

HUPO Proteomics Standards Initiative

Molecular Interactions

Henning Hermjakob

Edinburgh

October 2010

EMBL-EBI





HUPO Proteomics Standards Initiative



- **Mission Statement:**

The HUPO Proteomics Standards Initiative (PSI) defines community standards for data representation in proteomics to facilitate data comparison, exchange and verification.

- **Deliverables:**

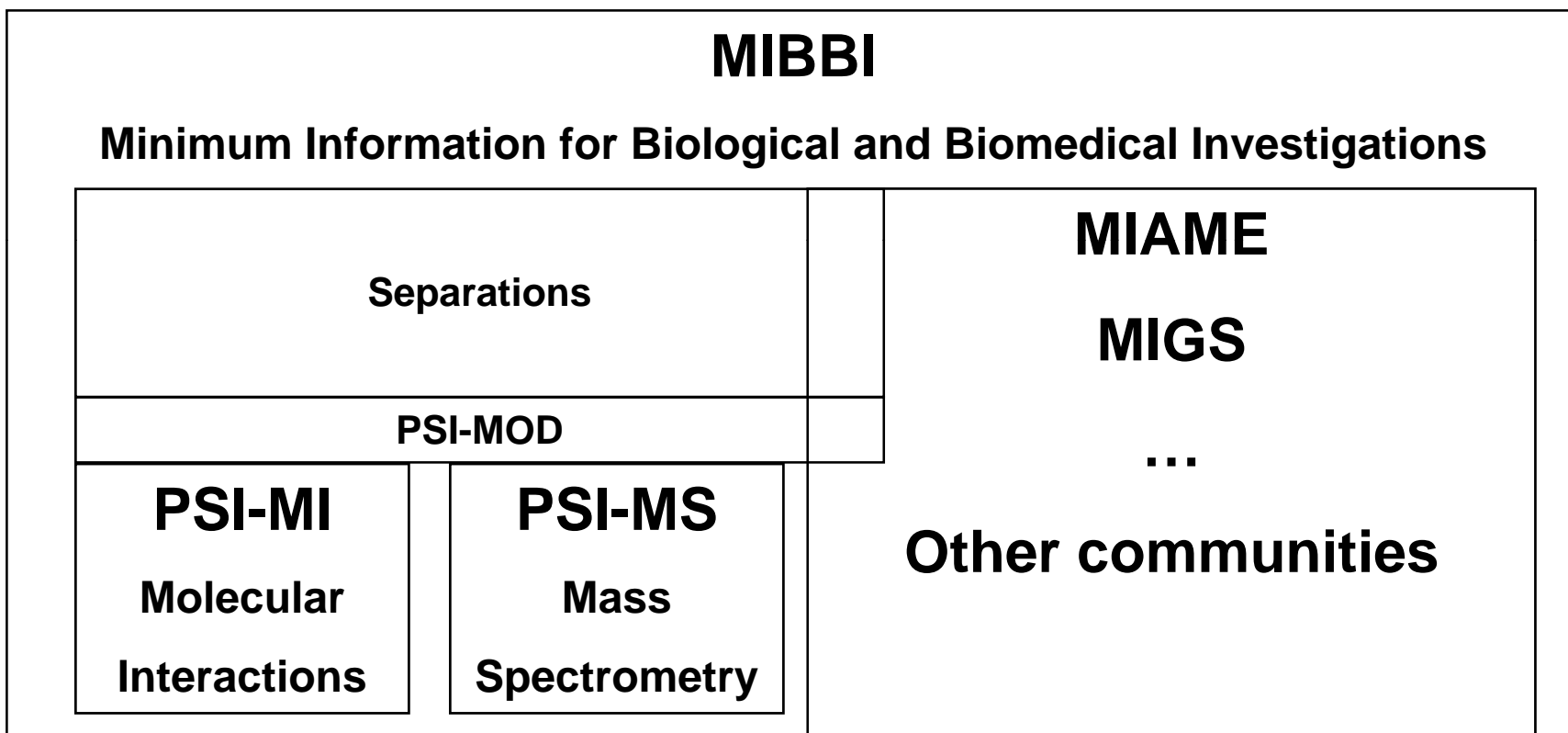
- Minimum information (MIAPE) specification: Format-independent specification of relevant data items
- Format: Usually an XML schema capable of representing the relevant Minimum Information, plus much more detailed data for the domain
- Controlled vocabulary: Usually an OBO-style hierarchical controlled vocabulary defining value ranges for attributes of the XML schema
- Databases and Tools: Foster software implementations to make the standards useful and practicable
- Community interaction to ensure deposition of data in public repositories

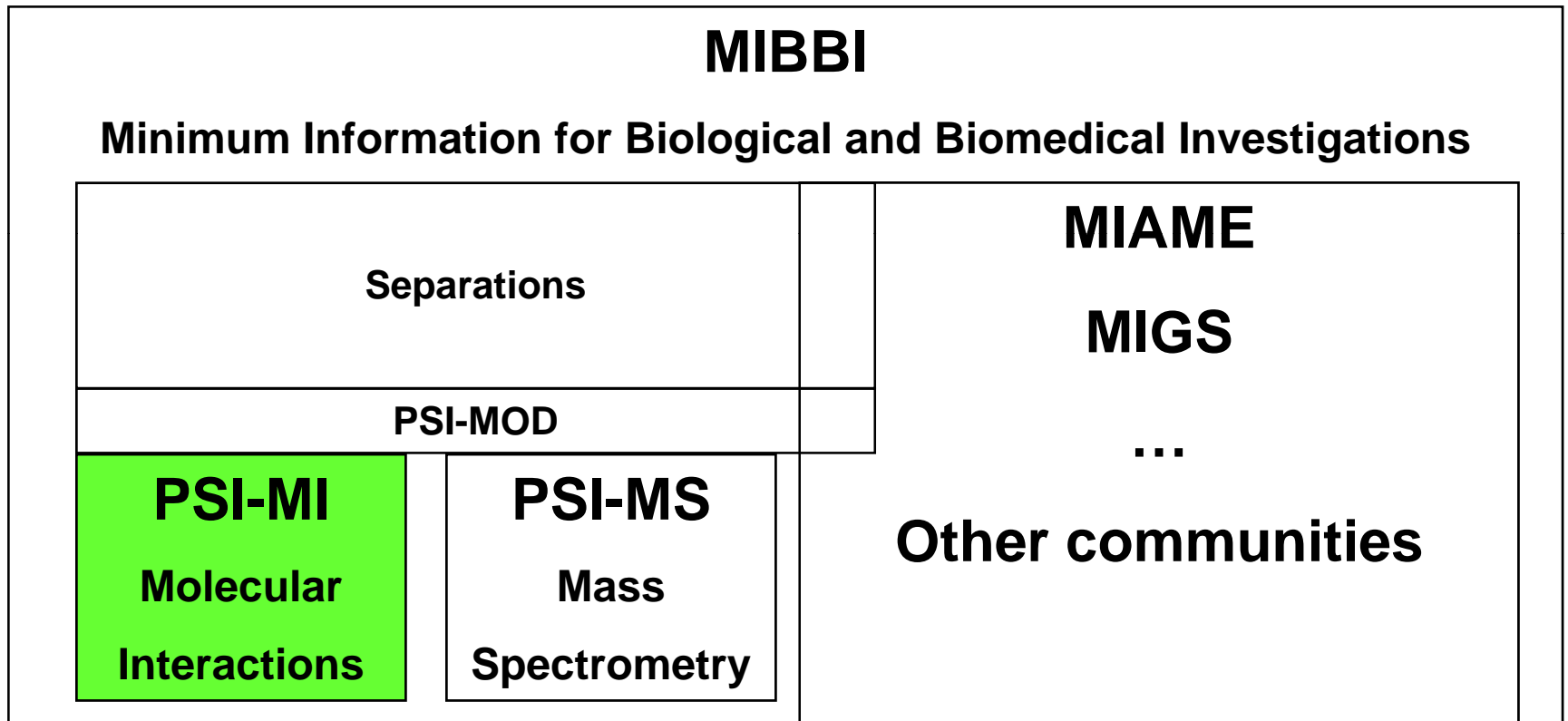


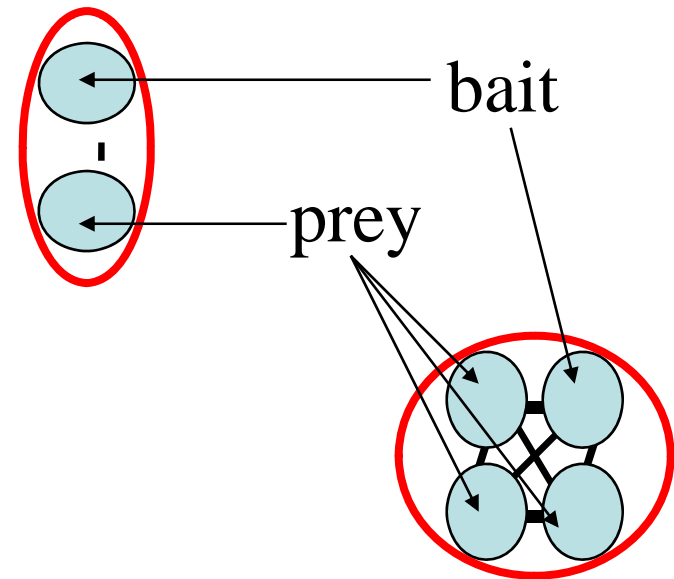
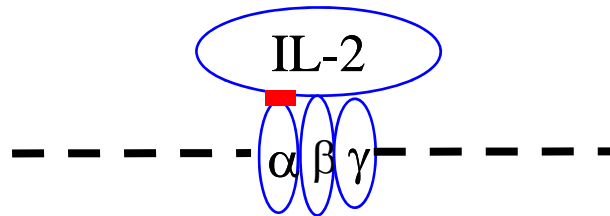
The PSI Document Process

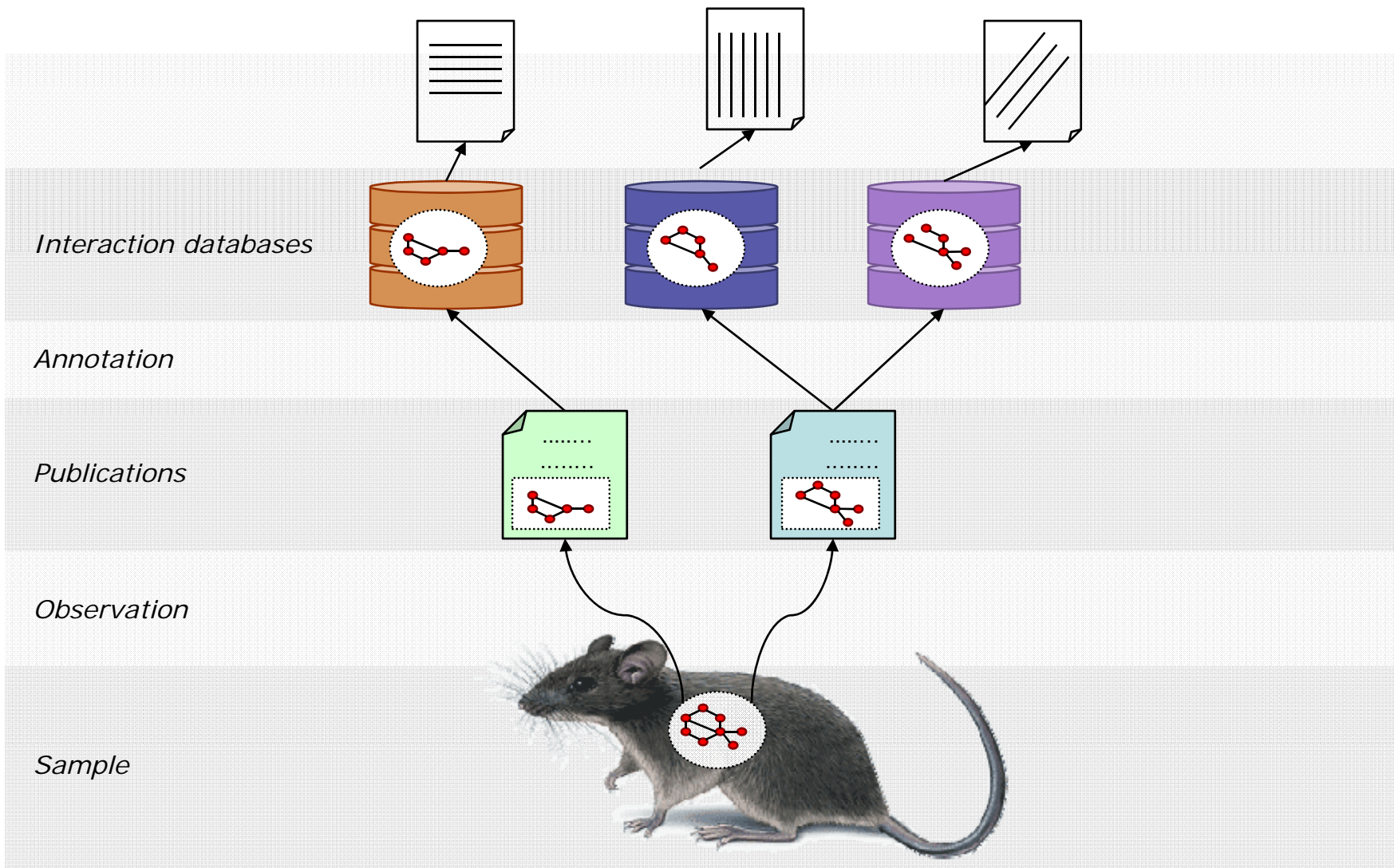


- **Standards influence data representation and software development, hence, ultimately, investment**
 - Need for transparent process
 - Need for community involvement
- **PSI has developed a well-defined document process**
 - Vizcaino, J.A., Martens, L., Hermjakob, H., Julian, R.K., Paton, N.W.: The PSI formal document process and its implementation on the PSI website. *Proteomics* 7 (2007) 2355-2357
 - A PSI standard document requires
 - Specification document
 - Instance documents
 - Public review period
 - Editorial review
 - Approval











The Molecular Interaction Standard (PSI-MI)



- **Community format for the representation of interaction data**
- **XML schema**
- **Version 1.0: February 2004**
 - The HUPO PSI Molecular Interaction Format - A community standard for the representation of protein interaction data. Hermjakob, H., et al: Nature Biotechnology 2004, 22, 176-183.
- **Version 2.5: October 2007**
 - Quantitative parameters
 - Additional interactor types: DNA, RNA, small molecules
 - Also includes simplified tabular representation
 - Kerrien S., et al: Broadening the Horizon - Level 2.5 of the HUPO-PSI Format for Molecular Interactions. BMC Biol. 2007 Oct 9;5(1):44
- **Now widely implemented by DIP, MINT, MIPS, IntAct, HPRD, ...; Cytoscape, Pathway commons, iRefIndex, ...**

<http://www.psidev.info/MI>



MIMIx



- Orchard, S. et al: The minimum information required for reporting a molecular interaction experiment (MIMIx).
Nat Biotechnol. 2007 Aug;25(8):894-8.
- Experiment
 - Host system: The host organism in which the interaction took place, identified by the NCBI taxonomy identifier. Example: Yeast (TaxID:4932) Further specification of cell line or tissue is recommended
 - Interaction detection method. Example: two hybrid (MI:0018)
 - Participant identification method Example: nucleotide sequence (MI:0078)
- Interaction
 - Participant list: The list of all molecules participating in the interaction. The list can contain any number of elements. Each molecule should be characterized by:
 - Database. Example: UniProt (MI:0486)
 - Accession number from that database. Example: P48551
 - Version number (optional)
 - The common name of the molecule used in the manuscript. Example: IFN-alphaRbetaL
 -



Example



Experiment (100 interactions)

Accession: EBI-2509892
Name: sowa-2009-1
Host organism: [human-293t](#)
Interaction Detection Method: [anti tag coip](#)
Participant Identification Method: [fingerprinting](#)
Cross References:

imex	IM-12079	-	imex-primary
pubmed	19615732	-	primary-reference
newt	9606	human	target-species

Annotations:

curation request	Harper Wade: 28th Oct 2009, Department of Pathology, Harvard Medical School, 77 Avenue Louis Pasteur, NRB 940, Boston, MA 02115
confidence-mapping	Interactors scored using the D score which incorporates the uniqueness, the abundance of the interactor (TSC), and the reproducibility of the interaction to assign a score to each protein within each IP (the same protein in two different IPs may have distinct D scores since its TSC and reproducibility could differ). In this way, the highest scores are given to proteins in each IP that are found rarely, found in duplicate runs, and have high TSCs-all characteristics of proteins that would be considered candidate interactors. A global D score threshold (DT) is determined and all raw D scores (DR) are normalized to this value, producing DN scores.
full coverage	Only protein-protein interactions
imex curation	-

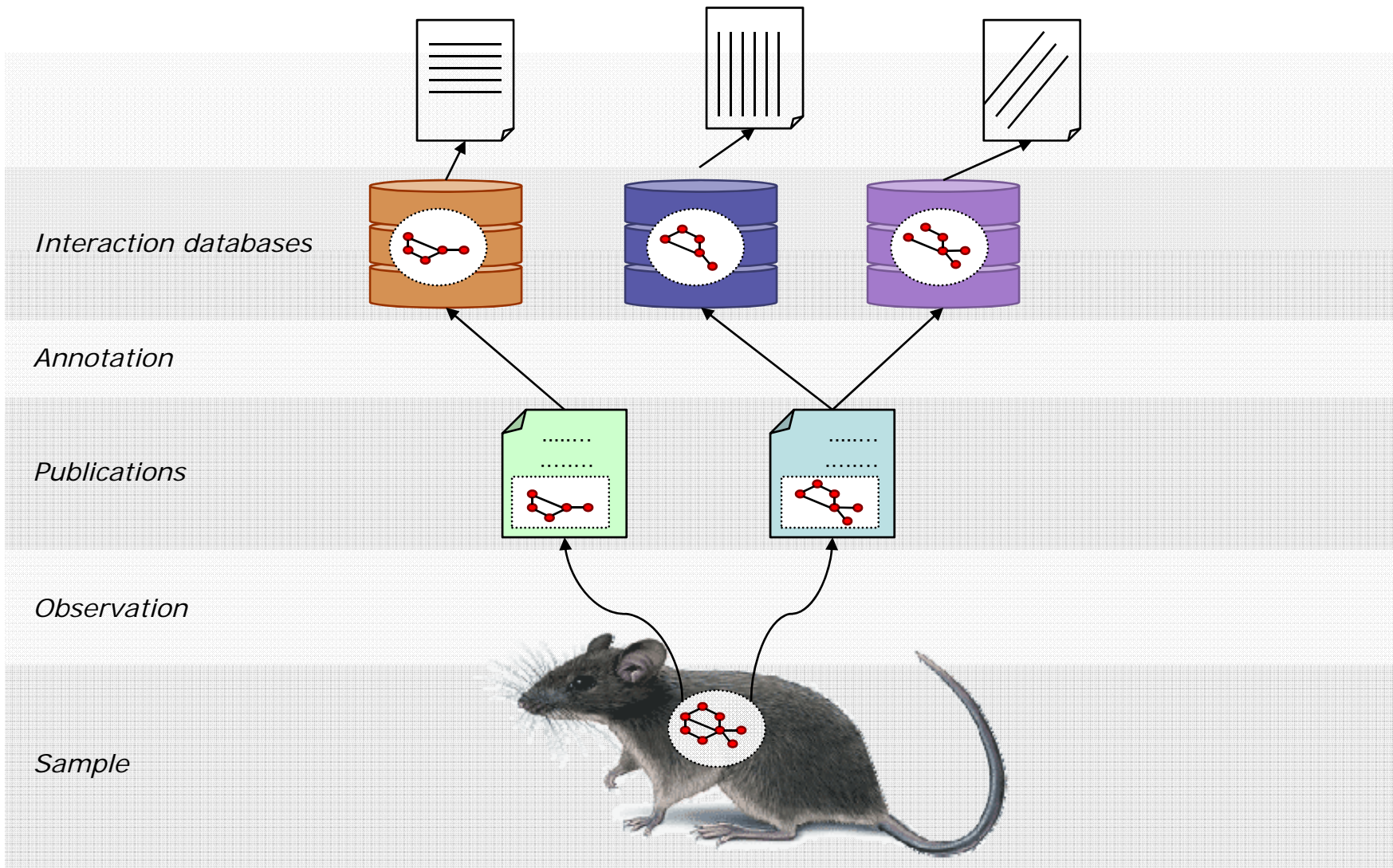
Publication

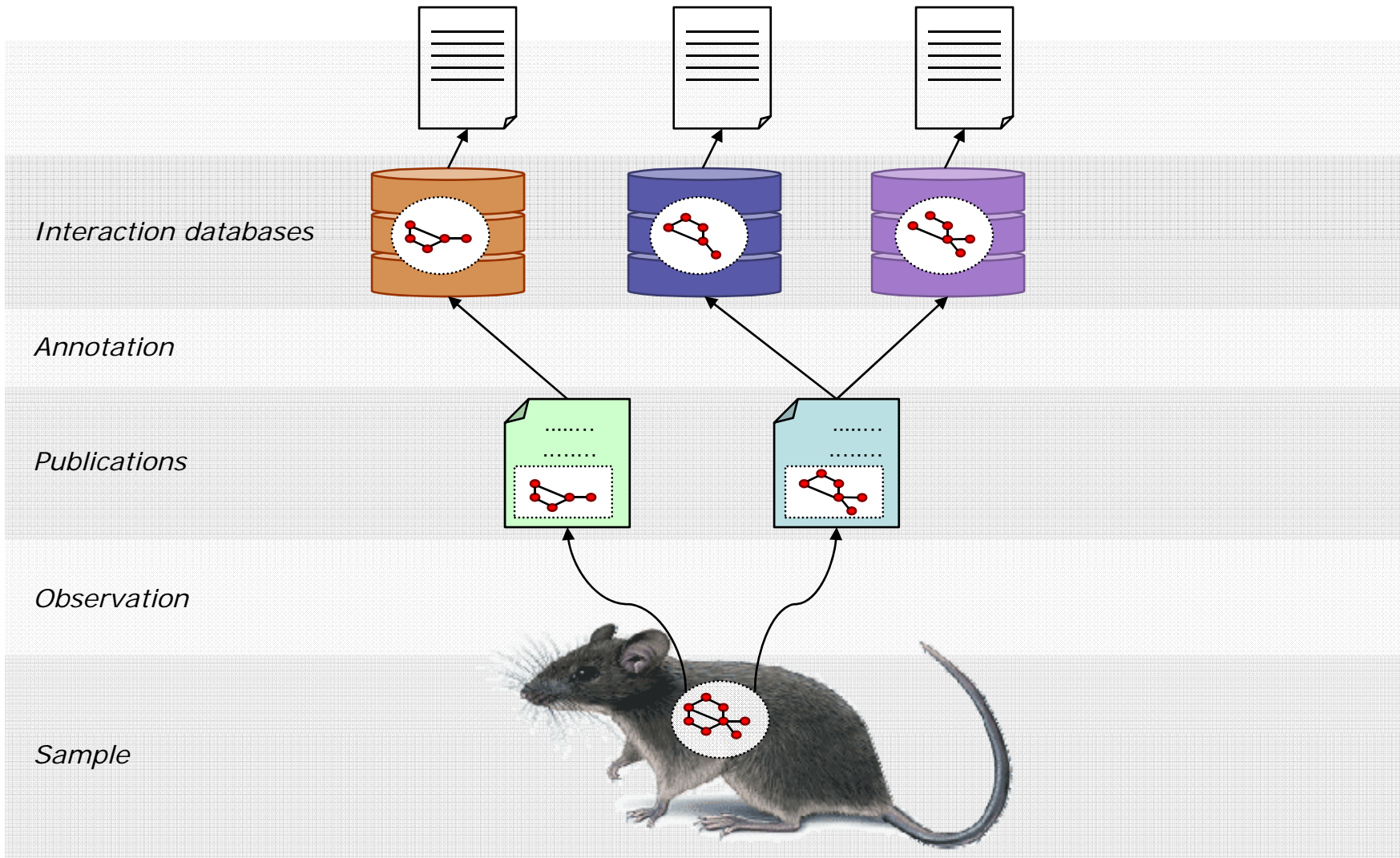
Author List: Sowa ME., Bennett EJ., Gygi SP., Harper JW.
Title: Defining the Human Deubiquitinating Enzyme Interaction Landscape
Journal: Cell (0092-8674)
Year of Publication: 2009
PubMed Id: 19615732
Cross References:

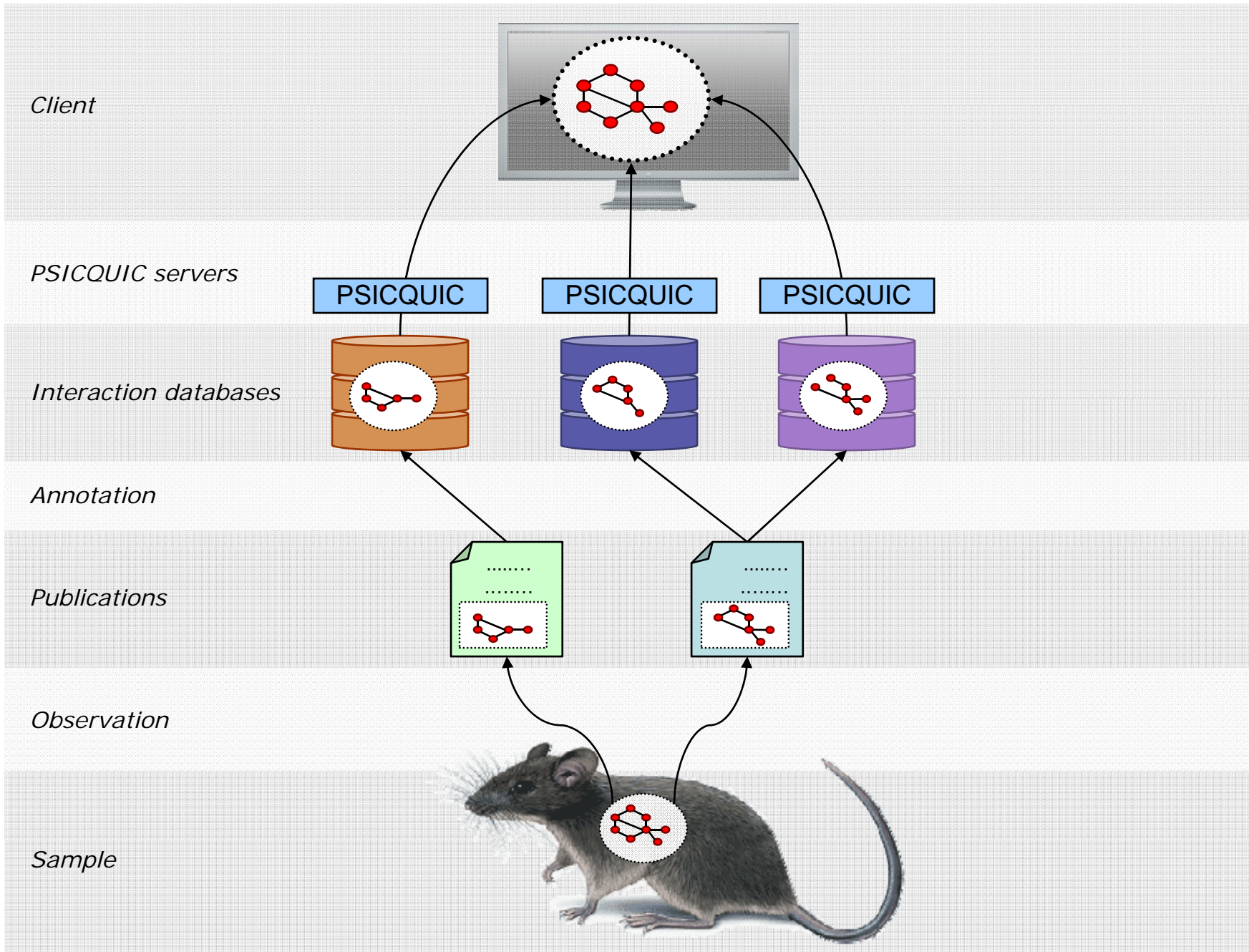
imex	IM-12079	-	imex-primary
----------------------	--------------------------	---	------------------------------

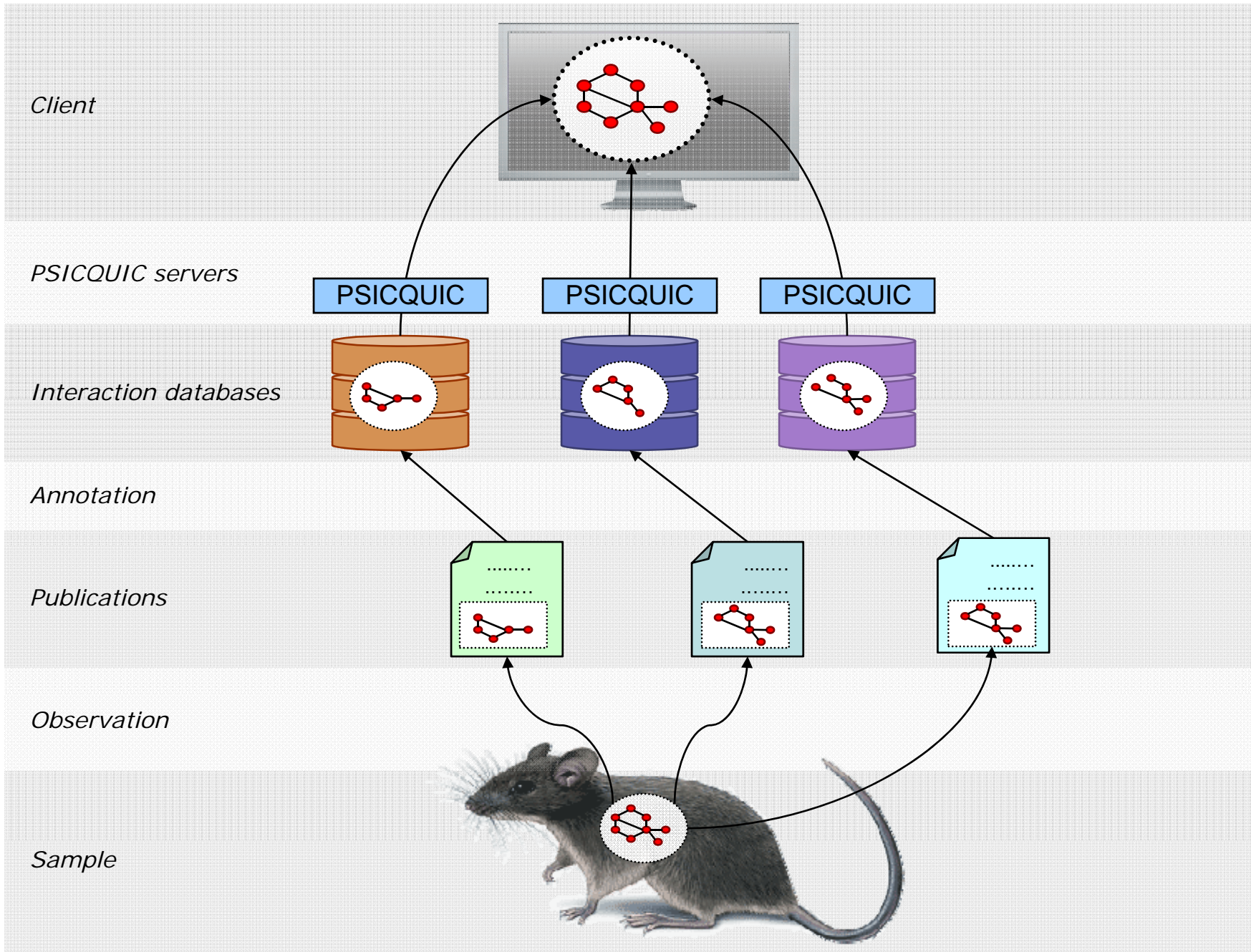
Annotations:

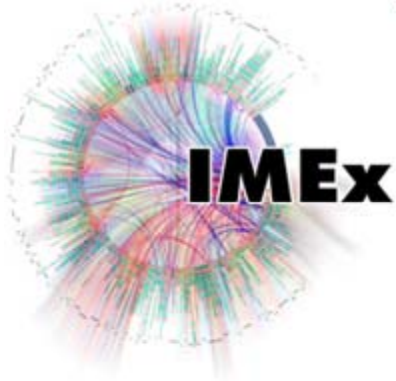
full coverage	Only protein-protein interactions
imex curation	-
last-imex-assigned	100











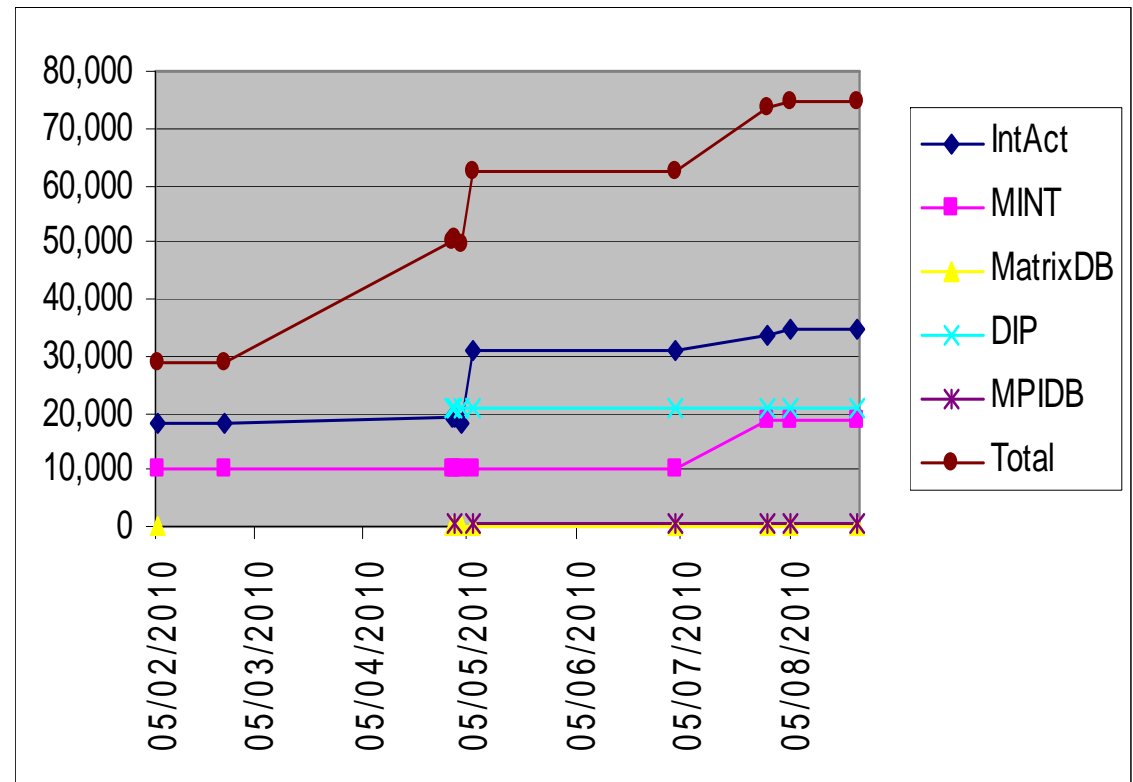
The International Molecular Exchange Consortium (IMEx)



- Major interaction databases are co-ordinating literature curation in PSI-MI format
 - DIP, Los Angeles, US
 - IntAct, Cambridge, UK
 - MINT, Rome, Italy
 - MatrixDB, Lyon, France
 - MPIDB, La Jolla, US
 - Molecular Connections Inc, India
 - BioGRID, Edinburgh, UK/Toronto, Canada (Observer)

- Aims:
 - Avoid redundant curation
 - Consistent body of public data
 - Detailed joint curation guidelines
 - Provide a network of stable, comprehensive resources for molecular interaction data

- Now in production mode (since 2/2010)





Acknowledgements



- **PSI MI**

- Samuel Kerrien
- Bruno Aranda
- Sandra Orchard
- Lukasz Salvinski, DIP
- Gianni Cesareni, MINT
- Johannes Goll, MPIDB
- Silvie Ricard-Blum, MatrixDB
- Gary Bader, MSKCC

- **Editors**

- Mike Dunn, Proteomics
- Kathy Aschheim, NBT

- **Funding**

- EU FW6 Felics
- EU FW6 Enfin
- EU FW7 PSIMEx

- **All participants** of the HUPO Proteomics Standards Initiative, IMEx, and ProteomeXchange
- **All authors** who make their interaction data publicly accessible, preferably through deposition to <http://www.imexconsortium.org>

