EMBL- EBI
Wellcome Trust Genome Campus
Hinxton, Cambridge, CB10 1SD, UK

carcinoGENOMICS
a Project of the European Union

EMBL-EBI
European Bioinformatics Institute is an Outstation of the European Molecular Laboratory

Standards and infrastructure for managing experimental metadata

Philippe Rocca-Serra, Marco Brandizi, Eamonn Maguire, Nataliya Sklyar, Chris Taylor, Manon Delahaye, Richard Evans & Susanna-Assunta Sansone

Standards to report complex studies

Today's researchers can perform biological and biomedical studies where the same material is run through a wide range of assays, comprising several technologies such as genomics, transcriptomics, proteomics and metabol/nomics (hereafter referred as '**omics**'). To enable others to correctly interpret the complex data sets that result, and the conclusions drawn, it is necessary to provide contextualizing **experimental metadata** at an appropriate level of granularity; see Figure 1.

Standards initiatives normally cater to particular domains. However, several **synergistic standards activities** foster cross-domain harmonization of the three kinds of reporting standard (minimum information checklists, ontologies and file formats). Some 29 groups participate in the **MIBBI** project, which offers a one-stop shop for those exploring the range of extant 'minimum information' checklists, and which fosters integrative development [1]. More than 60 groups participate in the **OBO Foundry** [2], which coordinates the orthogonal development of ontologies such as **OBI** [3] for describing experimental (meta)data. And several groups participate in the development of **ISA-Tab**, a tabular framework for presenting experimental metadata [4] (analogous to **FuGE**, a generic data model to underpin various XML file formats [5]).

BioInvestigation Index and ISA-based infrastructure

We have developed an infrastructure that leverages the aforementioned synergistic standards to create a **common structured representation and storage mechanism** for **experimental metadata** from biological and biomedical investigations ranging from simple single-assay studies to complex, methodologically diverse multi-assay studies. The infrastructure's components include: a data capture and editing tool (**ISAcreator**); validator (**ISAvalidator**); database (**BioInvestigation Index**); and converter (**ISAconverter**); and a BioConductor analysis package (**R-ISApackage**).

The components (shown in Figure 2) are designed for local installation, and can work independently, or as unified system.

An instance of the ISA infrastructure has been deployed at EBI as a public prototype: After automatically converting ISA-Tab formatted metadata into the various required formats, it dispatches data files to ArrayExpress (as MAGE-Tab for transcriptomics), PRIDE (as PRIDE-XML for proteomics) and ENA-Reads (as SRA-XML for high-throughput sequencing).
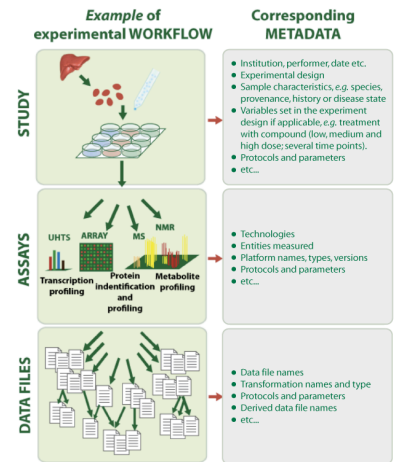

Figure 1.

Go to: http//www.ebi.ac.uk/bioinvindex to view the EBI instance and to http://isatab.sf.net to download the components for your local use.
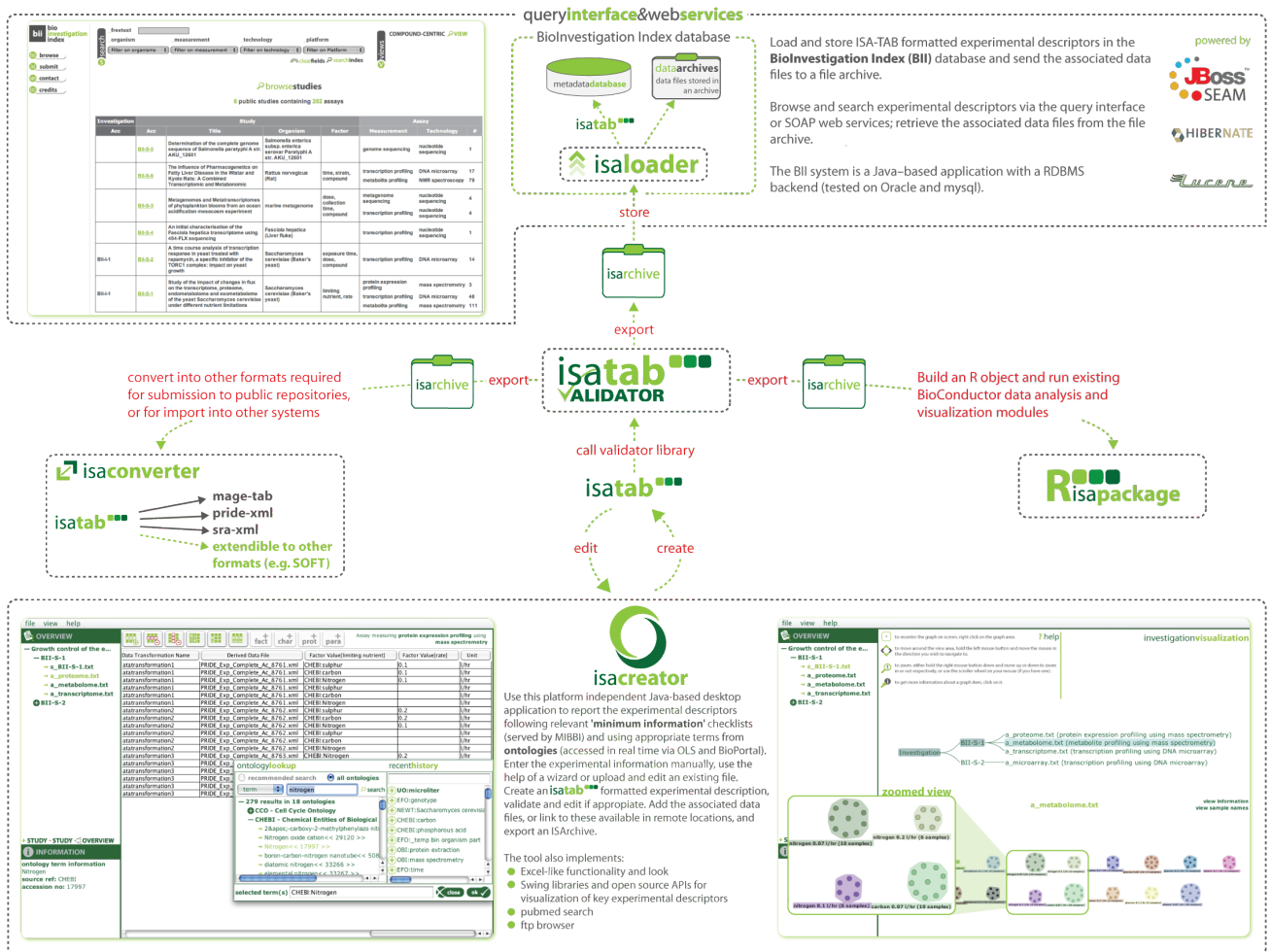

Figure 2.

References

1. Taylor CF, Field D, Sansone SA,... Rocca-Serra P et al. (2008) The MIBBI Project. *Nature Biotechnology* Aug;26(8):889-896. http://http://www.mibbi.org
2. Smith B, Ashburner M, Rosse C,... Rocca-Serra P, ...Sansone SA et al. (2007) The OBO Foundry. *Nature Biotechnology* Nov;25(11):1251-5. http://www.obofoundry.org
3. Ontology for Biomedical Investigations (OBI): http://obi-ontology.org
4. Sansone SA, Rocca-Serra P, Brandizi M,... Taylor CF et al. (2008) The First MGED RSBI (ISA-TAB) Workshop. *OMICS.* Jun;12(2):143-9. http://isatab.sf.net
5. Jones AR, Miller M, Aebersold R,... Sansone SA et al. (2007) The Functional Genomics Experiment model (FuGE). *Nature Biotechnology* Oct;25(10):1127-1133. http://fuge.sf.net