



Proteomics Standards Initiative

Standards in Proteomics

NIH workshop January 2005

Henning Hermjakob, European Bioinformatics Institute

HUPO Proteomics Standards Initiative

- **Founded at the HUPO meeting in Washington, April 28-29, 2002**
- **Major workshops:**
 - **Cambridge, UK, October 2002**
 - **Cambridge, UK, January 2003**
 - **Nice, France, April 2004**
 - **Siena, Italy, Spring 2005 (Planned)**
- **Satellite meetings to HUPO congresses**



HUPO Proteomics Standards Initiative

- Develop data *format* standards
- Data *representation* and *annotation* standards
- Involve data producers, database providers, software producers, publishers



PSI work groups

PSI-GPS

General Proteomics Schema

PSI-MI

**Molecular
Interactions**

PSI-MS

**Mass
Spectrometry**



MGED collaboration

FG-OM

Functional Genomics Experiment model

PSI-GPS

General Proteomics Schema

PSI-MI

Molecular
Interactions

PSI-MS

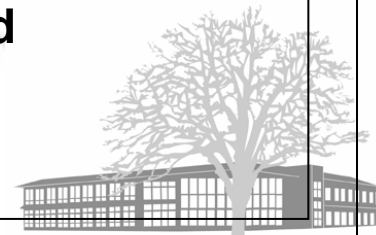
Mass
Spectrometry

MGED

MIAME

MAGE-OM

Microarray
Standard



PSI work groups: MI

FG-OM

Functional Genomics Experiment model

PSI-GPS

General Proteomics Schema

PSI-MI

Molecular
Interactions

PSI-MS

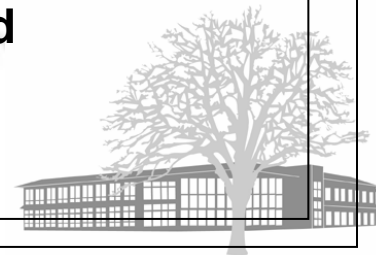
Mass
Spectrometry

MGED

MIAME

MAGE-OM

Microarray
Standard



PSI-MI XML format

- **Community standard for Molecular Interactions**
- **XML schema and detailed controlled vocabularies**
- ***Exchange* format, not internal format**
- **Jointly developed by major data providers:**
BIND, CellZome, DIP, GSK, HPRD, Hybrigenics, IntAct, MINT, MIPS, Serono, U. Bielefeld, U. Bordeaux, U. Cambridge, and others
- **Version 1.0 published in February 2004**

The HUPO PSI Molecular Interaction Format - A community standard for the representation of protein interaction data.

Henning Hermjakob et al, Nature Biotechnology 2004, 22, 176-183.



PSI-MI controlled vocabularies

- **PSI develops not only formats, but also controlled vocabularies/ontologies where necessary**
- **Example: > 20 ways to write:**
yeast two hybrid, Y2H, 2H, yeast-two-hybrid, two-hybrid, ...
- **PSI-MI CVs for interaction type, sequence feature type, feature detection, participant detection, and interaction detection.**
- **Ca. 200 terms, fully defined and cross-referenced**
- **GO format**



PSI-MI XML format support

- **Data**
 - DIP, MINT, IntAct, Hybrigenics, HPRD, BIND, MIPS, ...
- **Tools**
 - **Viewer**
 - Cytoscape (MSKCC, ISB, Whitehead)
 - PIMWalker (Hybrigenics)
 - ProViz (U. Bordeaux)
 - **Converter**
 - Tabular ↔ PSI MI (MINT)
 - PSI MI ↔ HTML (PSI)
 - **Database**
 - cPath (MSKCC)



PSI-MI XML benefits

- **Collecting and combining data from different sources has become easier.**
- **Standardized annotation through PSI-MI ontologies**
- **Tools from different organizations can be chained, e.g. analysis of IntAct data in Cytoscape**





- IntAct Home
- Search
- Aims
- Members
- Documentation
- Software
- Data
- Status
- Statistics
- Mailing Lists
- Publications
- Acknowledgements

IntAct Project

Search IntAct:

◆

◆ Examples:

- ◇ Gene name: Ism7
- ◇ IntAct Ac: EBI-141
- ◇ SPTR Ac: Q08162
- ◇ SPTR Id: rr44_yeast
- ◇ InterPro Ac: IPR001900
- ◇ GO Id: GO:0000176

The IntAct project aims are to:

- ◆ define a standard for the representation and annotation of protein-protein interaction data
- ◆ provide a public repository
- ◆ populate the repository with experimental data from project partners and curated literature data
- ◆ provide modular analysis tools
- ◆ provide portable versions of the software to allow installation of local IntAct nodes

Members of the IntAct consortium:

News

July 1, 2004:
New datasets:
Blagoev et al.

Next release will be July 28, 2004, 3 work days early due to the ISMB conference. We expect the server to be down from 10:00-10:30 British Summer Time.

June 1, 2004:
New datasets:
Bouwmeester et al,
Tewari et al,
Lachner et al,
Lehner et al.

May 1, 2004:
Statistics added.

C. elegans interactome datasets added:





HierarchView?

Interactor ?

EBI-141

Search Add

Graph ?

Expand

Clicking on a node of the graph will :

center the view ?

add a network ?

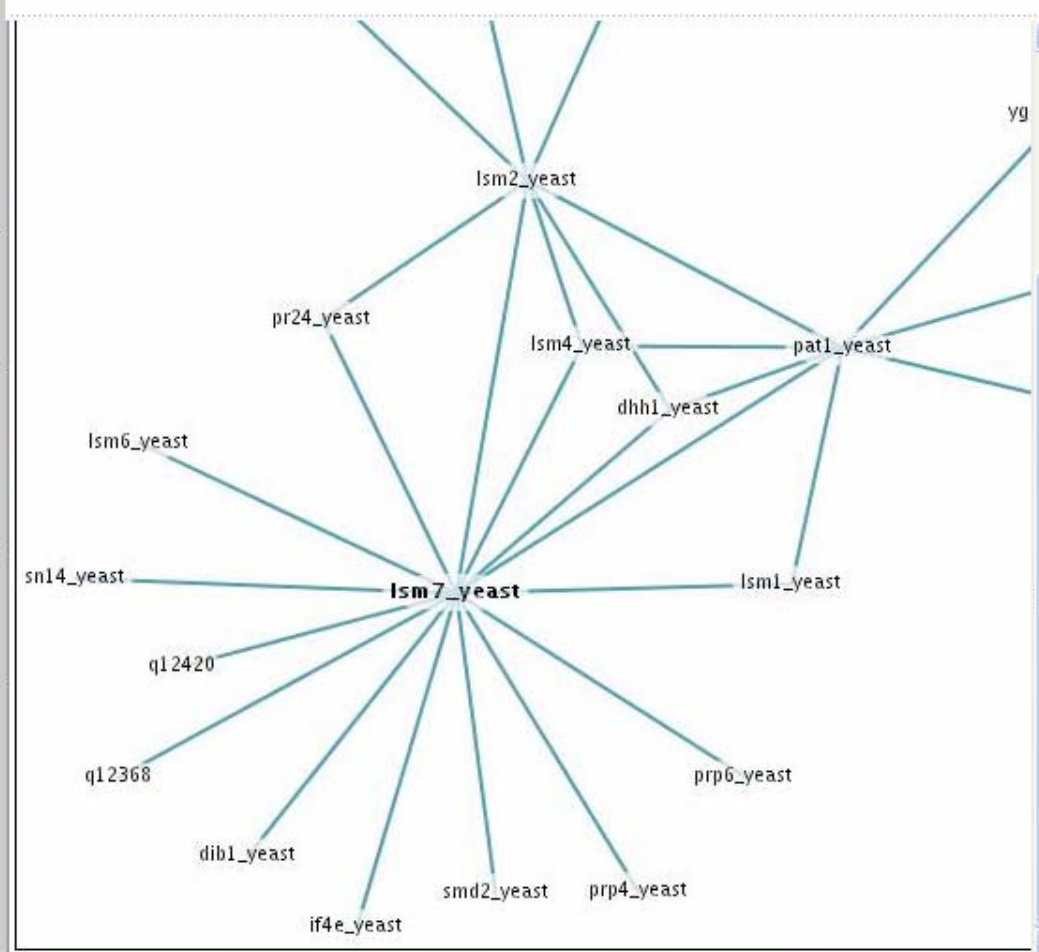
PSI-XML ?

Download

Links

[IntAct home](#)

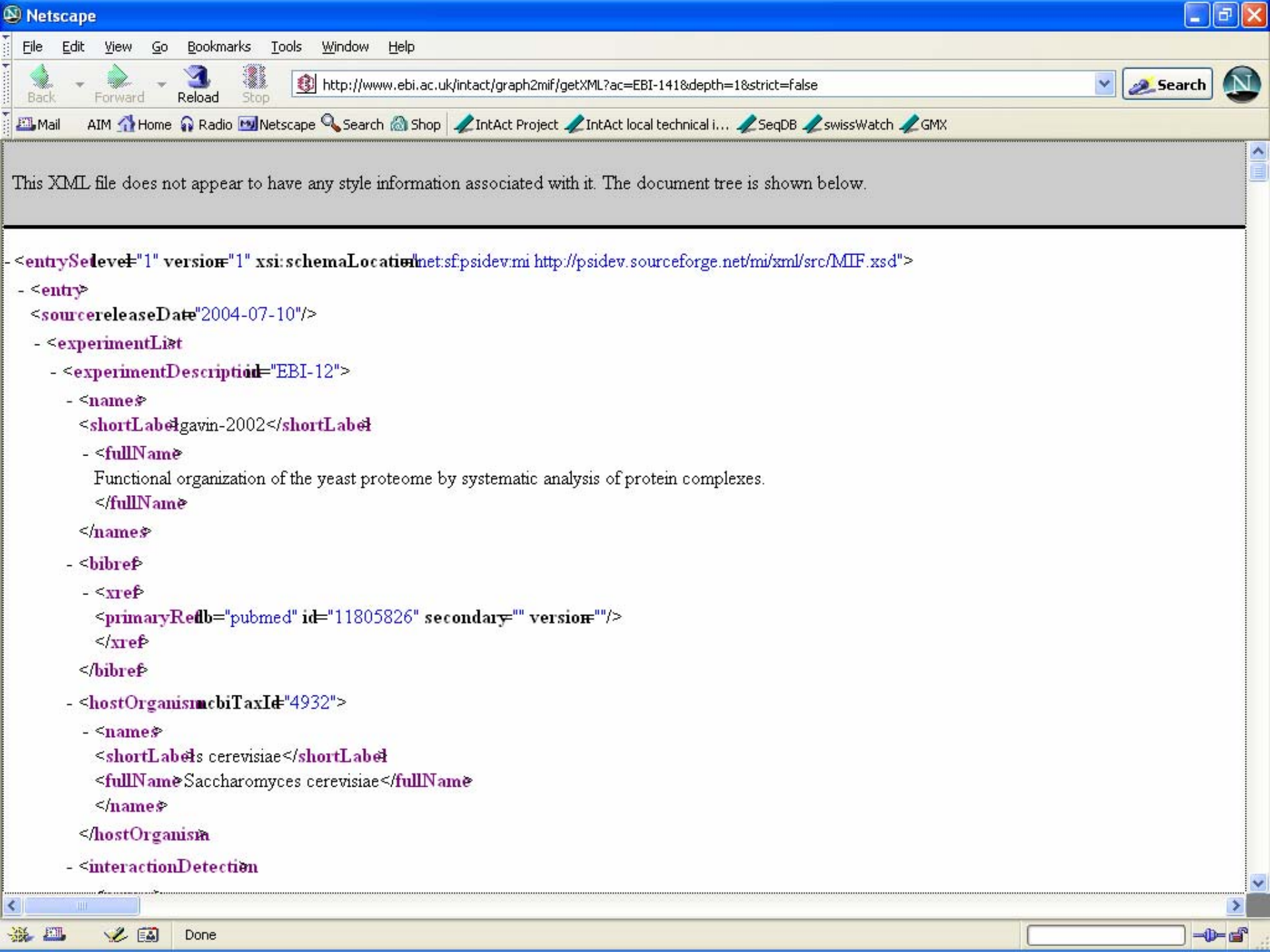
[Back to search](#)



Existing highlight source for the central protein(s).

ID	Description	Use
GO:0005688	C:snRNP U6	▲▼
GO:0017070	F:U6 snRNA binding activity	▲▼
GO:0006402	P:mRNA catabolism	▲▼
GO:0006371	P:mRNA splicing	▲▼
GO:0000398	P:nuclear mRNA splicing, via s	▲▼





This XML file does not appear to have any style information associated with it. The document tree is shown below.

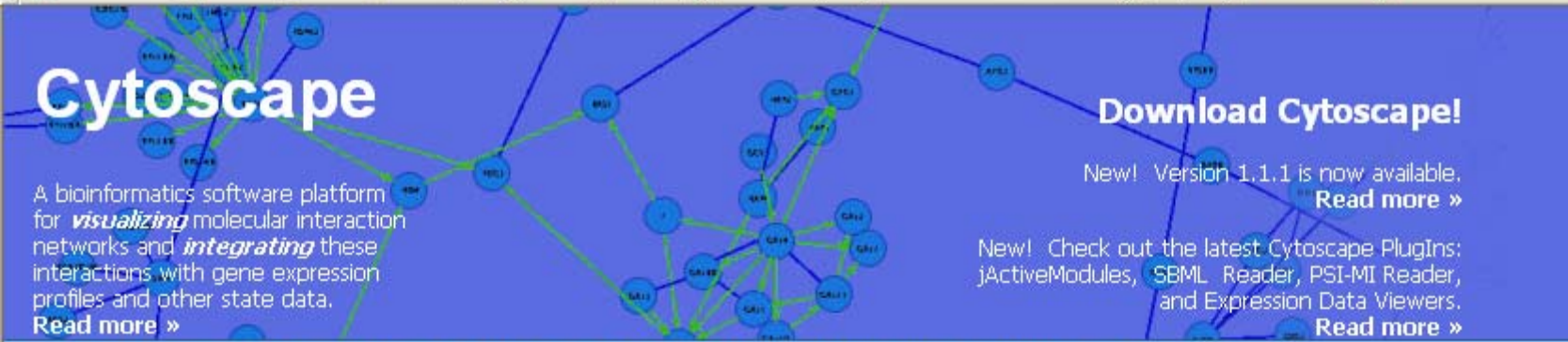
```
<entry level="1" version="1" xsi:schemaLocation="http://psidev.mif.sf.net/mif http://psidev.sourceforge.net/mif/xml/src/MIF.xsd">
- <entry>
  <source releaseDate="2004-07-10"/>
  - <experimentList>
    - <experimentDescription id="EBI-12">
      - <name>
        <shortLabel>gavin-2002</shortLabel>
        - <fullName>
          Functional organization of the yeast proteome by systematic analysis of protein complexes.
        </fullName>
      </name>
    - <bibref>
      - <xref>
        <primaryRef db="pubmed" id="11805826" secondary="" version="">
        </xref>
      </bibref>
    - <hostOrganism>
      - <name>
        <shortLabel>S cerevisiae</shortLabel>
        <fullName>Saccharomyces cerevisiae</fullName>
      </name>
    </hostOrganism>
  - <interactionDetection>
```

Cytoscape: Analyzing and Visualizing Biological Network Data - Netscape

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop <http://www.cytoscape.org/>

Mail AIM Home Radio Netscape Search Shop IntAct Project IntAct local technical i... SeqDB swissWatch GMX



Cytoscape

A bioinformatics software platform for *visualizing* molecular interaction networks and *integrating* these interactions with gene expression profiles and other state data. [Read more »](#)

Download Cytoscape!
New! Version 1.1.1 is now available. [Read more »](#)

New! Check out the latest Cytoscape PlugIns: jActiveModules, SBML Reader, PSI-MI Reader, and Expression Data Viewers. [Read more »](#)

Cytoscape: Features | Screenshots | Dev team | PlugIns | Download | Report a Bug

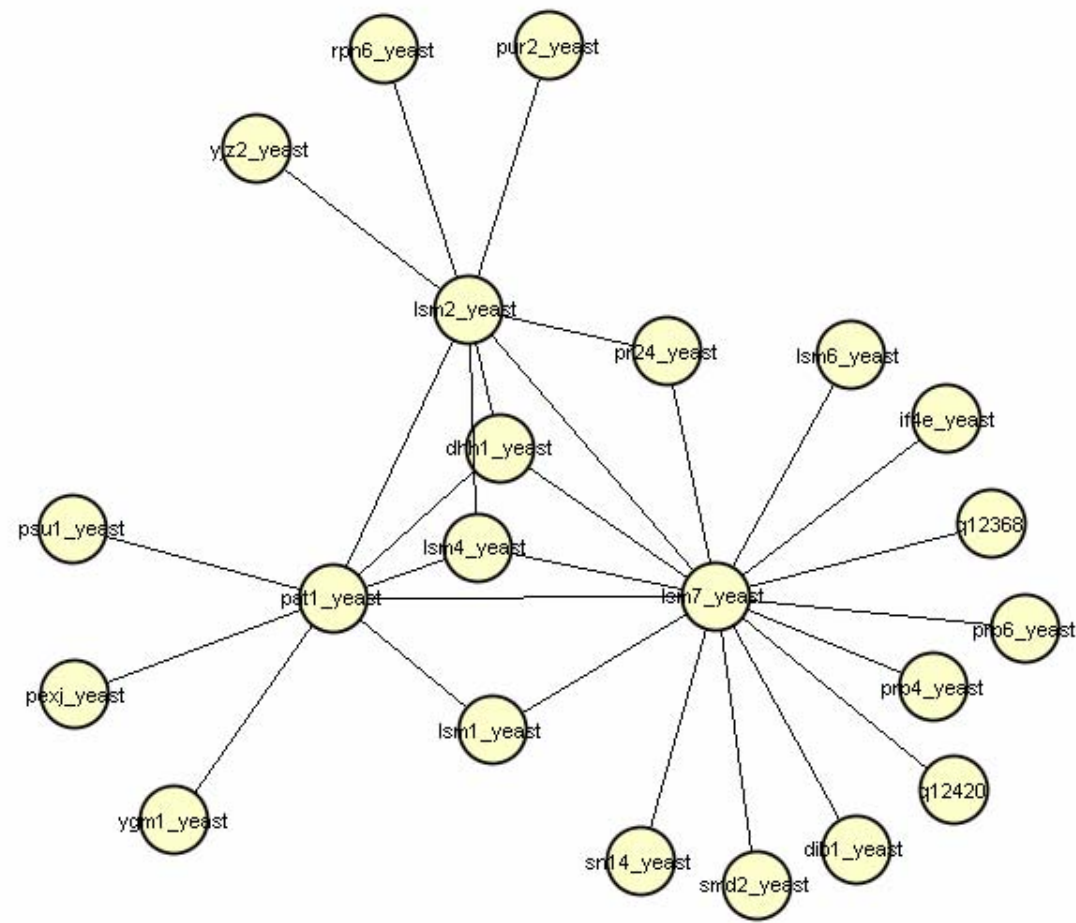
<p>Download Cytoscape Version 1.1.1 Requires Java 1.4.1</p>	<p>Download Source Version 1.1.1 Source code is written for Java 1.4.1</p>	<p>Click here to run Cytoscape from the web! (if you already have Java Web Start.)</p>	<p>Online Tutorial The tutorial uses Java Web Start.</p>	<p>Manual PDF format, explains all basic features of Cytoscape. Get Acrobat reader</p>
--	---	---	---	---



- Import PSI-MI Data from file...
- Import PSI-MI Data from web...
- Import GEO SOFT Data from file...
- Import GEO SOFT Data from web...
- Export Data to PSI-MI XML Format...

Input

Enter URL for PSI Document:



PSI-MI XML evolution

- Well-defined schema evolution, yearly releases
- Next release January 2005, version 2.0
 - Better identifier handling
 - New interactor types: DNA, RNA, small molecules
 - Beta: <http://psidev.sourceforge.net/mi/rel2/doc/>
- Well-defined evolution of controlled vocabularies
 - Modeled on GO procedures
 - Requests mailing list
 - Editorial board



The IMEx consortium

- *International Molecular-Interaction Exchange* consortium
- **BIND, DIP, IntAct, MINT, MIPS** will establish an exchange of user-submitted data in PSI-MI format from beginning of 2005 onwards to provide a network of stable, comprehensive resources for molecular interaction data
- **Aims:**
 - Consistent body of public data
 - Avoid redundant curation



PSI work groups: MS

FG-OM

Functional Genomics Experiment model

PSI-GPS

General Proteomics Schema

PSI-MI

Molecular
Interactions

PSI-MS

Mass
Spectrometry

MGED

MIAME

MAGE-OM

Microarray
Standard



Mass spectrometry: PSI-MS

- **mzData format as common instrument output format**
 - Format beta version accepted in Nice, April 2004
 - EBI workshop July 2004
 - Version 1.05 released January 4, 2005
 - Controlled vocabularies developed jointly with ASTM
 - Key concept:
Request direct vendor support to avoid version problems due to vendor API changes



Announced mzData support

- **SIB: Aldente (next release)**
- **GeneBio: Phenyx (next release)**
- **Matrix Science: Mascot (in release 2.1)**
- **Bruker Daltonics (next release)**
- **Kratos (next release)**
- **Thermo Electron (next release)**
- **Agilent, ABI, (exploratory phase)**
- **Protagen**
- **EBI/University of Gent: PRIDE: PRoteomics IDentifications database (February 2005)**
- **Randall Julian from Eli Lilly has a mzData version of X! Tandem**
- **Bioinformatics Solutions Inc.: Peaks**

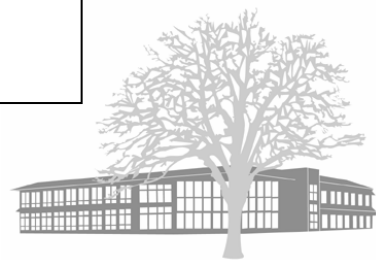
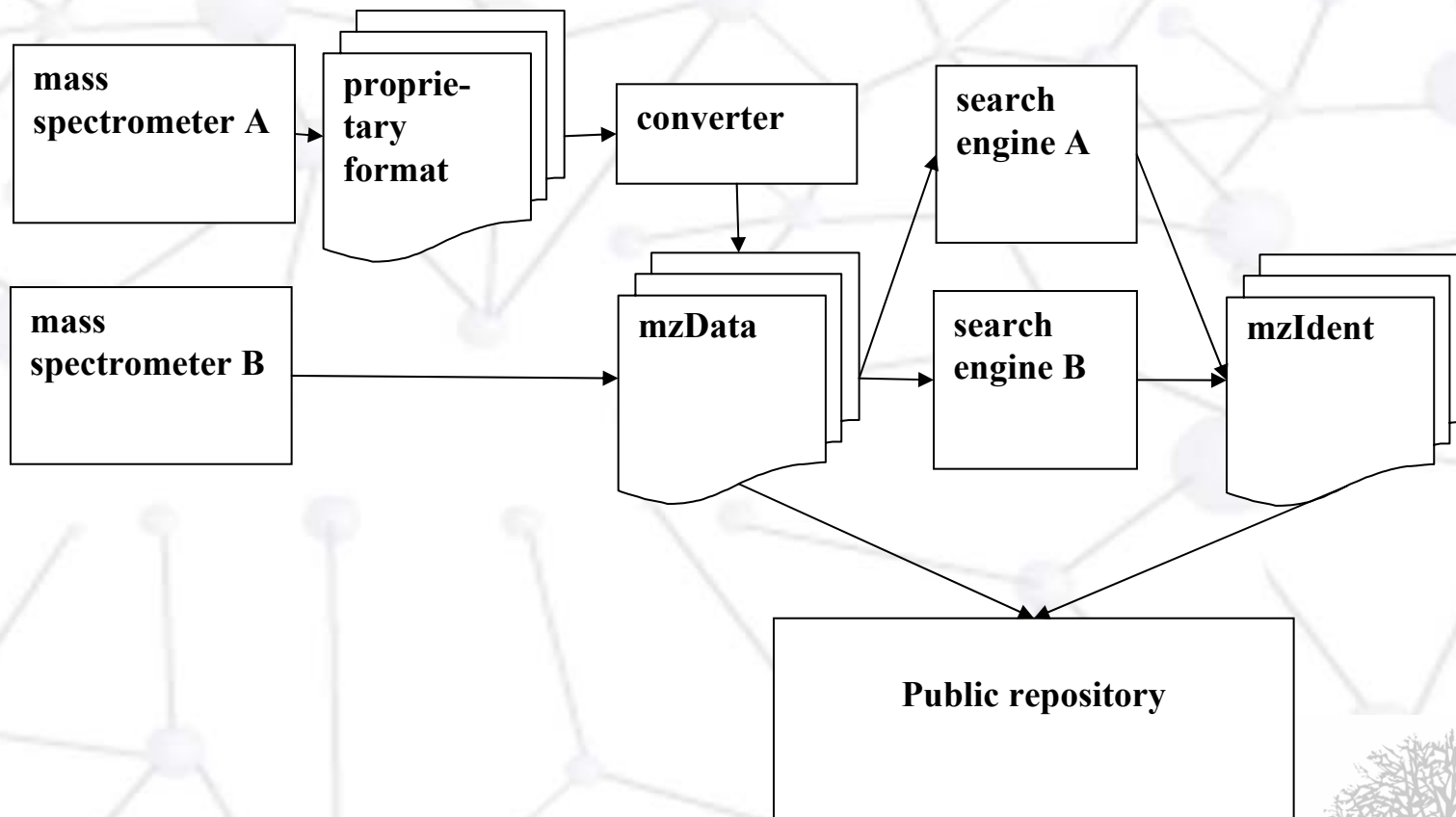


Mass spectrometry: PSI-MS

- **mzIdent format as common search engine output format**
 - **Suggested in Nice, April 2004**
 - **Aim: Facilitate comparison and archiving of search engine output, in particular in comparative projects like the HUPO PPP**
 - **Current beta version released October 2004**

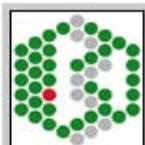


PSI-MS based data flow





PRIDE PRoteomics IDentifications database



Log in to PRIDE: Username: Password:

Search Results

The search was performed on sample, with 'unstimulated human platelets' as the parameter.

Experiments

- [COFRADIC methionine proteome of unstimulated human blood platelets](#)
- [COFRADIC cysteine proteome of unstimulated human blood platelets](#)
- [COFRADIC aminoterminal proteome of unstimulated human blood platelets](#)

COFRADIC methionine proteome of unstimulated human blood platelets

COFRADIC methionine proteome of unstimulated human blood platelets

Accession	Contact	Reference(s)	Short	Description	Location
-----------	---------	--------------	-------	-------------	----------

- Links
- [Home](#)
 - [Search](#)
 - [PRIDE](#)
 - [Register](#)
 - [About PRIDE \(Project Home\)](#)

Medium term vision

- **Establish data exchange and collaboration similar to PSI-MI/IMEx between PeptideAtlas, GPMDB, PRIDE, ...**
- **Provide a set of compatible, synchronized, public resources for protein identification data**



PSI work groups: GPS

FG-OM

Functional Genomics Experiment model

PSI-GPS

General Proteomics Schema

PSI-MI

Molecular
Interactions

PSI-MS

Mass
Spectrometry

MGED

MIAME

MAGE-OM

Microarray
Standard



Global Proteomics Standards (GPS)

- Capture all relevant aspects of a proteomics experiment
- Iterative, long-term development
- Building on PEDRo* work
- Intensively discussed at PSI meeting in Nice, April 2004



Taylor et al: A systematic approach to modeling, capturing, and disseminating proteomics experimental data. Nat Biotechnol. 2003 Mar;21(3):247-54.



GPS documents

- **Design Principles**
 - Scope and purpose of GPS
- **MIAPE**
 - Minimum data requirements, similar to MIAME
 - “Parent” document for all experiment types
 - “Modules” - technology-specific guidelines
- **PSI-OM**
 - UML model and documentation
- **PSI-ML**
 - XML schema



GPS status

- **MIAPE parent document and MS guidelines will be released in spring**
- **Gel-based and MS-informatics documents in advanced stage**
- **Joint ontology project initiated with MGED**
- **Next milestone: PSI spring workshop, Siena**



Summary: PSI process

- **Open, community-based process**
- **Visibility through meetings, web site and journal articles**
- **Active invitations to organisations and individuals active in an area**
- **Open for participation**
- **Very little formal structure**



Summary: PSI process problems

- Lengthy process of consensus-building
- Might be overtaken by organisations just “going ahead”
- New participants require frequent re-introductions
- Funding problem:
 - PSI activities “hobby” for most participants
 - Difficult to find funding for truly international activity
 - Important to maintain organisational independence



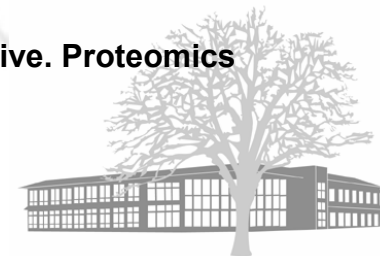
Summary: PSI success

- **PSI-MI now de facto standard for molecular interaction data**
- **PSI-MS on the verge of broad vendor acceptance**
- **Intensive community discussion process with broad participation from academic and commercial sector**
- **Strong support by journals and scientific community**



Publications

- **The HUPO PSI Molecular Interaction Format - A community standard for the representation of protein interaction data.**
Henning Hermjakob, Luisa Montecchi-Palazzi, Gary Bader, Jérôme Wojcik, Lukasz Salwinski, Arnaud Ceol, Susan Moore, Sandra Orchard, Ugis Sarkans, Christian von Mering, Bernd Roechert, Sylvain Poux, Eva Jung, Henning Mersch, Paul Kersey, Michael Lappe, Yixue Lix, Rong Zeng, Debashis Rana, Macha Nikolski, Holger Husi, Christine Brun, K. Shanker, Seth G.N. Grant, Chris Sander, Peer Bork, Weimin Zhu, Akhilesh Pandey, Alvis Brazma, Bernard Jacq, Marc Vidal, David Sherman, Pierre Legrain, Gianni Cesareni, Ioannis Xenarios, David Eisenberg, Boris Steipe, Chris Hogue, Rolf Apweiler.
Nature Biotechnology 2004, 22, 176-183.
- **Orchard, S. et al. Common interchange standards for proteomics data: Public availability of tools and schema.** *Proteomics* 4, 490-1 (2004).
- **Orchard, S., Hermjakob, H. & Apweiler, R. Proteomics and Data Standardisation.** *Drug Discovery Today: BIOSILICO* 2, 91-93 (2004).
- **Hermjakob, H. & Apweiler, R. Maximising proteomics data for the scientific community.** *European Pharmaceutical Review* 9, 23-28 (2004).
- **Orchard, S., Kersey, P., Hermjakob, H. & Apweiler, R. The HUPO Proteomics Standards Initiative meeting: towards common standards for exchanging proteomics data.** *Comparative and Functional Genomics* 4, 16-19 (2003).
- **Orchard, S., Hermjakob, H. & Apweiler, R. The proteomics standards initiative.** *Proteomics* 3, 1374-6 (2003).



THANKS:

•MGED:

- Alvis Brazma, EBI

•PRIDE:

- Phil Jones, EBI
- Lennart Martens, U Ghent

•PSI participants, in particular:

•PSI-MI:

- Luisa Montecchi-Palazzi, MINT
- Gary Bader, BIND/MSKCC
- Lukasz Salvinski, DIP

•PSI-GPS:

- Chris Taylor, EBI
- Pierre-Alain Binz, SIB

•PSI-MS:

- Weimin Zhu, EBI
- Randall Julian, Lilly

•Sponsors:

•EU

(QLRI-CT-2001-00015)

•Lilly, Novartis, Roche, Serono, Syngenta

•EBI Industry Group

•HUPO

•Samir Hanash

•Gil Omenn

•Rolf Apweiler

•YOU!

[Http://psidev.sf.net](http://psidev.sf.net)