).

Metadata requirements for MS/MS submissions

Proteomics data is substantially enriched when it is accompanied by sufficient metadata. The repositories in PX are quite flexible about how much experimental metadata they will accept. However, there are some minimal requirements. Below we summarise what you must supply when submitting your data to PX. We recommend that you add extensive metadata in the interest of making your dataset as useful as possible in the future.

The following details are mandatory for any PX submission:

- Contact name and e-mail for the submission. The contact details of the data submitters need to be provided, allowing interested users to contact the original authors if desired;
- Lab Head or Principal Investigator;
- Project title of the PX dataset;
- Project description: it could be considered as the abstract information of the dataset (provided as free text);
- A summary for each of the Sample and Data Processing Protocols (provided as free text);
- Experiment type (chosen from a drop-down menu);
- Keywords: A list of keywords that describe the content and type of the experiment being submitted. Multiple entries should be comma separated;
- Sample annotation:

  - Species. At least one NEWT ontology Controlled Vocabulary (CV) term is mandatory per dataset.
  - Tissue. At least one 'BRENDA Tissue' Ontology (BTO) CV term is mandatory per dataset;
  - Instrument. At least one 'Mass Spectrometry' (MS) CV term is mandatory per dataset.

  - Quantification method (if applicable).

The following details are also optional:

- Sample annotation:

  - Cell type. Use the “Cell Type” ontology (CL);
  - Disease. Use the “Human Disease” ontology (DOID).

- Dataset details:

  (a) Your dataset is part of a bigger project/effort (for instance the Human Proteome Project or ‘PRIME-XS’). It is a way to tag your dataset to enable grouping this way;
  (b) There is already a PubMed ID associated with it (the data has been already published);
  (c) Your dataset represents a reanalysis of an earlier public PX dataset;
  (d) There are other related “omics” datasets (for instance transcriptomics, metabolomics data present in other repositories) that can be associated.
Submission to ProteomeXchange via PRIDE using the PX submission tool

ProteomeXchange submission tool

The PX submission tool is a stand alone desktop application that handles both complete and partial MS/MS proteomics submissions to ProteomeXchange (Figure 7). This tool will guide you through the whole submission process.

To run the tool, choose the download option directly from the ProteomeXchange submission tool page.

MS/MS Proteomics Submission

The ProteomeXchange submission tool assists you with the submission. You can run the tool using the download option below:

![Download button](download-button.png)

Figure 7 The PX submission tool setup page gives a download link.

Create a PRIDE account

To start your submission, you will be prompted to enter details for a PRIDE user account. If you haven't already done so, you will need to create a new PRIDE account (Figure 8):
Creating a PRIDE [12] account is free and only takes a few minutes to register. You should receive an automated confirmation email more-or-less straight away and be able to use your new account. Please contact us by email [53] if your login information is not valid after 24 hours following registration.

Figure 8 The PRIDE user account registration page.

Select your workflow

When you launch the ProteomeXchange submission tool, you will be prompted to choose between
complete submission and partial submission (Figure 9).

Figure 9 The ProteomeXchange submission tool's welcome screen.

You can choose between two main submission types depending on the availability of mzIdentML [43] or PRIDE XML [44] files as "Result" files for complete submissions. The recommended submission subtype is a complete submission, but alternatively partial submissions are accepted as well. For details about this, please see the PRIDE Help Submission section [54].

**Log-in with PRIDE account**

After choosing your submission type, you will be prompted to log in with your PRIDE [12] account (Figure 10). Please enter your registered email address and password to continue.
Figure 10 PX submission tool, PRIDE account log-in

Dataset details

Here mandatory details about the dataset are required, such as a title for your submission, experimental and data processing protocol, and experiment type (Figure 11).
ProteomeXchange submissions via PRIDE

Figure 11 PX submission tool, dataset details.

Please note that the description and protocol fields need at least 50 characters to be entered in order to continue.

Add your files

On the ‘Add files’ screen, select all the files you wish to submit and choose which file type they belong to (Figure 12). The tool will automatically guess what the type is for each file, which you can manually change. You can also remove any files you have accidentally included.

Depending on what type of submission you have made, different files will be mandatory or optional. An example of a complete submission with an mzIdentML [43] result file, along with the related peak list [23] file, and raw data [16] file is shown in the figure below.
Again, for details about the submission types, please see the PRIDE Help submission section [54].
ProteomeXchange submissions via PRIDE

Figure 12 The PX submission tool's ‘Add files’ screen. Here, files that are going to be part of the submission are selected and tagged.

For a full list of all the different file types, see Appendix 1 [55].

Create relationships between files

Raw files need to be mapped to the corresponding result (complete submission) or search engine output files [49] (partial submission). The tool will try to do this automatically depending on the file naming structure, and this can be edited manually using the "+ Relation" button for each result or search output file (Figure 13).
In Figure 14, below, an example for a complete submission is shown. In order to reach this screen, the "+ Annotate" button for the result file was clicked (similarly to the previous step for mapping relationships). A partial submission will go straight to this screen instead, after performing the mapping for relationships.

To enhance your dataset you should add as much available metadata [24] as possible.
Mandatory information needed for a complete submission: species, tissue, and instrument.

A partial submission also requires modifications to be entered [4] [21].

*Ontology* [56] terms have to be selected from each of the drop-down menus which have the most common options listed. If a required term is not present: select the bottom "Other..." option from the drop-down which will then allow you to search across various ontologies to select a desired term.

**Add Lab Head information**

Lab Head (Principal Investigator) details are required to continue (Figure 15).
The mandatory fields are: name, email, and affiliation.

**Figure 15 PX submission tool, lab head details.**

The mandatory fields are: name, email, and affiliation.

**Final checks and data submission**

Once all the relevant information has been added you will have an opportunity to do a final check before you start the submission (Figure 16).

After you click "Submit", your files will be uploaded to [PRIDE](#) [12]. This will take some time depending on how large your dataset is, and how fast your internet connection is. When all files have finished transferring to PRIDE, you will receive an e-mail including a submission reference. This is not an [accession][45] number to be quoted in manuscripts. It should be used in all communications between you (the submitter) and the PRIDE team. An automated confirmation e-mail will also be sent.
to your registered e-mail address.

Your submission will then be processed by expert biocurators in the PRIDE team, which may need to contact you in order to improve the quality of the provided metadata [24]. Once everything is corrected (if applicable), your dataset will be loaded into the PRIDE database. After this happens you will receive another automated e-mail with a unique ProteomeXchange identifier [19] (PXD accession number) and the private reviewer account details. If you had made a complete submission, you will also receive a data DOI [57].
In the top-right corner of this screen there is a button called "Export summary file". This will save all the meta-data for your dataset into a summary "submission.px" file. This is useful if making a "bulk" submission through the PX Submission tool, or a direct manual submission to PRIDE via our private Aspera/FTP server. Such manual submissions are useful when dealing with large datasets that cannot be easily managed with the PX Submission tool alone. Full details about this can be found on the PRIDE Help section.

There are a few options possible post-submission, e.g. making the dataset public. These are listed in our PRIDE Help FAQ.

Summary
• ProteomeXchange provides a standard way to submit MS-based proteomics data to public repositories.

• It is developed by the ProteomeXchange consortium, a collaboration among the groups that develop the main MS-based proteomics repositories.

• Submitting proteomics data to ProteomeXchange is a simple way of ensuring that your data is made available to the research community and disseminated to the major data proteomics data resources.

• Submission to a public repository is recommended, if not mandated, by many of the major proteomics journals as part of the publication of a paper.

• ProteomeXchange ensures that published (archival), raw and reprocessed (secondary) data are made available, disseminated and appropriately cross-linked.

• PRIDE is the initial submission point of MS/MS data.

• There are two pipelines for submission of MS/MS data: one for complete submissions (containing raw data and PRIDE XML or mzIdentML files) and one for partial submissions (containing raw data and search output files). The ProteomeXchange submission tool can be used for both routes.

• PASSEL is the initial submission point of SRM data.

Your feedback

Please tell us what you thought about this course. Your feedback is invaluable and helps us to improve our courses and thus enhance your learning experience.

Learn more

Find out more

• ProteomeXchange concept [60] - For more background knowledge on this project.

• PRIDE/ PX tutorial document [61] - Detailed information about how to perform submissions.

• PRIDE - Quick tour [62] - For a quick overview on PRIDE see the Quick tour.

Get help and support on ProteomeXchange
Where to get help and support

- More information about [ProteomeXchange](#) [63] and [PRIDE](#) [64] is available on their websites.
- [PRIDE Help](#) [58] details submitting data, generating data, and searching PRIDE. There is also an up-to-date [FAQ](#) [65].
- You can also contact the [PRIDE](#) [5] team by [email](#) [53].

References


Other recommended reading


Appendix 1: Formats supported by PRIDE for PX MS/MS submissions

a) as [raw data](#) [16]

Formats supported:
ProteomeXchange submissions via PRIDE

- mzML [71], mzXML, mzData. These files must not be heavily processed to be considered 'raw'
- Thermo .RAW, ABSCIEX .wiff, .wiff.scan, Agilent .d/, Waters .raw/
- imzML, Shimadzu .run/, Bruker .baf, .yep

All peak lists formats (mgf, dta, ms2, pkl) can be supported but they will not be considered raw data. They will be considered as 'peak list processed files' or simply 'peak'.

b) as processed identification 'results'

b.1) PRIDE XML [44]: For performing a complete submission, different search engine output files [49] need to be converted to PRIDE XML using existing tools like PRIDE Converter 2 [72] and others. Formats supported:

- Tandem XML
- OMSSA .csv.
- Mascot .dat
- Sequest Crux .txt
- SpectraST .xls
- ProteomeDiscoverer .msf files
- All accompanying peak lists formats

b.2) mzIdentML [43] (version 1.1): There are a number of tools that can export mzIdentML 1.1. Formats supported this way:

- Tandem XML (using mzidLibrary [73])
- OMSSA .csv (using mzidLibrary [73])
- Mascot .dat (direct export functionality available from Mascot 2.4)
- Sequest .out files (using the ProCon tool [74])
- ProteomeDiscoverer .msf files (using the ProCon tool [74])
- ProteinScape 2.1 (Bruker) database content (using the ProCon tool [74])
- MS-GF+ (direct export functionality available)
- Phenyx (direct export functionality available)
- Trans-Proteomic Pipeline (pep.xml files). The idConvert tool from can be downloaded from ProteoWizard, or is bundled with the TPP directly starting with version 4.6.3
- Scaffold (direct export functionality available). From version 4.0
- OpenMS output
- MIAPE MSI Extractor output (ProteoRed [75], Madrid)
- PAnalyzer output: Tool to perform protein inference analysis [76]
- Output files from Myrimatch, Pepitome (spectral library search), TagRecon and IDPicker
- All accompanying peak lists formats

c) as search engine output files
Only those data formats that cannot be converted to PRIDE XML or mzIdentML can be submitted as partial submissions.

d) quantification output files

Any format is supported.

Contributors

Juan Antonio Vizcaíno [1]

EMBL-EBI
PRIDE Group Coordinator
Juan is the Group Coordinator of EMBL-EBI's PRIDE resource for mass spectrometry proteomics data, and he has authored more than 65 papers. He joined the PRIDE team as a postdoc in 2006 and has been involved in numerous projects that span research, development of tools, and training and support for PRIDE users. Since 2011, he has been managing the ProteomeXchange consortium of proteomics resources.
He has a background in the life sciences, with a PhD in Molecular Biology from the University of Salamanca, Spain, a Masters in Microbiology and undergraduate degrees in Pharmacy and in Biochemistry.

[1]
Tobias Ternent [77]

EMBL-EBI
Bioinformatician - Hermjakob group

Tobias is a curator for the PRIDE team (2013), and previously a curator for the ArrayExpress team (2011). Before joining the EMBL-EBI, he studied a Bachelor's in Computer Science from the University of Manchester (2005), and later a Master's in BioHealth Informatics in the same department (2010) which involved collaboration with the KNH-Centre for Biomedical and Forensic Egyptology. Along the way, his working experience has included a few software companies, and in particular managing clinical trial data for AstraZeneca for translational informatics.

Mindi Sehra [78]

EMBL-EBI
Scientific Training Officer - eLearning Content Developer

Mindi Sehra is the Scientific Training Officer (eLearning) for the Outreach and Training Team at EMBL-EBI. Mindi is responsible for expanding and consolidating the EBI’s range of online training materials and monitoring and maintaining the portal, which includes investigating ways to exploit electronic technologies. Mindi completed a Genetics Degree at Sheffield University before moving into genome analysis at the Wellcome Trust Sanger Institute. She completed a MSc in Medical Genetics and Immunology at Brunel University, her thesis on the Swine Leukocyte Antigen secured her a position in the Human and Vertebrate Annotation [79] and Analysis group as a computer biologist. She then joined the UniProt [25] team at the EBI as a protein curator [80] working on automatic and manual annotation [81].

All course materials in Train online are free cultural works licensed under a Creative Commons Attribution-ShareAlike 4.0 International license

Source URL: http://www.ebi.ac.uk/training/online/course/proteomexchange-submissions-pride

Links
[1] http://www.ebi.ac.uk/training/online/trainers/juan
torial.pdf
[62] http://www.ebi.ac.uk/training/online/course/pride-quick-tour
[64] http://www.ebi.ac.uk/pride/archive/
[65] http://www.ebi.ac.uk/pride/help/archive/faq
[70] http://europepmc.org/abstract/MED/26527722
[71] http://www.ebi.ac.uk/training/online/glossary/mzml
[76] https://code.google.com/p/ehu-bio/wiki/PAnalyzer
[77] http://www.ebi.ac.uk/training/online/trainers/tobias
[78] http://www.ebi.ac.uk/training/online/trainers/mindi
[79] http://www.ebi.ac.uk/training/online/glossary/annotation
[80] http://www.ebi.ac.uk/training/online/glossary/curator
[81] http://www.ebi.ac.uk/training/online/glossary/manual-annotation