



## PROJECT FINAL REPORT

**Grant Agreement number:** 260558

**Project acronym:** PROTEOMEXCHANGE

**Project title:** International Data Exchange and Data Representation Standards for Proteomics

**Funding Scheme:** Coordination and support action

**Period covered:** from 1.1.2011 to 30.06.2014

**Name of the scientific representative of the project's co-ordinator<sup>1</sup>, Title and Organisation:** Henning Hermjakob, Team Leader Proteomics Services, EMBL-European Bioinformatics Institute

**Tel:** +44 (0)1223 494671

**Fax:** +44 (0)1223 494468

**E-mail:** [hhe@ebi.ac.uk](mailto:hhe@ebi.ac.uk)

**Project website address:** <http://www.proteomexchange.org/>

---

<sup>1</sup> Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

## 1. Final publishable summary report

### 1.1 Executive summary

Over the last few years the field of mass spectrometry proteomics has evolved into a prolific data producer. As a result, various databases that collect and redistribute the acquired data were established. This simultaneous creation of multiple repositories and databases caused confusion for data submitters and users alike. Proteomics data resources such as PRIDE (EMBL-EBI, Cambridge, UK) and PeptideAtlas (Institute for Systems Biology, Seattle, USA) had accepted data submissions and handled MS proteomics data for many years, but until the ProteomeXchange (PX) Consortium started they had been acting independently with very limited global coordination. The overall aim of the PX Consortium was to provide a common framework and infrastructure for the cooperation of proteomics resources by defining and implementing consistent, harmonized, user-friendly data deposition and exchange procedures among the members.

In the current consortium structure PRIDE is the point of submission for tandem MS/MS experiments, while PeptideAtlas provides a repository for SRM (Selected Reaction Monitoring) experiments called PASSEL. The MassIVE repository (University of California San Diego) has recently joined PX, thus demonstrating PX's unifying role in the proteomics community by inclusion of members that were not part of the initial Consortium. The main common access point of the consortium is ProteomeCentral (<http://proteomecentral.proteomexchange.org>). It provides a sufficient set of experimental and technical metadata for the datasets.

Another main focus in the consortium was the maintenance and further development of proteomics data standards. While data format standards for qualitative proteomics data had been defined and implemented, standards for quantitative proteomics were still lacking. The new standard called mzQuantML was developed in a modular way. It consists of a core and different modules adapted for the different quantitative techniques. In addition, minimum annotation guidelines for quantification experiments (called MIAPE-Quant) were also developed.

ProteomeXchange was a Coordination Action project that consolidated an emerging informal collaboration between major repositories into a production-quality data deposition and dissemination consortium on par with the systems so successfully employed by three-dimensional structure databases and nucleotide sequence databases, amongst others.

To date, ProteomeXchange has received more than 1,150 data submissions with a total volume of more than 50 TB. We are a major contributor to rapidly increasing availability and reuse of proteomics data globally. We have published 72 manuscripts, among them the recent major consortium publication in the high-impact journal *Nature Biotechnology*<sup>2</sup>. The PRIDE 2013 *Nucleic Acids Research* (NAR) publication<sup>3</sup>, which we used as a reference publication prior to the above main consortium publication, is one of the five highest cited 2013 publications for EMBL-EBI with 361 citations (Google Scholar, August 2014).

---

<sup>2</sup> Vizcaíno JA, Deutsch EW, Wang R, Csordas A, Reisinger F, Ríos D, Dianas JA, Sun Z, Farrah T, Bandeira N, Binz PA, Xenarios I, Eisenacher M, Mayer G, Gatto L, Campos A, Chalkley RJ, Kraus HJ, Albar JP, Martinez-Bartolomé S, Apweiler R, Omenn GS, Martens L, Jones AR, Hermjakob H. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol.* 2014 Mar;32(3):223-6.

<sup>3</sup> Vizcaíno JA, Côté RG, Csordas A, Dianas JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J, O'Kelly G, Schoenegger A, Ovelheiro D, Pérez-Riverol Y, Reisinger F, Ríos D, Wang R, Hermjakob H. The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 2013 Jan;41(Database issue):D1063-9.

## 1.2 Summary description of project context and objectives

ProteomeXchange had two major objectives:

1. Further development and implementation of data representation standards for proteomics.
2. The definition and implementation of consistent, harmonized data deposition and exchange procedures adhered to by the major public proteomics repositories.

The ProteomeXchange project was shaped as a three and a half-year Coordination and Support Action that was developed along 6 work packages, executed by 13 scientific partners. A contract amendment was approved on 13/02/2014 by the commission for the addition of the partner SNM to manage the project.

**WP1**, “management”, ensured efficient communication in the ProteomeXchange consortium, maintained the consortium website, organized regular management phone conferences, and maintained a comprehensive documentation of all consortium decisions in the form of minutes available to the partners via the website.

**WP2**, “standards development”, contributed to the maintenance and further development of the qualitative Proteomics Standards Initiative (PSI) standards mzML, mzIdentML, and TraML, together with the development of a standard for the representation of quantitative mass spectrometry data called mzQuantML.

**WP3**, “Data management system”, was devoted to the implementation of software support for exporting the data formats needed to perform a ProteomeXchange submission. The developments included adaptations of existing data management software (ProCon, OmicsHub), development of an open source LIMS system (ms\_limsX, later renamed as colims), and redevelopment of the open source PRIDE Converter tool.

**WP4**, “Data deposition”, was the key collaborative work package establishing co-ordinated data management between the repositories, receiving feedback from the journals as essential stakeholders. The developments within the ProteomeXchange grant were centered on PRIDE and PeptideAtlas. At present, PRIDE acts as the single point of contact for data deposition for tandem MS data, while the PASSEL component of PeptideAtlas has an equivalent role for Selected Reaction Monitoring (SRM) data. Once the data is public, it is disseminated from the submission repositories through the ProteomeCentral resource to PeptideAtlas, UniProt and to the proteomics community as a whole.

**WP5**, “Dissemination pipelines”, had two major tasks: 1) the broadcast of the metadata related to the datasets via an RSS feed and 2) the development of ProteomeCentral as the PX identifier service and an archive of all the ProteomeXchange datasets and related metadata.

**WP6**, “Outreach”, ensured close contact between the ProteomeXchange consortium and the wider proteomics community. Beyond the usual media website, mailing lists, discussion forums, and request tracking systems, a key outreach event was the annual PSI spring meeting. Four meetings have taken place during the entire project: the first was held in Heidelberg (Germany) on April 14-16, 2011; the second in San Diego (CA, USA) on March 15-16, 2012; the third was held in Liverpool (United Kingdom) on April 18-19, 2013 and the latest one in Rudesheim (Germany) on April 16-17, 2014.

The network was coordinated by Henning Hermjakob ([hhe@ebi.ac.uk](mailto:hhe@ebi.ac.uk)) and managed by Juan Antonio Vizcaíno ([juan@ebi.ac.uk](mailto:juan@ebi.ac.uk)) assisted by Pascal Kahlem ([pkahlem@s-n-m.eu](mailto:pkahlem@s-n-m.eu)) until 28/02/2014. From 01/03/2014 to 31/08/2014 the project management was overtaken by the partner SNM, where Pascal Kahlem worked together with Patricia Carvajal ([pcarvajal@s-n-m.eu](mailto:pcarvajal@s-n-m.eu)). The ProteomeXchange executive committee was formed of all participating PIs or their deputies, plus the project manager. The executive committee took all major decisions, and communicated via direct emails, a mailing list

and minutes, and annual in-person meetings.

The executive committee consisted of the project coordinator Henning Hermjakob (WP1/WP4 coordinator), Andrew Jones (WP2 coordinator), Lennart Martens (WP3 coordinator), Eric Deutsch (WP5 coordinator) and Ioannis Xenarios (WP6 coordinator).

Because of the distributed structure of a European Coordination and Support Action such as ProteomeXchange with 13 Partner institutions across 6 countries, an important effort was engaged by the project managers to maintain a good level of communication within the consortium by the use of all available modern techniques of “work at distance”:

- Phone conferences chaired by the project manager were held regularly in order to discuss issues between partners located at distance.
- A general mailing list [px-all@ebi.ac.uk](mailto:px-all@ebi.ac.uk). It allowed anyone of the project to send a message to all ProteomeXchange partners and stakeholders.
- A mailing list for principal investigators and project managers, [px-pi@ebi.ac.uk](mailto:px-pi@ebi.ac.uk) was also available.
- All the communication related to the standards development (WP2) took place in the dedicated PSI phone conferences and mailing lists. The relevant mailing lists were: [psidev-ms-dev@lists.sourceforge.net](mailto:psidev-ms-dev@lists.sourceforge.net) (MS workgroup), [psidev-pi-dev@lists.sourceforge.net](mailto:psidev-pi-dev@lists.sourceforge.net) (Proteomics Informatics workgroup) and [psidev-ms-vocab@lists.sourceforge.net](mailto:psidev-ms-vocab@lists.sourceforge.net) (Controlled Vocabularies/ ontologies activities).

The ProteomeXchange **website** was developed by the external services team at the EMBL-EBI, in coordination with the project managers and receiving feedback from the project partners. It is now deployed under the URL <http://www.proteomexchange.org>. The website is built upon the Content Management System (CMS) “Drupal” and is used by the project members to share documents, i.e. minutes of meetings or deliverables (a password protects access-restricted pages). It also gives public access to the project documentation, the PX submission tool and the list of publications.

An important part of the management **was reviewing agendas** of the different work packages to ensure the efficient organization and progress in adequacy with the plan exposed in the Description of Work.

### 1.3 A description of the main S&T results/foregrounds

The project reporting was structured in two periods, the first one on months 1-18, and the second one on months 19-42. The project progressed well, in line with the planned objectives of the work packages (months 1-42).

The ProteomeXchange two major objectives were subdivided in five milestones:

#### Objective 1: Further development and implementation of data representation standards for proteomics.

This objective was attained with the achievement of the MS2 at month 15:

- MS2: Framework for representation of the first full quantitative datasets (WP2) due at month 15.

Originally, two competing standards for the representation of mass spectra existed, mzData developed by the HUPO PSI (Proteomics Standards Initiative), and mzXML, developed by the Institute for Systems Biology (ISB). Subsequently, both jointly developed and released the **mzML** standard for mass spectra, which unites the best of both approaches, and is now widely implemented by instrument vendors and search engine providers. The next standard in the chain, **mzIdentML** for the representation of search engine parameters and results, was originally released in June 2009. Both standards, as well as the **TraML** standard for the representation of SRM transitions, were maintained and further developed during the first half of the project (Deliverable 2.1).

For the representation of quantitative mass spectrometry data, the HUPO PSI decided in 2008 to focus on the finalization of the **mzIdentML** format for peptide and protein identifications, deferring the

representation of quantitative information to a different data format called **mzQuantML**. The first released version of **mzQuantML** established a framework for the representation of quantitative mass spectrometry and quantitative results (Deliverable 2.2). Then small technology-specific modules were developed (Deliverable 2.3). This allowed providing data producers with a highly relevant standard quickly, while maintaining the flexibility to rapidly react to new or evolving quantitation technologies. During the second period, modules for different quantitation techniques (e.g. SRM, absolute quantification) were developed in **mzQuantML** (Deliverables 2.4, 2.5, 2.6) and the PSI standards **mzML**, **mzIdentML**, and **TraML** continued to be maintained (Deliverable 2.7).

Objective 2: The definition and implementation of consistent, harmonized data deposition and exchange procedures adhered to by the major public proteomics repositories

*Milestones involved:*

- MS1: ProteomeXchange format framework complete (WP1)

During the first period (months 1-18), the consortium established its concept, documentation and website (Deliverables 1.1 and 1.2 were submitted) and during the second period (months 19-42) the consortium documentation from year 2, 3 and 4 were delivered (Deliverables 1.3, 1.4 and 1.5 were submitted).

- MS3: Basic ProteomeXchange support across four LIMS systems (WP3)

During the first period (months 1-18), ProteomeXchange data export functionalities have been developed for the data management systems OmicsHub (Deliverable 3.1), for ProteinScape (Deliverable 3.2) and for Phenyx (Deliverable 3.3). The submission tool PRIDE Converter was also redeveloped to deal more efficiently with ProteomeXchange submissions (Deliverable 3.4). The development of the LIMS system originally called ms-limsX was started and its final name is 'colims' (Deliverables 3.5 and 3.6). In the second period (months 19-42), the colims system development was completed (Deliverable 3.7) and officially released, including full documentation and installation guide (Deliverable 3.9). The ProteomeXchange export functionality in colims (Deliverable 3.8) and its export functionality for quantification data in colims (Deliverable 3.10) were also implemented.

- MS4: Definition and implementation of ProteomeXchange data deposition process (WP4 and WP6)

During the first period (months 1-18), data producers were provided with a unified submission procedure ensuring maximal exploitation of data across proteomics resources with different aims and approaches (Deliverables 4.1, 4.2 and 4.3). A data producer can deposit data either in PRIDE (tandem MS data) or PASSEL (SRM data), who provides a static, immutable view of the data as supporting a specific publication (Deliverable 4.5). The potentially very large volumes of raw data can be stored in raw data repositories set up at EMBL-EBI and ISB, respectively (Deliverable 4.4). In the second period (months 19-42) the quantitation option was added to the ProteomeXchange repository data flow (Deliverable 4.6) and to MIAPE (Minimum Information About A Proteomics Experiment) guidelines for quantitative techniques were developed (Deliverable 4.7).

WP6 ensured a close connection between the ProteomeXchange consortium and the wider proteomics community. During the first period (months 1-18), the Stakeholder meeting (Year 1) was conducted (Deliverable 6.1) and the web-based tutorial 1 about "Proteomics Data Deposition and Dissemination through ProteomeXchange" deliverable was submitted (D6.4). In the second period (months 19-42), the second-, third- and fourth-year stakeholder meetings were conducted (Deliverables D6.2 and D6.3); the web-based tutorial 2 deliverable was submitted (D6.5) and the second and third ProteomeXchange training workshops deliverables were also submitted (D6.6 and D6.7).

- MS5: First view of full ProteomeXchange data flow from submitter to secondary repositories

During the second period (months 19-42), the central ProteomeXchange dataset identifier lookup service “ProteomeCentral” was created (Deliverable 5.1). The reprocessed-data broadcast mechanism was implemented (RPXD datasets, Deliverable 5.2) as well as the ProteomeXchange data import in UniProt and PeptideAtlas (Deliverables 5.3 and 5.4). And the implementation of extra functionality in ProteomeCentral was also submitted (Deliverable 5.5).

**In summary, the major achievements of the project were as follows:**

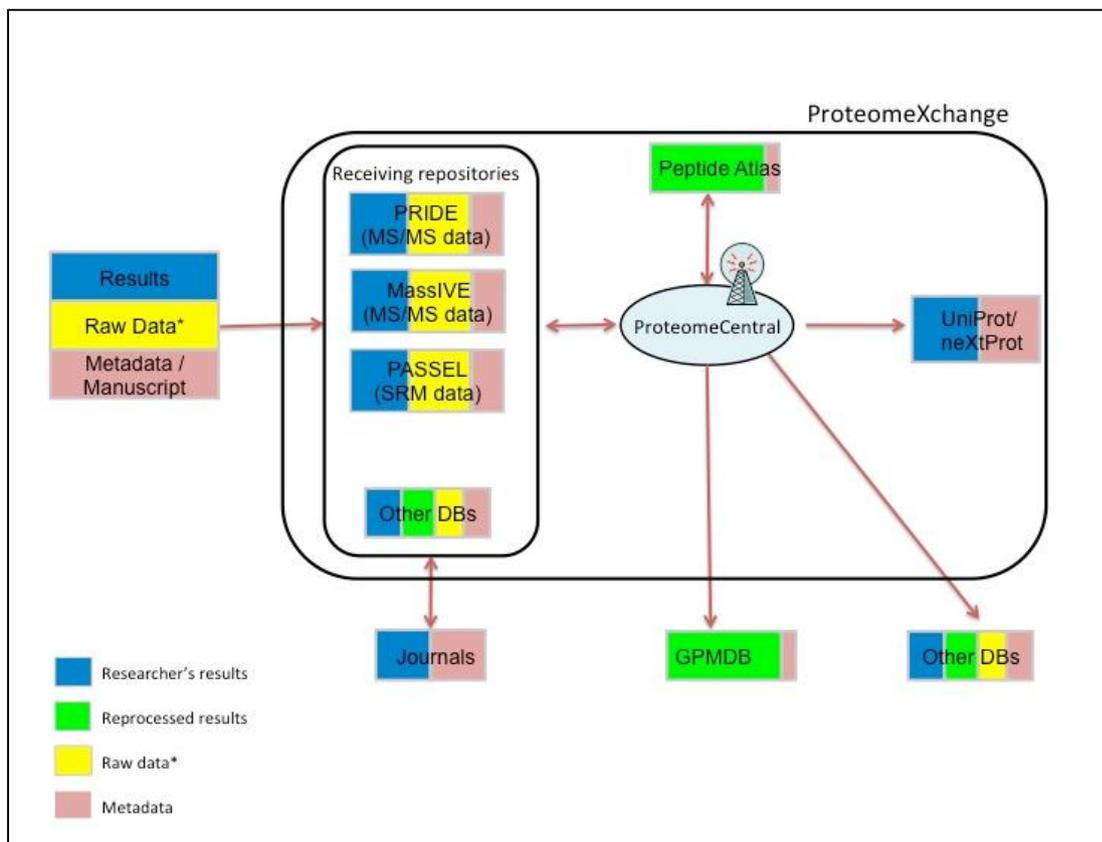
The project progressed well, in line with the planned objectives of the work packages (months 1 - 42). Overall, the work plan was designed to reach major milestones at month 18 of the project (Milestones 3 and 4). Through completion of a series of deliverables, the ProteomeXchange consortium is now fully operational. A qualitative dataset can be submitted, being received by one repository, and then automatically distributed and incorporated in other participating repositories, where it is made public, and such incorporation reported through the resource ProteomeCentral (Figure 1). ProteomeXchange started to accept regular submissions in June 2012 (month 18). By June 30th 2014, more than one thousand datasets have had been submitted (1,053 datasets). Of those, 47% have been already been made publicly available through PX.

The core of the standard for quantitative data (mzQuantML) was developed during the first period. During the second period, mzQuantML was extended with modules to support different quantification techniques. The second phase of the project was devoted to fine-tuning and production mode for the ProteomeXchange data flow. In addition, tracking of reprocessed datasets was developed. Since the start of the project, the consortium has published so far a total of 72 scientific publications acknowledging ProteomeXchange.

In June 2014, the MassIVE repository for mass spectrometry data (led by Prof. Nuno Bandeira at University of California, San Diego) became an official full member of the ProteomeXchange consortium (it was the first member to join that was not originally included in the FP7 grant). MassIVE’s integration with the ProteomeXchange consortium extended the range of options for sharing for proteomics mass spectrometry data and facilitated dataset submissions from institutions within and to the United States. It is also an impressive testimony to the attraction of the ProteomeXchange concept, and to its long-term viability.

All deliverables were submitted on time and are available on the ProteomeXchange website at the following URL, going through the different work packages:

<http://www.proteomexchange.org/project-organization/workpackages>



**Figure 1.** Representation of the ProteomeXchange workflow for MS/MS and Selected Reaction Monitoring data (adapted from (1)). \*Raw data represents mass spectrometer output files.

#### 1.4 The potential impact (including the socio-economic impact and the wider societal implications of the project so far) and the main dissemination activities and exploitation of results

Data sharing in the field of proteomics has been on the rise for several years now, with supporting infrastructure built at several locations worldwide. However, for end users, data dissemination used to be a challenge, as they were confronted with two important questions: (i) “where do I deposit my own data” and (ii) “where can I easily access all publicly available data”. ProteomeXchange addressed these two concerns directly and effectively: coordinating large-scale proteomics data sharing between the different resources, while allowing convenient access to all publicly available data in any participating resource. Additionally, the development of a novel standard for quantitative proteomics data ensures the availability of minimum annotation guidelines (MIAPE-Quant) (3) and a standard format (mzQuantML) (4,5). Data submitters now are able to benefit from a broad platform of software tools tailored to make submission easy (6).

Overall, building on well-established collaborations prior to the start of the project, the ProteomeXchange consortium developed a network of resources for MS proteomics data, spanning from data generation via data representation in scientific journals and databases, to data re-analysis and exploitation in dedicated tools and third party resources.

The ProteomeXchange project improved the current situation for stakeholders in the proteomics field, and also increased the value of proteomics to general molecular biology. In this context, Data Producers benefit from:

- Further development of data representation standards, allowing more and easier reuse of their data, leading to increased citations. This was also supported by introduction of a co-ordinated ProteomeXchange dataset identifier, used by all participating resources, and allowing better credit attribution to the original data producers. Implementing a suggestion from journal representatives

at our stakeholders meetings, we are issuing DOIs (Digital Object Identifiers) for complete PX submissions, making them citable and compatible with established academic impact metrics.

- Data representation standards through an increased ability to combine different elements of a proteomics dataflow (instrument, search engine, results interpretation, and back to instrument configuration), rather than being tied to a monolithic provider-specific data flow.
- Easier data deposition process through implementation of standards and ProteomeXchange formats in commercial and freely available data management (LIMS) software, the major focus of the project.
- Unified data deposition procedures, now it is possible to interact with one data repository, rather than with potentially multiple repositories, plus individual requests from research projects for specific data or metadata items

Primary data repositories benefit from rapidly increasing data content. PRIDE data volume has approximately doubled each year of the grant period.

Journals benefit from a streamlined and consistent interaction with repositories, and from stable availability of public proteomics data (7). The Wiley *Proteomics* journal already mandates data deposition in ProteomeXchange or comparable repositories for the “Dataset Brief” manuscript type. ProteomeXchange representatives Eric Deutsch and Henning Hermjakob presented PX at a recent proteomics repositories workshop organised by *Molecular and Cellular Proteomics*, the leading journal in the field.

Editors, reviewers and authors benefit from automated and systematic curator consistency checks implemented by the repositories, as well as from metadata capture and transmission after data deposition (8,9).

Secondary resources like GPMDB, PeptideAtlas, UniProt, Ensembl, commercial users like large pharmaceutical companies, and systems biology efforts already benefit from simplified access to more, and more consistent data, as well as automated notification of new data sets through the ProteomeXchange RSS feed. As an example, the majority of GPMDB’s “Dataset of the week” are now based on ProteomeXchange submissions (<http://www.thegpm.org/>).

Funders already start to benefit from re-use of already funded datasets, as exemplified in the recent *Nature* publication of the “Mass-spectrometry-based draft of the human proteome”, which to a significant part reused PX datasets to increase the coverage of the new ProteomicsDB resource (10).

#### **1.4.1. Additionally expected longer term impact**

- Method developers will benefit from easier, more consistent, and timely access to relevant data, benefitting approaches such as the prediction of SRM transitions and search engines based on spectral libraries.
- Proteomics as a field will benefit from increased credibility due to improved public scrutiny and validation of published results.
- The wider molecular biology community in general and systems biology in particular, will benefit from easier access to qualitative, quantitative and dynamic proteomics data.
- The inclusion of MassIVE as a full member of the ProteomeXchange consortium strongly emphasizes that the computational proteomics community is united in its vision for freely-accessible data sharing. By adopting and contributing to the PX, PSI and related efforts for standard open formats and data sharing protocols, MassIVE further reinforces the enabling utility of these standards.

## 1.4.2 Main dissemination activities and exploitation of results

### Dissemination activities

The dissemination activities of the ProteomeXchange project were described in WP6. A major part of these activities were focused to reach out to proteomics data producers, instructing them on data submission guidelines and practical data deposition strategies. It is also included information on the use of the freely available local data management system for day-to-day lab data management. The Beneficiaries have produced so far since January 2011 72 scientific publications acknowledging ProteomeXchange's funding in international journals, as well as contributed to 91 (39 + 52 in the first and second period respectively) international meetings. The mentioned scientific publications are listed in the ProteomeXchange, EC Participant Portal and OpenAIRE web site (<http://www.openaire.com>).

The training and dissemination activities described in WP6 are only part of the overall training activities of the ProteomeXchange partners. They were integrated in the context of other existing training activities, and both benefited from existing activities as well as complementing each other.

For example, planned text- and web- tutorials were integrated or at least referred-to in general courses on proteomics and proteomics data management. In the same vein, the planned workshops integrated already existing material on the respective core resources of ProteomeXchange participants, including information on how to use tools that reuse publicly available proteomics data such as spectral libraries or predictions for targeted proteomics.

Four training courses have been organised under the auspices of ProteomeXchange:

- a) Bioinformatics for MS analysis 2012. CMU-Faculty of Medicine, Geneva, Switzerland. 15-19/10/2012. Organisers: EuPA Education Committee in collaboration with the SIB Swiss Institute of Bioinformatics, the Swiss Proteomics Society and the ProteomeXchange Consortium.
- b) Wellcome Trust Proteomics Bioinformatics Course 2012. EBI IT room, Hinxton, Cambridge, United Kingdom. 05-09/11/2012. Organizers: EMBL-EBI and The Wellcome Trust.
- c) Wellcome Trust Proteomics Bioinformatics Course 2013. EBI IT room, Hinxton, Cambridge, United Kingdom. 11-15/11/2013. Organizers: EMBL-EBI and The Wellcome Trust.
- d) Data sharing in proteomics: databases, repositories, standards, data submission 2013. Swiss Institute of Bioinformatics, Lausanne, Switzerland. 20-22/11/2013. Organisers: SIB Swiss Institute of Bioinformatics and the ProteomeXchange Consortium.

### Exploitation of results

Interestingly, the availability of the above mentioned pipelines is clear evidence that the exploitation of results is already in large part in place. There is a growing trend towards public data reuse which is facilitating the assessment, reuse, comparative analyses and extraction of new findings from published proteomics data. PX has also contributed to the improvement of the traceability of this public data reuse, thanks to the universal dataset identifier. In fact, one prominent example of reusability of PX datasets took place in the elaboration of the recently published 'draft of the human proteome' (10), which re-used a significant number of PX datasets, but there have been already other examples. In addition, proteomics data in public resources is used routinely for the generation of spectral libraries or the design SRM transitions. Furthermore, data from the PeptideAtlas and PRIDE resources is also already in active use in the annotation of proteins in UniProt, a collaborative effort that was expanded and consolidated during ProteomeXchange.

Through the education of data submitters, and the education of data consumers as highlighted above, ProteomeXchange not only covered end-to-end knowledge dissemination, but also created added value and responsibility towards other researchers. In order to reach an audience as large as possible, there was a strong exchange and coordination of overall training activities that was fostered by the members of ProteomeXchange.

In June 2014, the Mass spectrometry Interactive Virtual Environment (MassIVE) repository became an official full member of the ProteomeXchange consortium. MassIVE's integration with the

ProteomeXchange consortium extended the range of options for sharing proteomics mass spectrometry data and will facilitate dataset submissions from institutions with access within and to the United States. The inclusion of MassIVE as a full member of the ProteomeXchange consortium strongly emphasizes that the computational proteomics community is united in its vision for freely-accessible data sharing. By adopting and contributing to the PX, PSI and related efforts for standard open formats and data sharing protocols, MassIVE further reinforces the enabling utility of these standards.

ProteomeXchange partners invited other proposal participants for presentations or training events at their own facilities, and they also aim at co-ordinating activities of ProteomeXchange partners at conferences and similar events, for example through proposal of joint or co-ordinated instead of parallel presentations, together contributing to the dissemination of activities as well as general expertise in proteomics data management. The global spread of ProteomeXchange partners thus was optimally exploited in reaching as wide a target audience as possible.

We developed high quality training material for ProteomeXchange data management and deposition, comprising:

- Tutorial-style publications: “Standardization and Guidelines: How to submit MS proteomics data to ProteomeXchange via the PRIDE database” (11).
- Interactive video clips, similar to the ones developed for data deposition in the PRIDE database, both for data deposition, and for the use of the free tool for local data management.
- A complete module on proteomics data for the EMBL-EBI e-learning platform.
- Four workshops on proteomics data deposition and analysis, available to any interested organisation. Normally the host is requested to pay the travel expenses of the trainers. These workshops are “marketed” both from the ProteomeXchange partner websites, and through existing training programmes like the EMBL-EBI “roadshow” (<http://www.ebi.ac.uk/training/roadshow/>).
- Short versions of the above, to be offered as half day tutorials in the framework of conferences, such as the ISMB and HUPO conferences.

## 1.5 Public website address (if applicable), as well as relevant contact

The project website has the address: [www.proteomexchange.org/](http://www.proteomexchange.org/)

### List of relevant URLs

This is a list of URLs to the most updated documentation and resources:

#### A) General

- PX home page: <http://www.proteomexchange.org>

#### B) Data submission

- PX submission guidelines:

[http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/px\\_general\\_guidelines.pdf](http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/px_general_guidelines.pdf)

- PX submission tool: <http://www.proteomexchange.org/submission>

- PX submission tool tutorial:

[http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/px\\_submission\\_tutorial.pdf](http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/px_submission_tutorial.pdf)

- Web course “How to submit MS/MS data to PX via PRIDE” (EBI E-learning platform):

<http://www.ebi.ac.uk/training/online/course/proteomexchange-submissions-pride>

- How to do bulk submissions for MS/MS data:

<http://www.proteomexchange.org/bulk-submission>

- PX submission tool file format:

[http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/proteomexchange\\_submission\\_summary\\_file\\_format.pdf](http://www.proteomexchange.org/sites/proteomexchange.org/files/documents/proteomexchange_submission_summary_file_format.pdf)

- PASSEL submission form:

[https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/PASS\\_Submit?datasetType=SRM](https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/PASS_Submit?datasetType=SRM)

### C) Data access

- ProteomeCentral: [www.proteomecentral.proteomexchange.org/](http://www.proteomecentral.proteomexchange.org/)

- PX XML schema:

<http://code.google.com/p/proteomexchange/source/browse/schema/proteomeXchange-1.1.0.xsd>.

### *Contractors involved*

European Bioinformatics Institute	EMBL	Henning Hermjakob
The University of Liverpool	UofL	Andrew Jones
VIB	VIB	Lennart Martens
Eidgenössische Technische Hochschule Zurich	ETH Zurich	Ruedi Aebersold
Institute for Systems Biology	ISB	Eric Deutsch
The Regents of the University of Michigan	U of Michigan	Phil Andrews
Integromics SL ITG	ITG	Eduardo Gonzalez
Ruhr-Universitaet Bochum	RUB	Christian Stephan/Martin Eisenacher
Swiss Institute of Bioinformatics	SIB	Ioannes Xenarios
GeneBio	GeneBio	Pierre-Alain Binz
Wiley-VCH Verlag GmbH & Co KGaA	Wiley-VCH	Hans Joachim Kraus
Agencia Estatal Consejo Superior de Investigaciones Científicas	CSIC	Juan Pablo Albar
Scientific Network Management SL	SNM	Pascal Kahlem

### *Coordinator's contact details*

Henning Hermjakob

Team Leader Proteomics Services

EMBL-European Bioinformatics Institute

Wellcome Trust Genome Campus

Hinxton

Cambridge CB10 1SD

United Kingdom

Tel: +44 (0)1223 494671

Fax: +44 (0)1223 494468

E-mail: [hhe@ebi.ac.uk](mailto:hhe@ebi.ac.uk)

Project managers: Juan A. Vizcaíno/Pascal Kahlem/Patricia Carvajal

## References

1. Vizcaino, J.A., Deutsch, E.W., Wang, R., Csordas, A., Reisinger, F., Rios, D., Dianes, J.A., Sun, Z., Farrah, T., Bandeira, N. *et al.* (2014) ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat Biotechnol*, **32**, 223-226.
2. Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J. *et al.* (2013) The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic acids research*, **41**, D1063-1069.
3. Martinez-Bartolome, S., Deutsch, E.W., Binz, P.A., Jones, A.R., Eisenacher, M., Mayer, G., Campos, A., Canals, F., Bech-Serra, J.J., Carrascal, M. *et al.* (2013) Guidelines for reporting quantitative mass spectrometry based experiments in proteomics. *Journal of proteomics*, **95**, 84-88.
4. Walzer, M., Qi, D., Mayer, G., Uszkoreit, J., Eisenacher, M., Sachsenberg, T., Gonzalez-Galarza, F.F., Fan, J., Bessant, C., Deutsch, E.W. *et al.* (2013) The mzQuantML data standard for mass spectrometry-based quantitative studies in proteomics. *Mol Cell Proteomics*, **12**, 2332-2340.
5. Qi, D., Krishna, R. and Jones, A.R. (2014) The jmzQuantML programming interface and validator for the mzQuantML data standard. *Proteomics*, **14**, 685-688.
6. Cote, R.G., Griss, J., Dianes, J.A., Wang, R., Wright, J.C., van den Toorn, H.W., van Breukelen, B., Heck, A.J., Hulstaert, N., Martens, L. *et al.* (2012) The PRoteomics IDentification (PRIDE) Converter 2 framework: an improved suite of tools to facilitate data submission to the PRIDE database and the ProteomeXchange consortium. *Mol Cell Proteomics*, **11**, 1682-1689.
7. Csordas, A., Wang, R., Rios, D., Reisinger, F., Foster, J.M., Slotta, D.J., Vizcaino, J.A. and Hermjakob, H. (2013) From Peptidome to PRIDE: public proteomics data migration at a large scale. *Proteomics*, **13**, 1692-1695.
8. Wang, R., Fabregat, A., Rios, D., Ovelleiro, D., Foster, J.M., Cote, R.G., Griss, J., Csordas, A., Perez-Riverol, Y., Reisinger, F. *et al.* (2012) PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat Biotechnol*, **30**, 135-137.
9. Csordas, A., Ovelleiro, D., Wang, R., Foster, J.M., Rios, D., Vizcaino, J.A. and Hermjakob, H. (2012) PRIDE: quality control in a proteomics data repository. *Database : the journal of biological databases and curation*, **2012**, bas004.
10. Wilhelm, M., Schlegl, J., Hahne, H., Moghaddas Gholami, A., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H. *et al.* (2014) Mass-spectrometry-based draft of the human proteome. *Nature*, **509**, 582-587.
11. Ternent, T., Csordas, A., Qi, D., Gomez-Baena, G., Beynon, R.J., Jones, A.R., Hermjakob, H. and Vizcaino, J.A. (2014) How to submit MS proteomics data to ProteomeXchange via the PRIDE database. *Proteomics*, *in press*.