Chapter 53

The Minimal Information About a Proteomics Experiment (MIAPE) from the Proteomics Standards Initiative

Salvador Martínez-Bartolomé, Pierre-Alain Binz, and Juan P. Albar

Abstract

During the last 10 years, the Proteomics Standards Initiative from the Human Proteome Organization (HUPO-PSI) has worked on defining standards for proteomics data representation as well as guidelines that state the minimum information that should be included when reporting a proteomics experiment (MIAPE). Such minimum information must describe the complete experiment, including both experimental protocols and data processing methods, allowing a critical evaluation of the whole process and the potential recreation of the work. In this chapter we describe the standardization work performed by the HUPO-PSI, and then we concentrate on the MIAPE guidelines, highlighting its importance when publishing proteomics experiments particularly in specialized proteomics journals. Finally, we describe existing bioinformatics resources that generate MIAPE compliant reports or that check proteomics data files for MIAPE compliance.

Key words Proteomics, HUPO-PSI, MIAPE, Standards, Reporting, Guidelines, Semantic validator

1 Introduction

With the growing number of variations of high-throughput proteomics techniques in the last 10 years, the scientific community soon detected the need to define reporting guidelines that ensure some minimal data and meta-data quality and consistency as a requirement prior to publish protein and peptide identification data. Following this, a group of experts proposed the so-called PARIS guidelines in 2004 [1]—later revised in 2005—and then implemented in the instructions for authors by the journal Molecular and Cellular Proteomics. Then, during 2006 and 2007, the Proteomics Standards Initiative (PSI) [2] from the Human Proteome Organization (HUPO), developed the concept of the set of Minimal Information about a Proteomics Experiment (MIAPE) guidelines, based on the experience of the micro-array community with their MIAME guidelines. A MIAPE parent document [3, 4] which defines the principles and objectives of MIAPE

Jesus V. Jorrin-Novo et al. (eds.), *Plant Proteomics: Methods and Protocols*, Methods in Molecular Biology, vol. 1072,
DOI 10.1007/978-1-62703-631-3_53,

guidelines laying the foundations for a set of MIAPE guideline modules would be published during the following years.

The MIAPE guidelines, as defined by HUPO-PSI, aim at listing the information that should be provided while describing a proteomics experiment. As a complete proteomics experiment can be divided into smaller experimental and data analysis steps, the various HUPO-PSI working groups, composed of experts from different proteomics fields, has defined one or more MIAPE modules each covering one of these steps. The different MIAPE modules are the result of discussions between working group experts, software developers, hardware developers and end users. The documents have further followed the formal review process internally defined by PSI [3] and are available on the HUPO-PSI Web site (www.psidev.info). In most of the cases, these guidelines have also passed the review process of Nature Biotechnology as they have been published there as well.

Subsequently, several joint sessions between HUPO-PSI and publication committees were held [4, 5], in which journals fed back their opinions about how to ensure proteomics data quality before publishing it; how to allow the reprocessing of the data; whether to require raw data deposition in a public repository or not; and how to tackle the adoption of MIAPE guidelines in their respective instructions for authors. Most importantly, proteomics journals have been taking part in these meetings, such as Nature Biotechnology, Molecular and Cellular Proteomics, Proteomics, Journal of Proteomics and Journal of Proteome Research. As a result, several of these journals have been encouraging MIAPE compliance [1, 8, 10] to submit manuscripts for several years.

It is quite important for the scientific community that MIAPE guidelines become a standard that regulates the minimum information compiled by every proteomics scientific publication, since they will turn into a quality stamp to ensure a critical review and a potential repetition of the results. In addition, it is also crucial that tools are available to help authors of proteomics manuscripts as well as proteomics data producers to reach compliance with minimal technical burden.

2 The Proteomics Standards Initiative

Initiated on 2002, the Proteomics Standards Initiative (PSI) is an open community and one of the 13 scientific initiatives of HUPO that involves researchers, database providers, vendors, software developers and publishers. Its mission statement is: HUPO-PSI defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification. Five main

working groups divide the vertical topics: Molecular Interactions (MI), Proteomics Informatics (PI), Mass Spectrometry (MS), Protein Separations (PS) and Protein Modifications (MOD). Additionally, a proposal for the creation of a new working group dedicated to proteomics experimental protocol standardization was presented in the last PSI meeting in San Diego (USA) on March 2012. It also has a steering committee that takes care of horizontal administrative and logistical tasks. Every year, a workshop [6–16] is organized in which the participants report on the status of the projects and split in different tracks to address specific outstanding issues. Developments and definitions are also carried out by individuals during the year, followed up and coordinated via regular teleconferences and in some cases, by face-to-face workshops on a specific topic. Additionally, progresses of HUPO-PSI standard developments are presented every year in the HUPO World Congress where a satellite meeting or a specific session [17–20] is dedicated; this also provides a wider panel to make the deliverables visible and to allow for additional discussions with the proteomics community.

HUPO-PSI is producing the following types of standards:

– *Standard formats for proteomics data representation*. These data formats are primarily XML formats defined by XML schemas (xsd) (Extensible Markup Language). In addition to the schema, a semantic validator [21] is provided in most of the cases to check the schema but, more importantly, also to ensure the proper usage of controlled vocabulary terms as well as the compliance with specific constraints (*see* Subheading 5.2.3). These standard data formats are meant to allow seamless exchange of data all the way along a proteomics workflow. They also, with the help of converter tools, aim to solve the problem of the difficulty of handling different proprietary data formats from various vendors while proposing a common way to represent data and meta-data. Current released formats are mzML [22], mzIdentML [23, 24], traML [25], gelML [26], and PSI-MI XML [27].
– *Controlled vocabularies*. A number of controlled vocabularies are developed to support the implementation of the data formats. They cover all concepts, procedures, materials, equipment, bioinformatics tools, etc. that are necessary to encode information into the data formats. They are integrated in *The Open Biological and Biomedical Ontologies (OBO) foundry* [28], a collaborative project for the establishment of a set of principles for ontology development in the biomedical domain. Proteomics information is coded into the standard data files using these ontology terms, which enormously facilitates the data interpretation by third-party bioinformatics tools.

– *MIAPE guidelines*. They describe the minimum information needed to appropriately report information and data about a proteomics experiment. Objectives and principles of MIAPE guidelines are defined in a parent document [4], which describes the minimum information that describes the experimental context, allows the understanding of the results and their interpretation sufficiently to permit a critical evaluation and, in principle, a potential recreation of the work. However, it is important to stress the fact that MIAPE guidelines intend neither to make any data quality judgement, nor to fix any data format for its representation or to try to establish the way to run an experiment. MIAPE guidelines have been defined in a modular way, and each MIAPE module defines the minimum information related to a certain part of the proteomics data flow. HUPO-PSI has currently produced nine different MIAPE documents/modules that will be described later in this chapter.

Data Formats, Controlled Vocabularies and MIAPEs are strongly interrelated since Controlled Vocabularies provide the way to represent proteomics data in the standard data files, and MIAPE modules define which data the file should contain to be MIAPE-compliant (Table 1).

Standards produced by HUPO-PSI are reported to the scientific community following a set of formal requirements defined by the initiative itself. Accordingly, documents can be one of the following types:

(a) *Community practice documents*, which inform and influence the community regarding an approach or process that is considered to be widely accepted by consensus and practice in the Proteomics community.

(b) *Informational documents*, which inform the community of an interesting and useful proteomics-related technology, architecture, framework or concept

(c) *MIAPE (minimum information about a proteomics experiment) documents*, which inform the community as to the minimal information that should be captured about an experiment to enable its results to be clearly interpreted and validated.

(d) *Recommendation documents*, which describe a particular technical specification or a particular set of guidelines for the application of a technical specification. Recommendations are intended to guide interoperability and promote standard approaches.

Table 1
The MIAPE module defined by each PSI working group is showed in each row, together with its corresponding standard data format (if any) and ontology

Working group	Reporting guide lines (MIAPE)	Data exchange format	Controlled vocabulary
Protein separations (PS)	Gel Electrophoresis (MIAPEGE)	gelML (vl.1)	Separation methods CV
	Gel Informatics (MIAPE GI)	–	
	Capillary Electrophoresis (MIAPE CE)	spML (milestone 1)	
	Column Chromatography (MIAPE CC)		
Mass spectrometry (MS)	Mass Spectrometry (MIAPE MS)	mzML (vl.1)	PSI mass spectrometry CV; Protein modifications CV
	–	traML (vl.1)	PSI mass spectrometry CV; Protein modifications CV
Proteomics informatics (PI)	Mass Spectrometry Informatics (MIAPE MSI)	mzIdentML (vl.1)	PSI mass spectrometry CV; Protein modifications CV
	Mass Spectrometry Quantification (MIAPE Quant)	mzQuantML (v1.0)	PSI mass spectrometry CV; Protein modifications CV
Molecular interactions (MI)	Molecular Interactions (MIMIx)	PSI-MI XML (v2.5)	Molecular interaction CV
	Protein Affinity Reagent (MIAPAR)	PSI-PAR	Protein affinity reagents CV
	Bioactive Entity (MIABE)	–	–

Additionally, standards are usually published in specialized journals (see references) for a greater spreading in the community.

Before being published, each one of the previously described document types must pass a formal review process also defined by the initiative itself [5]. Such a review process defines different review phases. The first review phase (Fig. 1a), is performed by the steering committee, and the second by reviewers selected by the PSI editor and, in parallel, by all the scientific community since the document is made public to receive any other comment or critical review. In the event of informational or community practice documents, the review simply consists of a public phase (Fig. 1b).

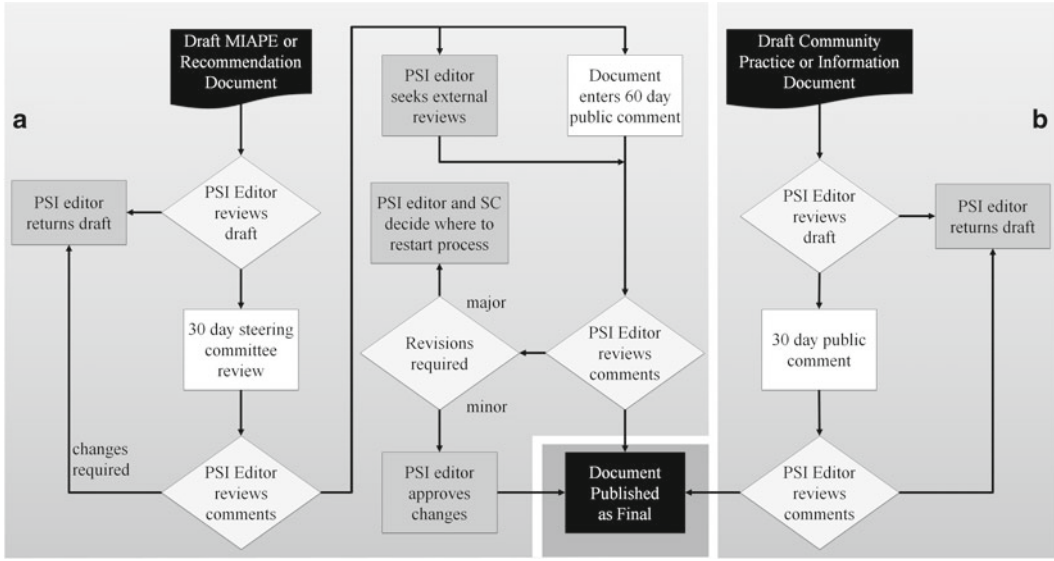

Fig. 1 Phases of the PSI formal document process by which standards are reviewed before to be published. MIAPE and recommendation documents (**a**); Community practice and information documents (**b**)

3 MIAPE Modules

As it has been already commented, MIAPE guidelines have been defined in a modular way, so each module contains a checklist of information to include when reporting a certain part of the workflow in a proteomics experiment.

In a typical proteomics experiment, we can differentiate different phases or steps: a sample separation in one or more dimensions in order to reduce its complexity; the acquisition of mass spectra by analyzing the sample in a mass spectrometer; and the bioinformatics analysis of the data generated in the previous step in order to identify or quantify proteins and/or peptides from the sample. Accordingly, with these proteomics workflow differentiations, HUPO-PSI defined the following MIAPE modules:

(a) MIAPE modules on complexity reduction of samples:

- MIAPE *gel electrophoresis* (MIAPE-GE) [29]: it describes the experimental protocol by which a sample is submitted to electrophoretic separation in a one or two dimensional gel matrix. This includes the gel matrix preparation and manufacturing, the run electrophoresis conditions, visualization techniques such as gel staining, as well as the scanning method performed to obtain the digitalized images of the gel matrixes.

- MIAPE *capillary electrophoresis* (MIAPE-CE) [30]: it describes the experimental protocol by which a sample is submitted to capillary electrophoresis, including different steps such as: pre-conditioning, injection, separation, and post-conditioning, and compiling different parameters such as temperature, detection length, pressure, voltages.
- MIAPE *column chromatography* (MIAPE-CC) [31]: it describes the experimental protocol by which a sample is submitted to a column chromatography separation step, including information such as column configuration, the appropriate mobile phase selection, the performed gradients for the chromatographic run, the sample fractioning collection.

(b) MIAPE module on mass spectrometry experiment:

- MIAPE *mass spectrometry* (MIAPE-MS) [32]: it describes the process in which a sample is analyzed by a mass spectrometer to generate the raw data files as well as the process that generates the processed spectra or peak lists. It also describes the employed equipment, including its configuration and mass spectra acquisition parameters.

(c) MIAPE modules about the bioinformatics analysis of data:

- MIAPE *gel informatics* (MIAPE-GI) [33]: it describes all the processes in which the digitalized images of the gels are analyzed by software to detect and quantify the spots or bands. This includes the experimental design description, including the description of selected group and/or replicates, the software description and parameters, and the detailed description of methods for the image alignment, detection, matching and quantification of spots, or the statistical analysis performed to determine the confidence of differential expression results.
- MIAPE *mass spectrometry informatics* (MIAPE-MSI) [34]: it describes the process by which acquired mass spectra are analyzed to identify proteins and peptides that exist in the sample. This includes all the database search engine processes, describing the software and parameters, any performed de-novo sequencing analysis, or any statistical analysis or post-processing of the identification data.
- MIAPE *mass spectrometry quantification* (MIAPE-Quant): it describes all the performed data analysis processes by way of a wide range of quantification techniques by mass spectrometry, such as chemical (iTRAQ, TMT, ICPL, …) or metabolic (SILAC) labelling-based techniques, label-free techniques (spectral counting, chromatogram alignment, etc.) or targeted quantification techniques such as

Multiple/Selected Reaction Monitoring (SRM/MRM). While writing this chapter, the guidelines for quantification were being reviewed under the PSI formal document process.

(d) MIAPE modules on molecular interaction experiments:

- *Minimum information about a molecular interaction experiment* (MIMIx) [35]: it describes the minimum information to report about molecular interaction experiments. It includes data, such as the host organism where the interaction has been detected, the interaction detection method, the list of participants in the interaction together with their identifiers in a public database, their biological and experimental roles.
- *Minimum information about a protein affinity reagent* (MIAPAR) [36]: it describes the protein affinity reagent characterization, such as antibodies used as protein identification tools. This includes the description of molecules participating in the interaction—i.e., the reagent molecule and the target molecule—features about the interaction, such as sensitivity, selectivity, the recognized epitope, binding constants, applications; and the characterization of the interaction by information such as kinetic constants, affinity measures, or the description of the characterization method of the reagent.
- *Minimum information about a bioactive entity* (MIABE) [37]: it describes drug-target data, including information such as molecule properties, molecule production, physicochemical properties, in vitro cell-free assays, whole organism studies and pharmacokinetic studies.

4 MIAPE Guidelines Evolution

Proteomics is a rapidly evolving field. New technologies and modifications of methods are regularly appearing in the literature and on the market that widen the range of tools available to proteomics scientists. Better understanding of the strength and limitations of these processes involves also the need for more precise requirements and SOPs that target optimized quality of the obtained results, As a consequence the MIAPE guidelines need to stay aligned with this evolution. Also, as already mentioned, scientific journals have actively participated in dedicated round tables to better align the generic MIAPE technical reporting guidelines with their own submission guidelines that include quality judgment aspects in addition to descriptive data and meta-data requirements [4, 5]. Representatives from each journal have shown statistics and data about how scientists

are able to comply with these guidelines. This has highlighted that in practice some fine tuning related to required information should be addressed before that MIAPE guidelines are fully adopted by journals as integral part of their rules. For example, the percentage of publication in which raw data has been made available by submitting it to a public repository is still low, often due to technical problems (data storage and/or bandwidth of transferring large amount of data) or a lack of bioinformatics resources to do it. Other explicitly required details in MIAPE guidelines were also difficult to compile by authors, due to ignorance or to lack of clarity on some definitions in the MIAPE modules.

All that reasons have been resulted in a further review of some MIAPE modules to facilitate their adoption as much as possible to the real life. So, the PSI document process actually allows the review of documents describing standards even if they were already published previously, provided that changes are well documented and justified [3]. It is the case of MIAPE MS and MSI modules, that were published on 2008 [32, 34], and they are currently under the PSI document process, after having changed some of their content. The reasons for these modifications have been the following:

- Some sections about quantification processes were present in 2008 MIAPE MS and MSI modules. These sections were deleted when a new MIAPE module (MIAPE-Quant) about quantification data was decided to be defined.
- A number of terms were removed and replaced by more generic terms to reduce over-specification.
- Some parts of the guidelines were restructured and definitions in the appendix were modified to improve the clarity of the checklist and make it easier to comply with journals requirements and PSI data standards.

So, hopefully, MIAPE MS and MSI documents will again be released after passing by a new PSI formal document process.

But, MIAPE modules are not the only standards that have been redefined. For example the standards for proteomics data representation mzML and mzIdentML were also slightly modified (from v1.0 to 1.1) after the detection of some technical issues by some developers that were implementing these formats in their tools.

MIAPE guidelines are not sufficient to describe all kinds of publishable data and results. They cover a large part of proteomics related experiments, but do not cover other fields such as micro array experiments, RNAi experiment, a T Cell Assay, a genotyping experiment. In that respect, MIAPE is registered to the MIBBI foundry [38], where over 30 similar projects covering a wide range of fields and technologies have deposited minimal information guidelines.

5 MIAPE Tools and Implementations

While reading through the documents, a proteomics scientist might feel that complying with MIAPE guidelines—i.e., providing all required experimental and analysis parameters—is not a trivial task. This is particularly true when some part of the experiment is done as a service, for instance when samples are analyzed by Mass Spectrometry by a core facility.

There is a real need for bioinformatics resources that help to search, extract, compile, check and store such minimal information in the most automated way. Below we will describe available resources related to the management of MIAPE information.

5.1 Semi-automatic Generation and Storage of MIAPE Reports

A MIAPE-compliant document describing a proteomics experiment can be included as a materials and methods section or as supplementary material in a manuscript for publication. To compile MIAPE information and generate a legible MIAPE document, storing them in a repository is just the purpose of these two bioinformatics tools: *MIAPEGelDB* [42] developed by the Swiss Institute of Bioinformatics (SIB) and the MIAPE Generator Tool [43] developed by the bioinformatics working group of ProteoRed-ISCIII [44, 45]

Both tools provide similar functionalities, being based on a set of Web forms dedicated to compile MIAPE information. In order to create a MIAPE-compliant report the user has to go through all requested fields typing in the data manually while it is stored in a database. MIAPE documents can be later viewed in HTML or plain text formats, exported to XML, and can be accessed by a permanent URL from remote sites.

MIAPEGelDB allows for the creation of MIAPE GE documents that describe gel-based proteomics experiments. The MIAPE Generator Tool allows for the generation of several MIAPE documents such as GE, GI, MS and MSI. Moreover, all the information stored in the repository can be easily reused as templates to generate new MIAPE-compliant reports, since the same protocols and/or parameters are commonly repeated through several experiments in a given laboratory, and solely the results or the data generated by those protocols are really specific for each experiment. Additionally, it provides a permission management system that allows users to share their documents with others, assigning them different levels of operational capabilities.

Regarding MIAPE GE documents that both tools can generate, it is important to stress that both tools allow exporting that information to its corresponding standard data schema, that is, the gelML [26] data format.

5.2 Automatic Extraction and Validation of MIAPE Information

If MIAPE documents describe what information is to be provided for a given experiment, the task of the Proteomics data representation formats as defined by HUPO-PSI is to provide a mean to store this information in a standardized manner. So for example, mzML data schema [22] presents all the information related to the processing of a sample as performed by a mass spectrometer, including instrument configuration, acquisition parameters, as well as the generated data itself, such as mass spectra and chromatograms. The mzIdentML data schema [23, 24] represents all the information related to the bioinformatics analysis of mass spectra that leads to the identification of peptides and proteins, which means the description of the process in which mass spectra are assigned to amino acid sequences belonging to protein databases. This includes all parameters used when submitting spectra to a search engine and possibly validation tools, as well as the results themselves, that is, the obtaining peptide and protein lists with attributed confidence levels. In addition to PSI standards, other data formats are aiming at storing relevant information from proteomics experiments. For example, the PRIDE data schema [39, 40] was developed by the European Bioinformatics Institute (EBI) in parallel to mzML and mzIdentML for the purpose of converting data and meta-data from proteomics experiments not expressed in PSI standard formats. PRIDE has become one of the most important repositories of protein identification data by mass spectrometry, currently containing almost 300 million of spectra, more than 50 million peptides and almost 9 million proteins in its database. In a near future, the PRIDE team will natively support mzML and mzIdentML data in their importing workflows.

More and more tools and software are capable of reading and/or writing these standard data formats (*see* http://psidev.info/mzml for mzML and http://psidev.info/mzidentml for mzIdentML), including free-tools, open-source tools and commercial tools. Proteomics data deposition in public repositories is enormously facilitated by these standards, which is required or strongly recommended to authors when publishing MS data and results in most of the proteomics journals.

An alignment between data contained by standard files and the information required by MIAPE guidelines is therefore of crucial importance. In other words, before to store standard files in a repository, it becomes necessary to check if the encoded information is MIAPE-compliant or not, and in case of no, it is also necessary to detect which information is missing and to provide the way to include it. The *ProteoRed MIAPE Web Toolkit* (PMWTK) [41] is able to do it, providing a MIAPE quality stamp to a given file indicating that the experiment is perfectly described and that could be evaluated and potentially reproduced by other scientists.

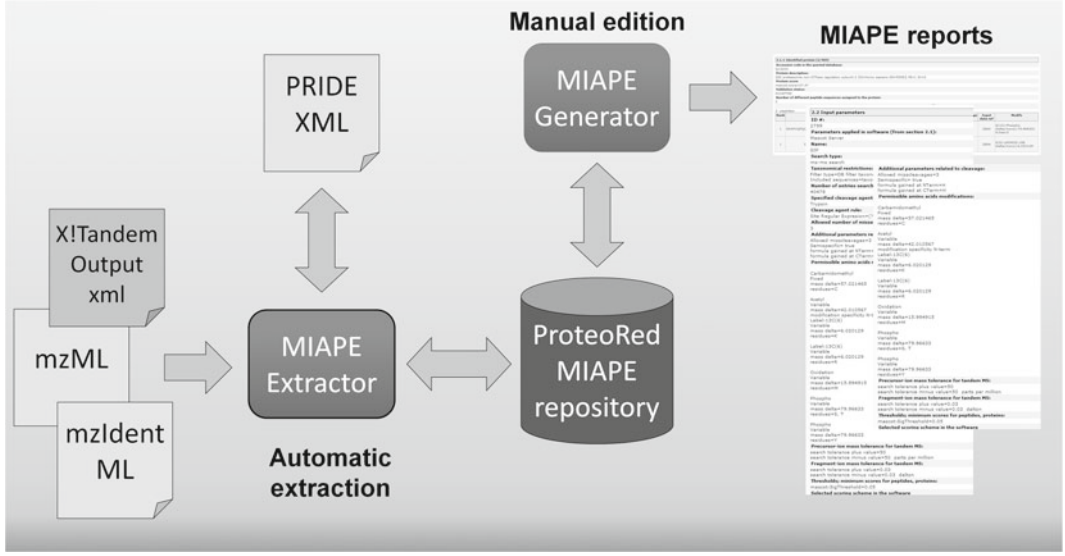

Fig. 2 Schematic workflow of the MIAPE Extractor and MIAPE Generator tools. MIAPE information is automatically extracted from mzML, mzIdentML, X!Tandem xml output file and PRIDE xml files, and it is stored in the ProteoRed MIAPE repository. This MIAPE information can then be manually checked and edited using the online MIAPE Generator tool which can generate MIAPE compliant reports that can serve as materials and methods section or as supplementary material in a manuscript for publication

5.2.1 The MIAPE Extractor

The *MIAPE extractor* tool provides automation to the PMWTK. It is a Java® stand-alone application that allows, in a fully automatic way, to extract MIAPE information from files such as PRIDE XML, mzML, or mzIdentML, and stores it in the ProteoRed MIAPE repository. So, the user can automatically generate MIAPE MS and MSI compliant reports from his proteomics data and then he can easily add missing MIAPE information by completing the corresponding Web forms from the MIAPE Generator tool. Additionally, the MIAPE Extractor tool is able to export MIAPE MS and/or MSI documents from the repository to a PRIDE XML file that will be MIAPE-compliant in case of these MIAPE reports were complete, and that can be sent to the PRIDE public repository for publication (Fig. 2).

5.2.2 The ProteoRed Java MIAPE-API

The MIAPE Extractor tool is developed using the ProteoRed Java MIAPE API, an Application Programming Interface (API) coded in Java that provides a data model for representing MIAPE GE, GI, MS and MSI documents; it includes a set of functionalities to extract MIAPE information from standard proteomics files; it also offers several programmatic ways to manage and store MIAPE information. This API licensed under an open-source license, is the unique programmatic resource that allows the development of tools managing MIAPE information.

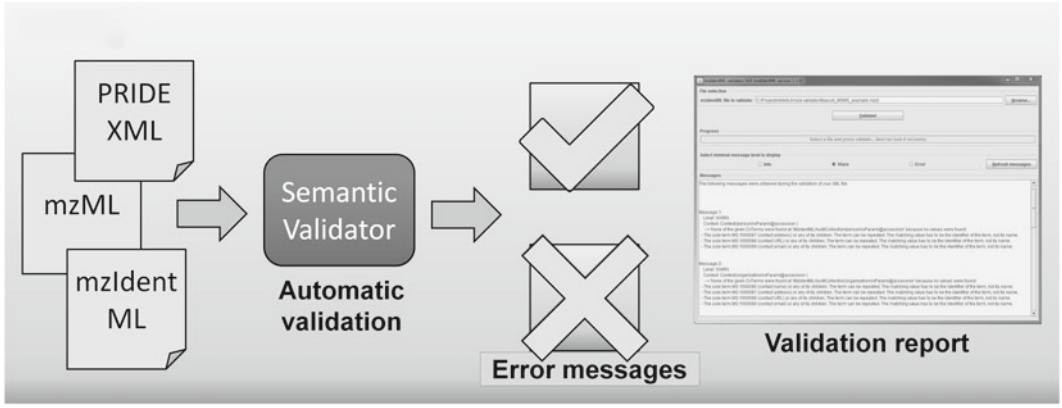

Fig. 3 Schematic workflow of the semantic validators. Standard data files are submitted to the validator software. If the file is MIAPE compliant, that is, it is semantically valid and contains all the data required by MIAPE guidelines, the validator will show a validation report confirming the MIAPE compliance of the file. However, if the file is not valid, the validation report will contain some error messages explaining which information is missing or missed placed in the file

Additionally, the bioinformatics working group of ProteoRed is also providing a set of Web-services that offers interaction with the ProteoRed MIAPE repository. So, it provides a remote and secure access to stored MIAPE documents, as well as a way to remotely store new MIAPE documents. The MIAPE Extractor tool uses these Web-services to interact with the ProteoRed MIAPE repository.

5.2.3 Semantic Validators

Together with the proteomics data representation schemas, a system for validating standard data files has been developed by HUPO-PSI. The called PSI semantic validator framework [21] not only checks the XML syntax but it also enforces rules regarding the use of an ontology class or Controlled Vocabulary (CV) terms by checking that the terms exist in the resource and that they are used in the correct location of a document. Moreover, this framework is extremely fast, even on sizable data files, and flexible, as it can be adapted to any standard by customizing the parameters it requires: an XML Schema Definition, one or more CVs or ontologies, and a mapping file describing in a formal way how the semantic resources and the format are interrelated. As such, the validator provides a general solution to the common problem in data exchange: how to validate the correct usage of a data standard beyond simple XML schema definition validation. It also tackles the issue of checking that experimental data reported using a specific format, CVs and public bio-ontologies (e.g., Gene Ontology, NCBI taxonomy) are compliant with the MIAPE recommendations. So, implementing the PSI semantic validator framework for each standard data format, HUPO-PSI provides a tool able to validate all the standard files (Fig. 3).

6 Discussion

The HUPO Proteomics Standards Initiative's developments are nowadays crucial for proteomics data interpretation, sharing and reprocessing. The definition of standards formats and the provision of tools that read and write these formats are particularly beneficial for highly collaborative scientific projects in which different proteomics laboratories with different equipment are exchanging data. This is also true for centralizing data and meta-data from variable sources in a repository such as PRIDE and therefore facilitates compliance with journals' requirements on data deposition. However, a minimal "quality control" (i.e., in the sense of MIAPE compliance) on proteomics data is desired in order to ensure that any reported, shared or published proteomics work has the minimum information that will allow a clear review of the performed procedures and produced data, as well as to allow a potential recreation of the work. We have described in this chapter a number of different bioinformatics resources that are available to help users to comply with required guidelines.

With this chapter, we present PSI standardization developments to the scientific community, in particular the MIAPE guidelines and related resources. These guidelines and associated tools are useful to help the community in producing, exchanging, interpreting and finally publishing proteomics data using a common language. In this era where datasets can live and exist by themselves, where they can obtain a DOI and become citable as proposed by the proteomeXchange consortium (www.proteomexchange.org), it is of interest for the whole community to appropriately generate well annotated datasets that can be reused and reinterpreted if needed.

Acknowledgments

This publication was supported by the Spanish National Research Council (CSIC), the Spanish National Proteomics Institute (ProteoRed-ISC III) (grant number 2005X747_3) and by the EU FP7 "ProteomeXchange" (EU FP7 grant number 260558) grants. The Swiss-Prot group is part of the SIB Swiss Institute of Bioinformatics. This activity is supported by the Swiss Federal Government through the The State Secretariat for Education, Research and Innovation SERI and by the European Union project "ProteomeXchange" (EU FP7 grant number 260558).

References

1. Carr S, Aebersold R, Baldwin M et al (2004) The need for guidelines in publication of peptide and protein identification data: working group on publication guidelines for peptide and protein identification data. Mol Cell Proteomics 3:531–533
2. Orchard S, Hermjakob H, Apweiler R (2003) The proteomics standards initiative. Proteomics 3:1374–1376
3. Vizcaino JA, Martens L, Hermjakob H et al (2007) The PSI formal document process and its implementation on the PSI website. Proteomics 7:2355–2357
4. Orchard S, Binz PA, Hermjakob H (2009) Second joint HUPO publication and proteomics standards initiative workshop. Proteomics 9:4426–4428
5. Orchard S, Ping P (2009) HUPO world congress publication committee meeting Aug 2008, Amsterdam, The Netherlands. Proteomics 9:502–503
6. Orchard S, Kersey P, Zhu W et al (2003) Progress in establishing common standards for exchanging proteomics data: the second meeting of the HUPO proteomics standards initiative. Comp Funct Genom 4:203–206
7. Orchard S, Kersey P, Hermjakob H et al (2003) The HUPO proteomics standards initiative meeting: towards common standards for exchanging proteomics data. Comp Funct Genom 4:16–19
8. Orchard S, Hermjakob H, Taylor CF et al (2005) Further steps in standardisation. Report of the second annual proteomics standards initiative spring workshop (Siena, Italy 17–20th Apr 2005). Proteomics 5:3552–3555
9. Orchard S, Apweiler R, Barkovich R et al (2006) Proteomics and beyond: a report on the 3rd annual spring workshop of the HUPO-PSI 21–23 Apr 2006, San Francisco, CA, USA. Proteomics 6:4439–4443
10. Orchard S, Hermjakob H, Taylor C et al (2006) Autumn 2005 workshop of the human proteome organisation proteomics standards initiative (HUPO-PSI) Geneva, 4–6 Sept 2005. Proteomics 6:738–741
11. Orchard S, Taylor CF, Jones P et al (2007) Entering the implementation era: a report on the HUPO-PSI fall workshop 25–27 Sept 2006, Washington DC, USA. Proteomics 7:337–339
12. Orchard S, Montechi-Palazzi L, Deutsch EW et al (2007) Five years of progress in the standardization of proteomics data 4th annual spring workshop of the HUPO-proteomics standards initiative Apr 23–25, 2007 Ecole Nationale Superieure (ENS), Lyon, France. Proteomics 7:3436–3440
13. Orchard S, Albar JP, Deutsch EW et al (2008) Annual spring meeting of the proteomics standards initiative 23–25 Apr 2008, Toledo, Spain. Proteomics 8:4168–4172
14. Orchard S, Deutsch EW, Binz PA et al (2009) Annual spring meeting of the proteomics standards initiative. Proteomics 9:4429–4432
15. Orchard S, Jones A, Albar JP et al (2010) Tackling quantitation: a report on the annual spring workshop of the HUPO-PSI 28–30 Mar 2010, Seoul, South Korea. Proteomics 10:3062–3066
16. Orchard S, Albar JP, Deutsch EW et al (2011) Enabling BioSharing—a report on the annual spring workshop of the HUPO-PSI Apr 11–13, 2011, EMBL-Heidelberg, Germany. Proteomics 11:4284–4290
17. Orchard S, Jones AR, Stephan C et al (2007) The HUPO pre-congress proteomics standards initiative workshop. HUPO 5th annual world congress. Long Beach, CA, USA 28 Oct–1 Nov 2006. Proteomics 7:1006–1008
18. Orchard S, Martens L, Tasman J et al (2008) 6th HUPO annual world congress—proteomics standards initiative workshop 6–10 Oct 2007, Seoul, Korea. Proteomics 8:1331–1333
19. Orchard S, Albar JP, Deutsch EW et al (2010) Implementing data standards: a report on the HUPOPSI workshop Sept 2009, Toronto, Canada. Proteomics 10:1895–1898
20. Orchard S, Albar JP, Deutsch EW et al (2012) From proteomics data representation to public data flow: a report on the HUPO-PSI workshop Sept 2011, Geneva, Switzerland. Proteomics 12:351–355
21. Montecchi-Palazzi L, Kerrien S, Reisinger F et al (2009) The PSI semantic validator: a framework to check MIAPE compliance of proteomics data. Proteomics 9:5112–5119
22. Martens L, Chambers M, Sturm M et al (2011) mzML—a community standard for mass spectrometry data. Mol Cell Proteomics 10(1):R110.000133. doi:10.1074/mcp.R110.000133–2
23. Eisenacher M (2011) mzIdentML: an open community-built standard format for the results of proteomics spectrum identification algorithms. Methods Mol Biol 696:161–177
24. Jones AR, Eisenacher M, Mayer G et al (2012) The mzIdentML data standard for mass spectrometry-based proteomics results. Mol Cell Proteomics 11:M111.014381
25. Deutsch EW, Chambers M, Neumann S et al (2011) TraML: a standard format for exchange

of selected reaction monitoring transition lists. Mol Cell Proteomics 10:O111.015446

26. Gibson F, Hoogland C, Martinez-Bartolome S et al (2010) The gel electrophoresis markup language (GelML) from the proteomics standards initiative. Proteomics 10: 3073–3081
27. Hermjakob H, Montecchi-Palazzi L, Bader G et al (2004) The HUPO PSI's molecular interaction format – a community standard for the representation of protein interaction data. Nat Biotechnol 22:177–183
28. Smith B, Ashburner M, Rosse C et al (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 25:1251–1255
29. Gibson F, Anderson L, Babnigg G et al (2008) Guidelines for reporting the use of gel electrophoresis in proteomics. Nat Biotechnol 26:863–864
30. Domann PJ, Akashi S, Barbas C et al (2010) Guidelines for reporting the use of capillary electrophoresis in proteomics. Nat Biotechnol 28:654–655
31. Jones AR, Carroll K, Knight D et al (2010) Guidelines for reporting the use of column chromatography in proteomics. Nat Biotechnol 28:654
32. Taylor CF, Binz PA, Aebersold R et al (2008) Guidelines for reporting the use of mass spectrometry in proteomics. Nat Biotechnol 26:860–861
33. Hoogland C, O'Gorman M, Bogard P et al (2010) Guidelines for reporting the use of gel image informatics in proteomics. Nat Biotechnol 28:655–656
34. Binz PA, Barkovich R, Beavis RC et al (2008) Guidelines for reporting the use of mass spectrometry informatics in proteomics. Nat Biotechnol 26:862
35. Orchard S, Salwinski L, Kerrien S et al (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat Biotechnol 25:894–898
36. Bourbeillon J, Orchard S, Benhar I et al (2010) Minimum information about a protein affinity reagent (MIAPAR). Nat Biotechnol 28:650–653
37. Orchard S, Al-Lazikani B, Bryant S et al (2011) Minimum information about a bioactive entity (MIABE). Nat Rev Drug Discov 10:661–669
38. Taylor CF, Field D, Sansone SA et al (2008) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotechnol 26:889–896
39. Martens L, Hermjakob H, Jones P et al (2005) PRIDE: the proteomics identifications database. Proteomics 5:3537–3545
40. Jones P, Cote RG, Martens L et al (2006) PRIDE: a public repository of protein and peptide identifications for the proteomics community. Nucleic Acids Res 34:D659–D663
41. Medina-Aunon JA, Martinez-Bartolome S, Lopez-Garcia MA et al (2011) The ProteoRed MIAPE web toolkit: a user-friendly framework to connect and share proteomics standards. Mol Cell Proteomics 10:M111.008334
42. Robin X, Hoogland C, Appel RD et al (2010) MIAPEGelDB, a web-based submission tool and public repository for MIAPE gel electrophoresis documents. Journal of proteomics 71:249–251
43. Martinez-Bartolome S, Medina-Aunon JA, Jones AR, et al (2010) Semi-automatic tool to describe, store and compare proteomics experiments based on MIAPE compliant reports. Proteomics 10:1256–1260
44. Martinez-Bartolome S, Blanco F, Albar JP (2010) Relevance of proteomics standards for the ProteoRed Spanish organization. Journal of proteomics 73:1061–1066
45. Paradela A, Escuredo PR, Albar JP (2006) Geographical focus. Proteomics initiatives in Spain: ProteoRed. Proteomics 6 Suppl 2:73–76