Standards and infrastructure for managing experimental metadata

BioInvestigation Index

(see also poster F4)

Susanna-Assunta Sansone, PhD

The European Bioinformatics Institute (EMBL-EBI)

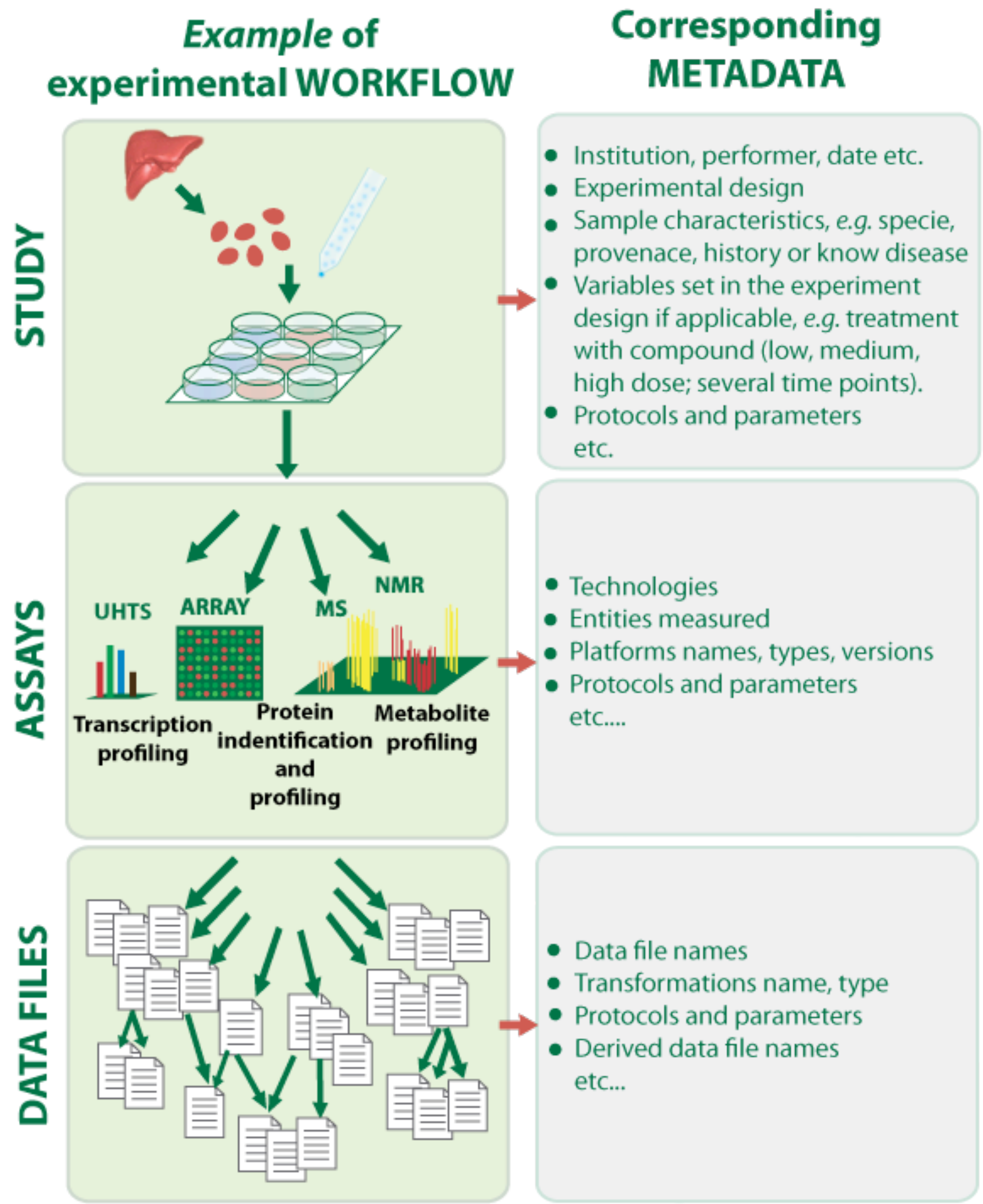
www.ebi.ac.uk/net-project

BioCurator Meeting – Berlin, April16-19, 2009

Outline

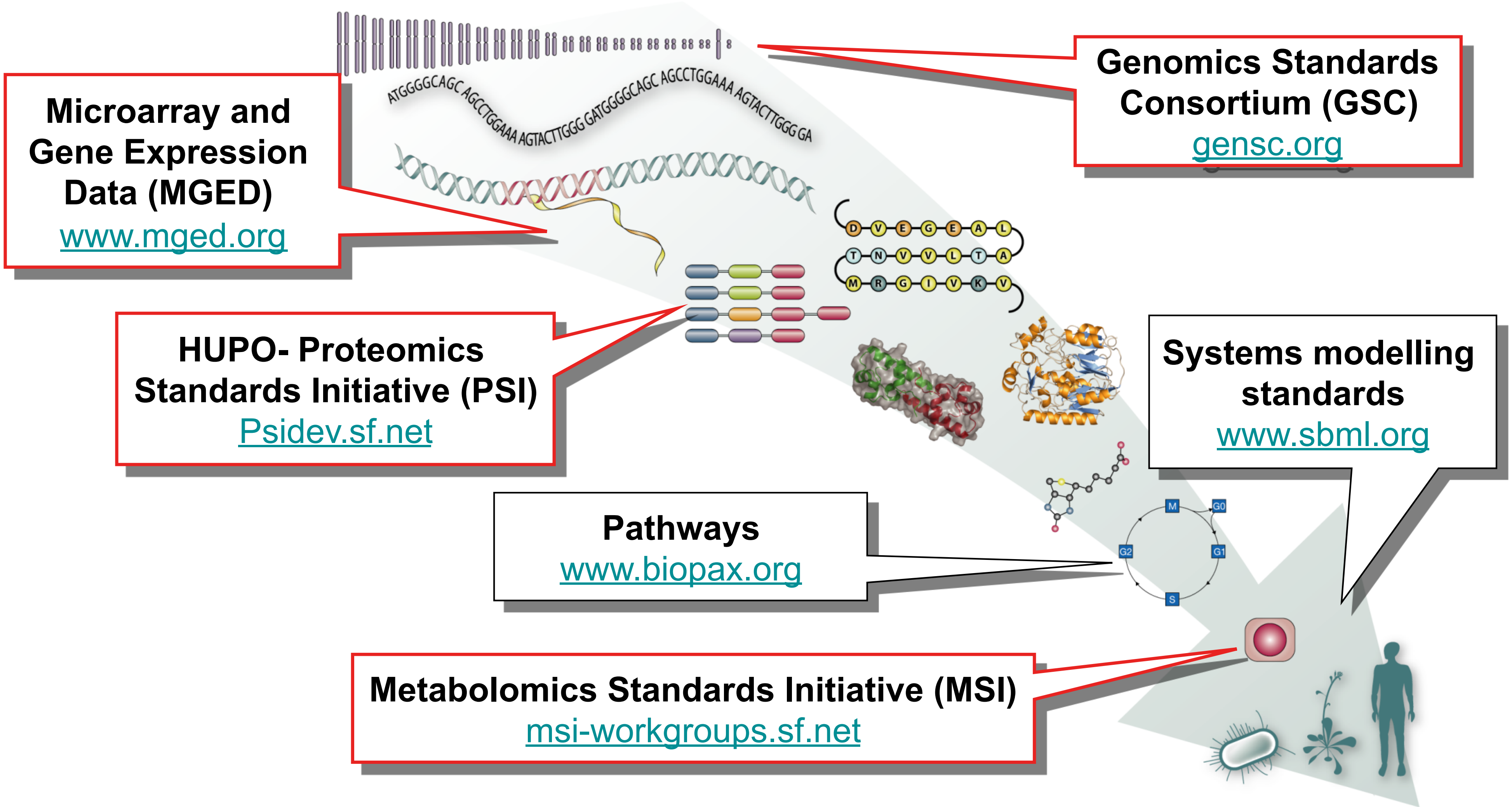
- Reporting standards
  - Hopes and hurdles
- Synergistic efforts
  - Overcome the fragmentation of standards
- Our standards-compliant implementation
  - Manage experimental metadata

Growing complexity of the experiments



Consistent reporting of the experimental metadata - <u>along with </u>the associated data- has a positive and long-lasting impact on the value of collective scientific outputs

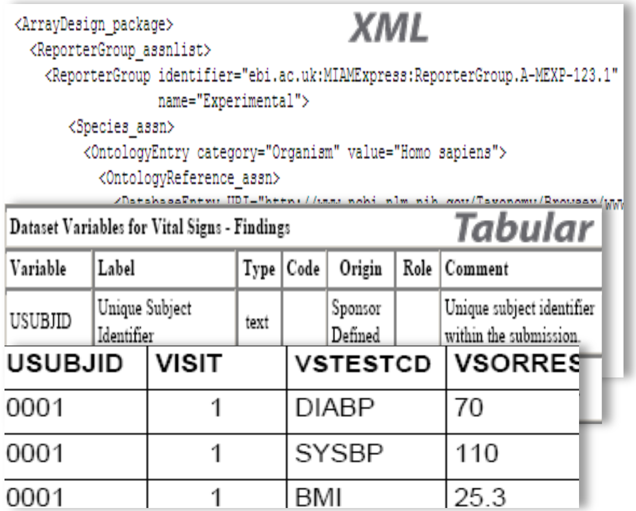
Grass root omics initiatives (*de facto* standards), e.g.:

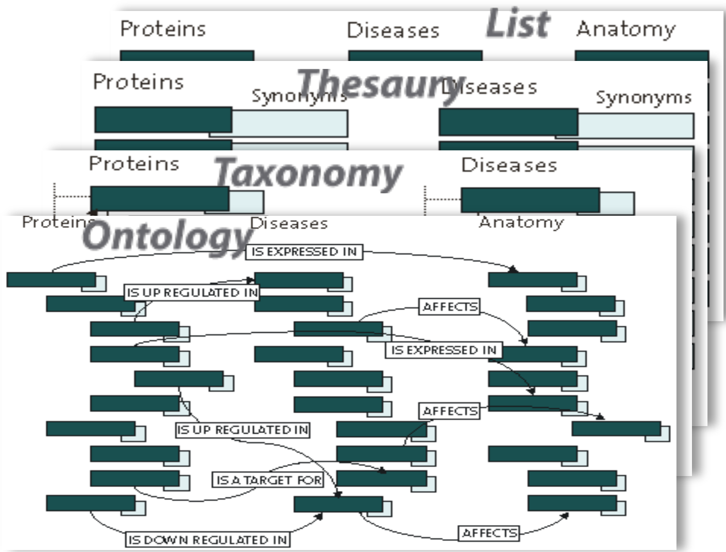


Some are <u>loosely</u> connected to regulatory/healthcare-driven initiatives and accredited Standards Developing Organizations (SDOs), *e.g.* CDISC, SEND, HL7, developing *de jure* standards.

Three types of reporting standards

Content
minimal information to be reported





Syntax
format for the communication

Semantics
terminology for the description

Fragmented standards, fragmented systems, *e.g.*:

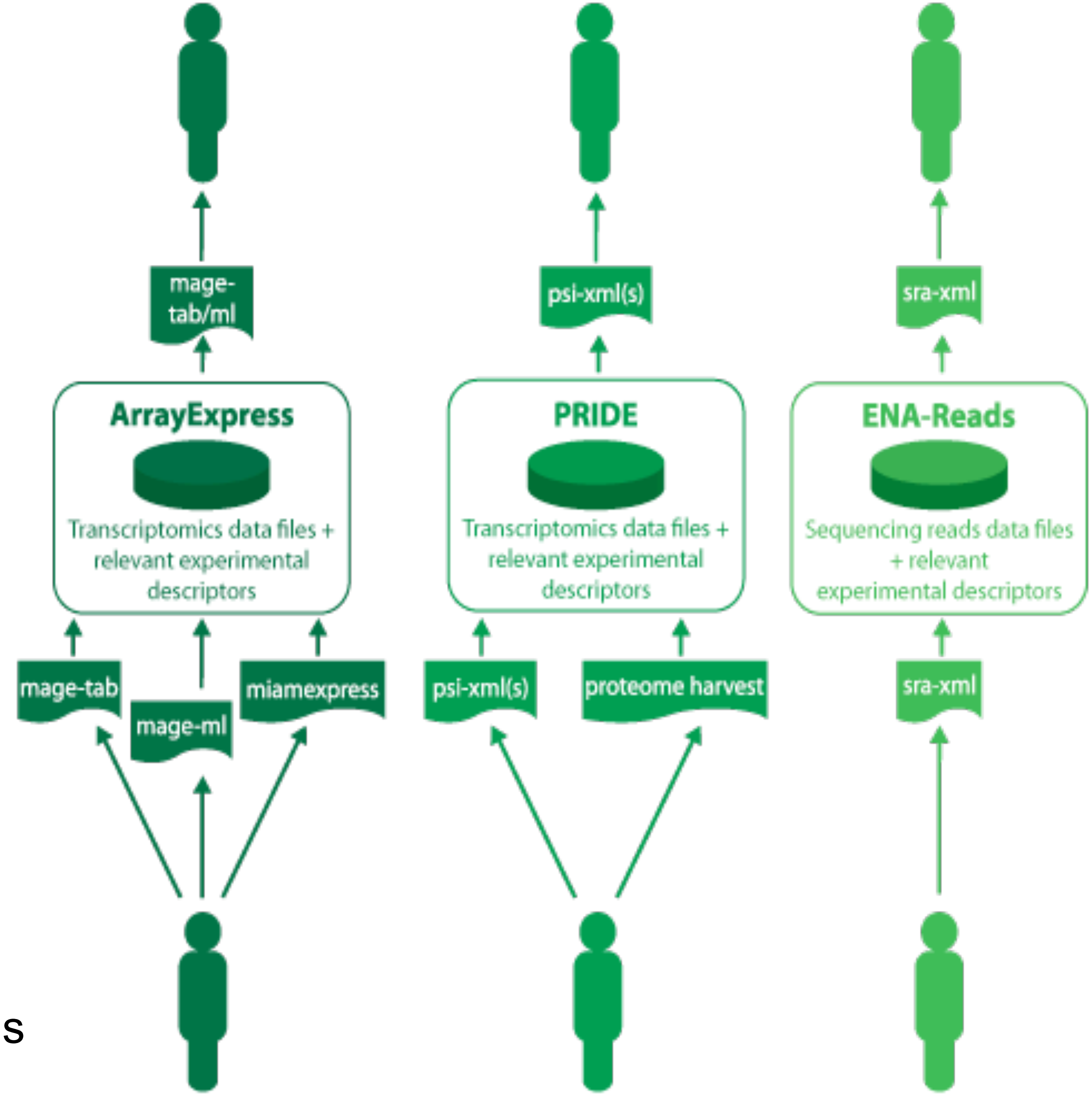
DIFFERENT
Access and exchange format

DIFFERENT
Core requirements captured

DIFFERENT
Terminologies

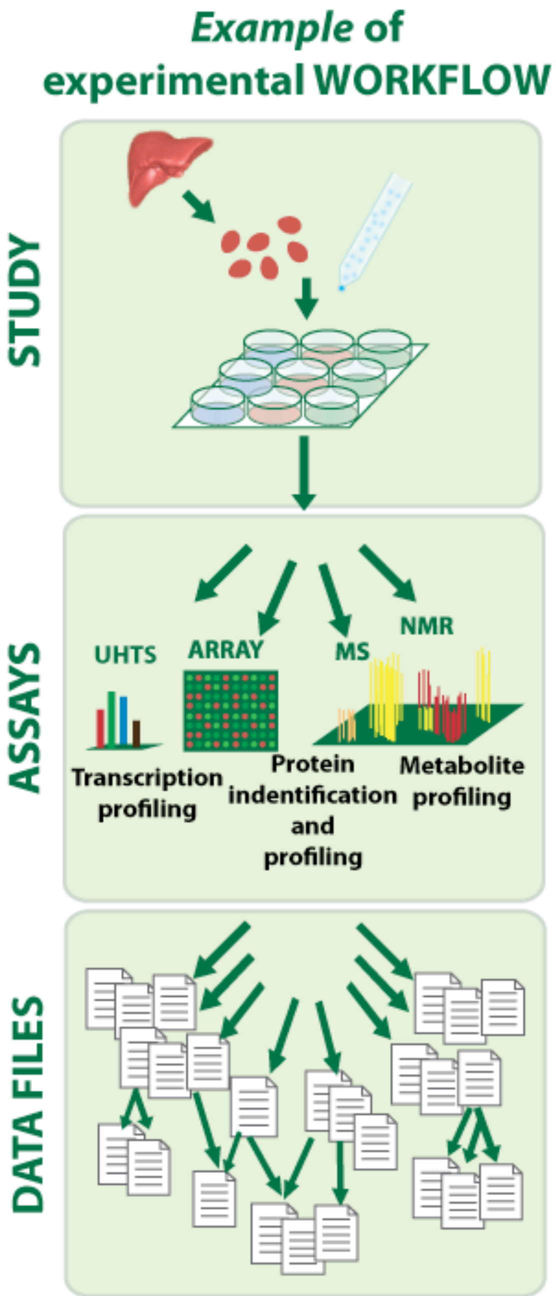
DIFFERENT
Deposition formats

DIFFERENT
Curation practices and tools



Three EBI omics systems

But....how do we manage complex experiments?



How do we **encourage** submissions of experimental metadata and data and **enable** consistent reporting and curation in the current scenario?

We need to address the fragmentation of standards

- Promote synergies among standards initiatives
  - ‘Limit’ the range and variability of formats, in particular
- Create interoperable reporting standards
  - Fit neatly into a jigsaw, resolving inconsistency and filling gaps
- Overcome several barriers
  - Technical, funds and (overall) sociological......

We need to address the fragmentation of standards

- Promote synergies among standards initiatives
  - 'Limit' the range and variability of formats, in particular
- Create interoperable reporting standards
  - Fit neatly into a jigsaw, resolving inconsistency and filling gaps
- Overcome several barriers
  - Technical, funds and (overall) sociological......
- Our* contribution to address these hurdles
  - ✓ Risen funds to hold workshops, supporting synergistic efforts
  - ✓ Initiated new synergistic efforts, where missing
  - ✓ Work with our data producers and collaborators to implement standards-compliant systems

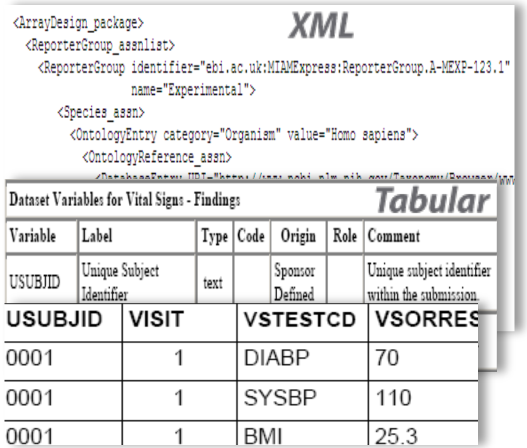
* *Sansone SA, Rocca-Serra P, Field D, Taylor C.*

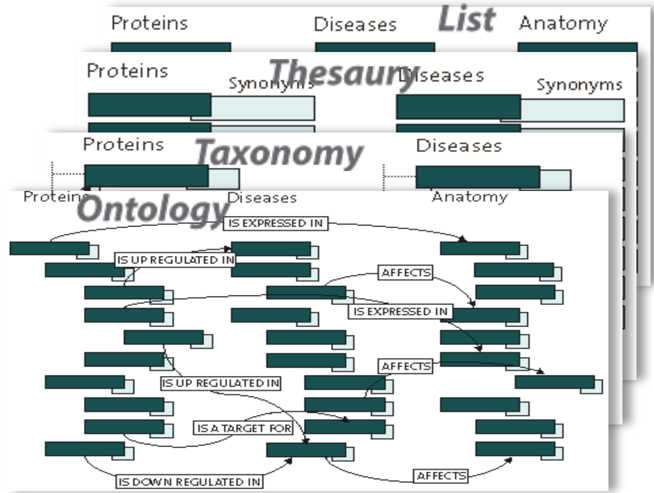
Synergistic efforts we contribute to

Scope

minimal information to be reported

MIBBI: http://mibbi.org





Syntax

format(s) for the communication

ISA-tab: http://isatab.sf.net

FuGE: http://fuge.sf.net

Semantics

terminology(s) for the description

OBO foundry: http://obofoundry.org

OBI: http://obi-ontology.org

Several stakeholders play pivotal role as enablers

EDITORIAL Volume 10, Number 10
October 2008

nature cell biology

Standardizing data

Biological research is benefiting from an explosion of data. There is an urgent need to invest in bioinformatic infrastructure and education to interpret this data and guarantee its archiving.

High-throughput research has helped fuel scientific progress at an unprecedented pace and left vast amounts of digital data in its wake. Even traditional hypothesis-driven research is now published at a rate that prohibits individuals from retaining the necessary overview. Bibliographic databases, such as PubMed, are key tools to navigate the information, but do not provide access to the primary data. The value of

How then can we ensure that researchers record the appropriate metadata to allow for accurate interpretation and repetition of data, and that the data are appropriately annotated before being deposited into databases? Several communities have developed data standards, but there has been little effort to coordinate these to avoid redundancy and incompatibility. Recently, the European Bioinformatics Institute (EBI) spearheaded the development of a shared platform for such standardization initiatives. The Minimum Information for Biological and Biomedical Investigations (MIBBI) project currently encompasses a collection of 22 minimum information guidelines on techniques such as microarrays, RNAi, quantitative PCR or FACS analysis. MIBBI aims to be a 'one-stop shop' for so-called checklist projects including MIAME (Minimum Information About a Microarray Experiment) and MIAPE (Minimum Information About a Proteomics Experiment). The 'Portal' section of MIBBI contains a growing number of links to the checklist projects, whereas the 'Foundry' invites input aimed at creating new, non-redundant checklists. Active community participation in the MIBBI Foundry will help ensure that the minimum information checklists remain relevant and thus, high-throughput data adequately annotated.

In addition to proper annotation, data must be described systematically in unambiguous language to make them machine-readable. To achieve this goal, communities must agree on ontologies (formal vocabularies for data and concepts). Ontologies allow semantic interoperability between various bioinformatics platforms and ensure that multiple repositories are compatible with each other. Although this remains a critical roadblock, the creation of the Open Biomedical Ontologies (OBO) Foundry, which acts as an umbrella organization for ontology projects, represents

BioMed Central's journals - with clinical content and BMC Bioinformatics - now include a link to the MIBBI in the instructions for authors and encourage data deposition

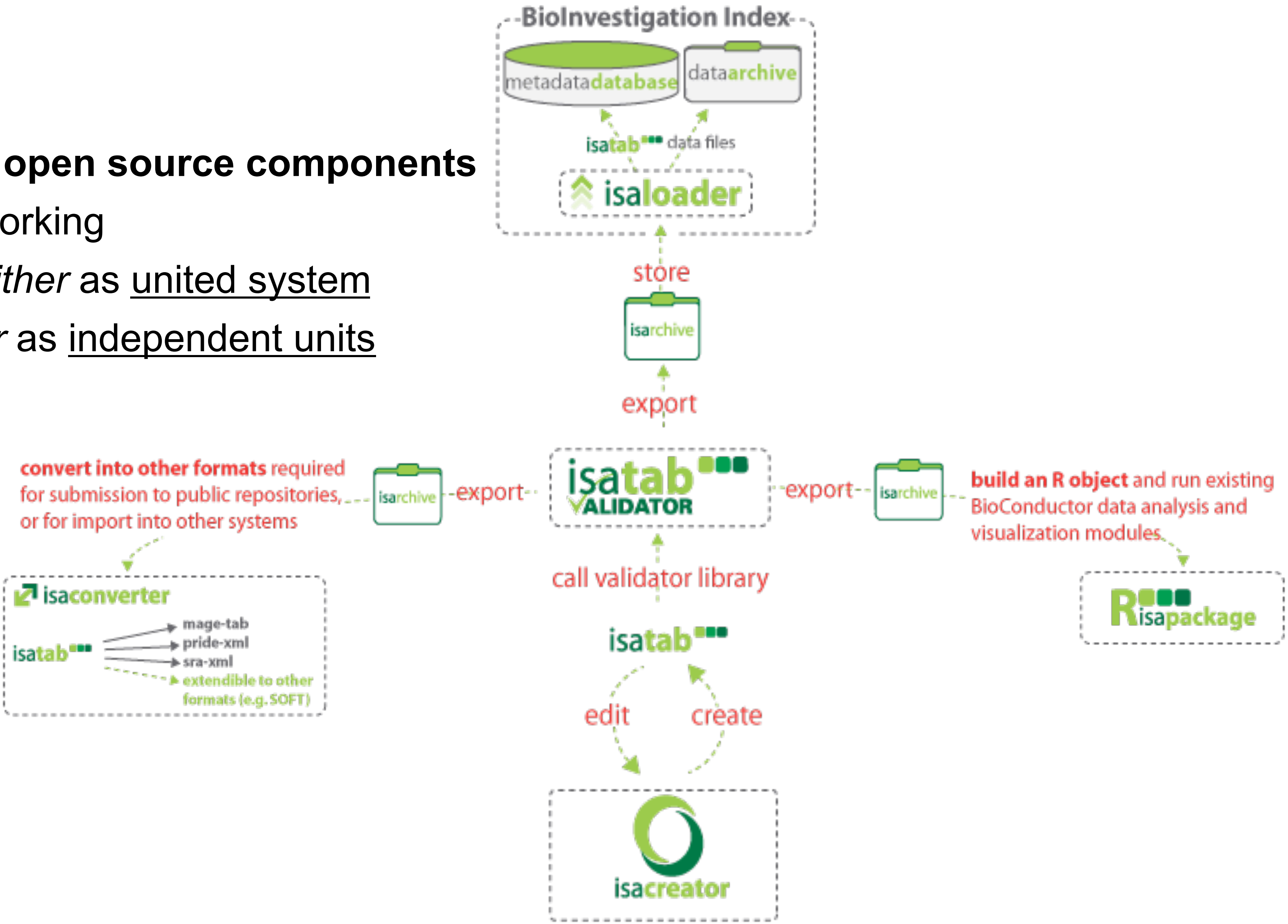
Our infrastructure based on synergistic standards

5 open source components

working

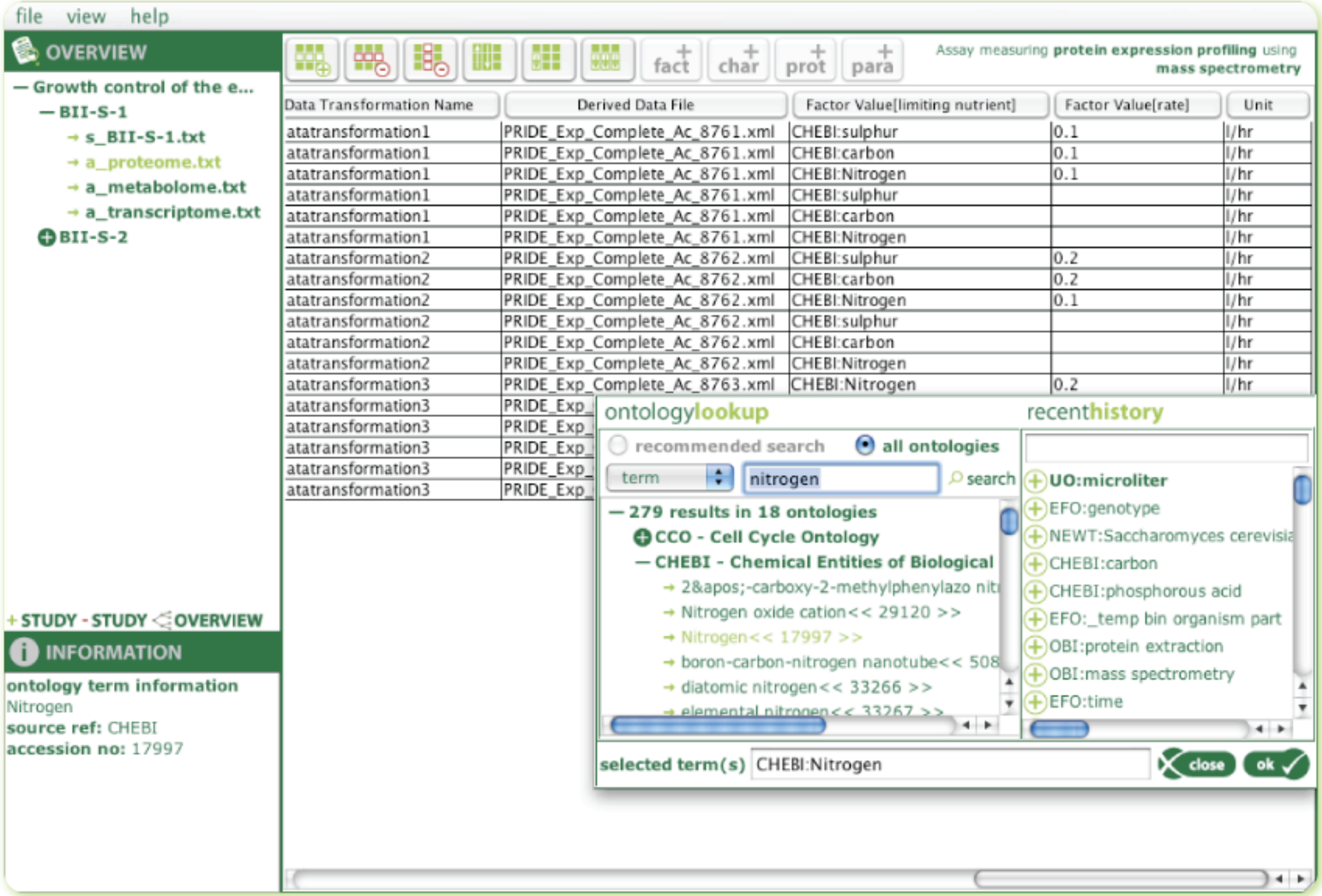
either as united system

or as independent units



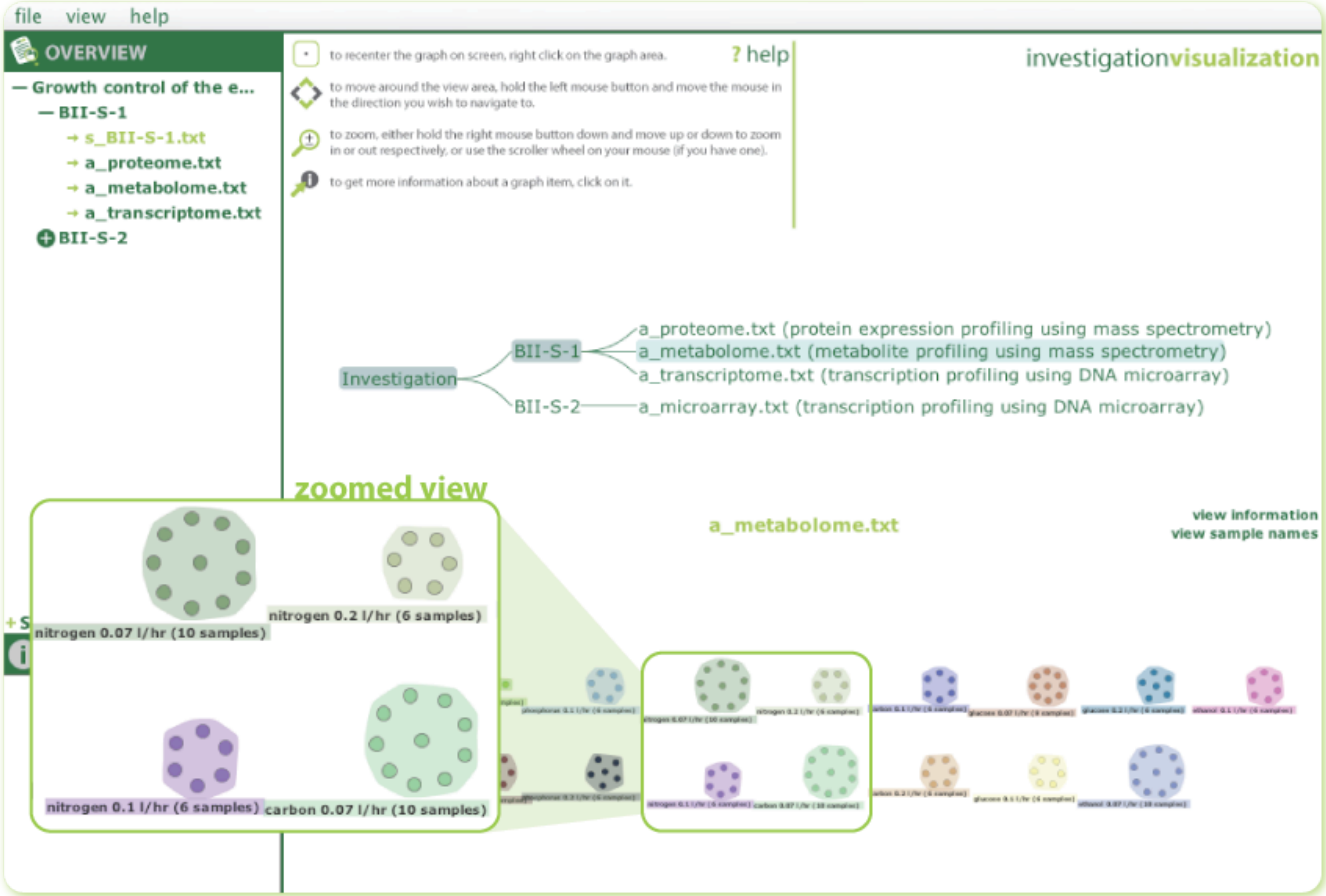
Component 1: ISAcreator, standalone editor tool

Component 1: ISAcreator, standalone editor tool

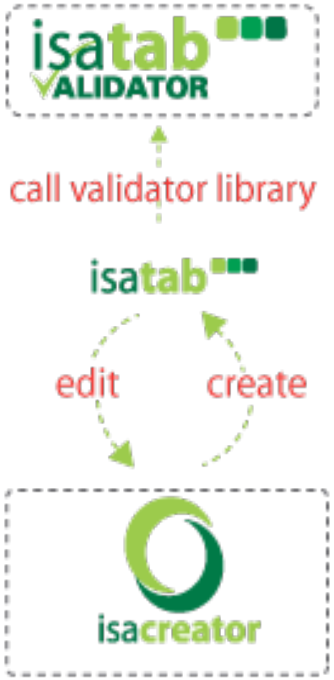


isacreator

Component 1: ISAcreator, standalone editor tool

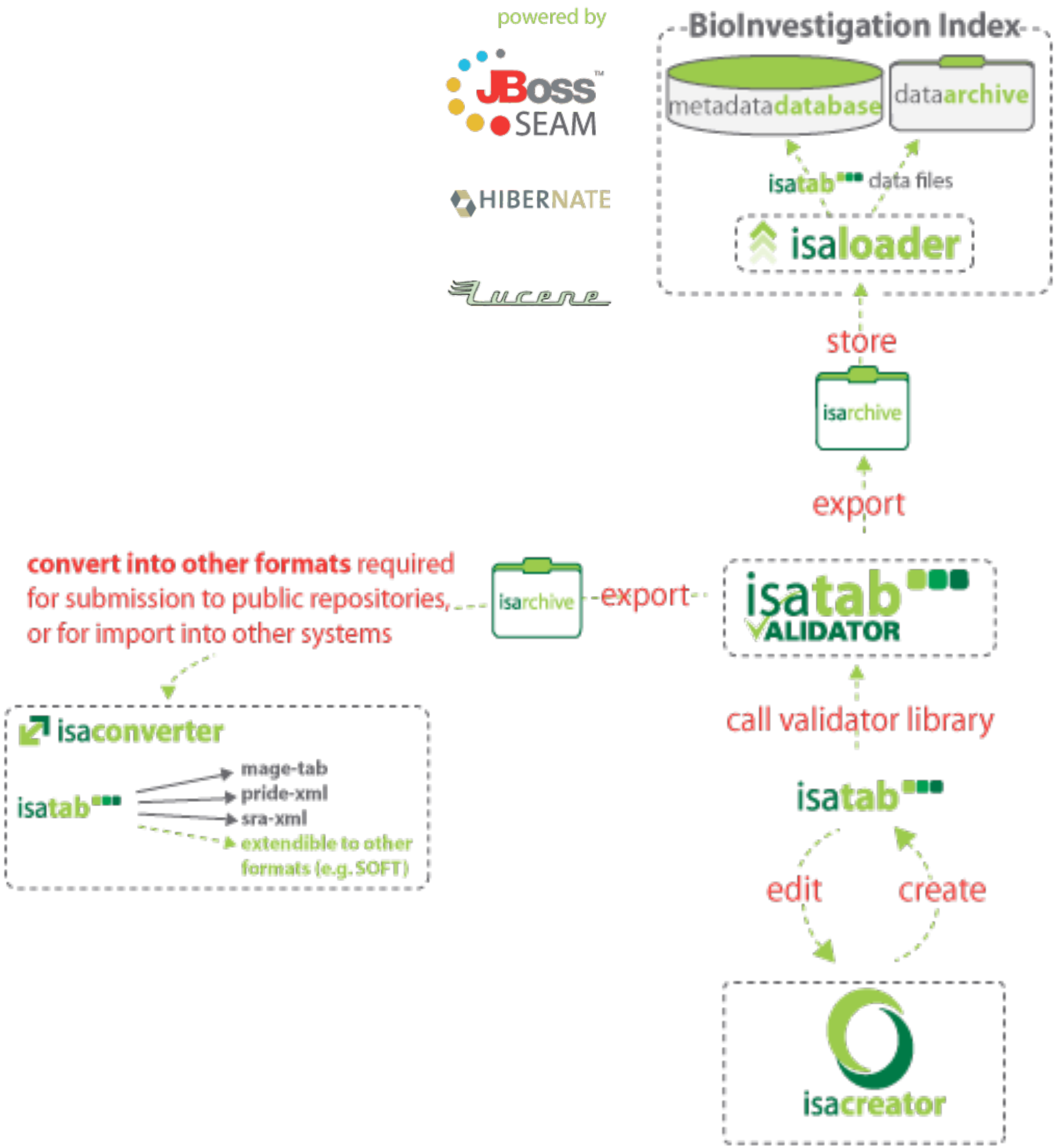


Component 2: ISAvalidator

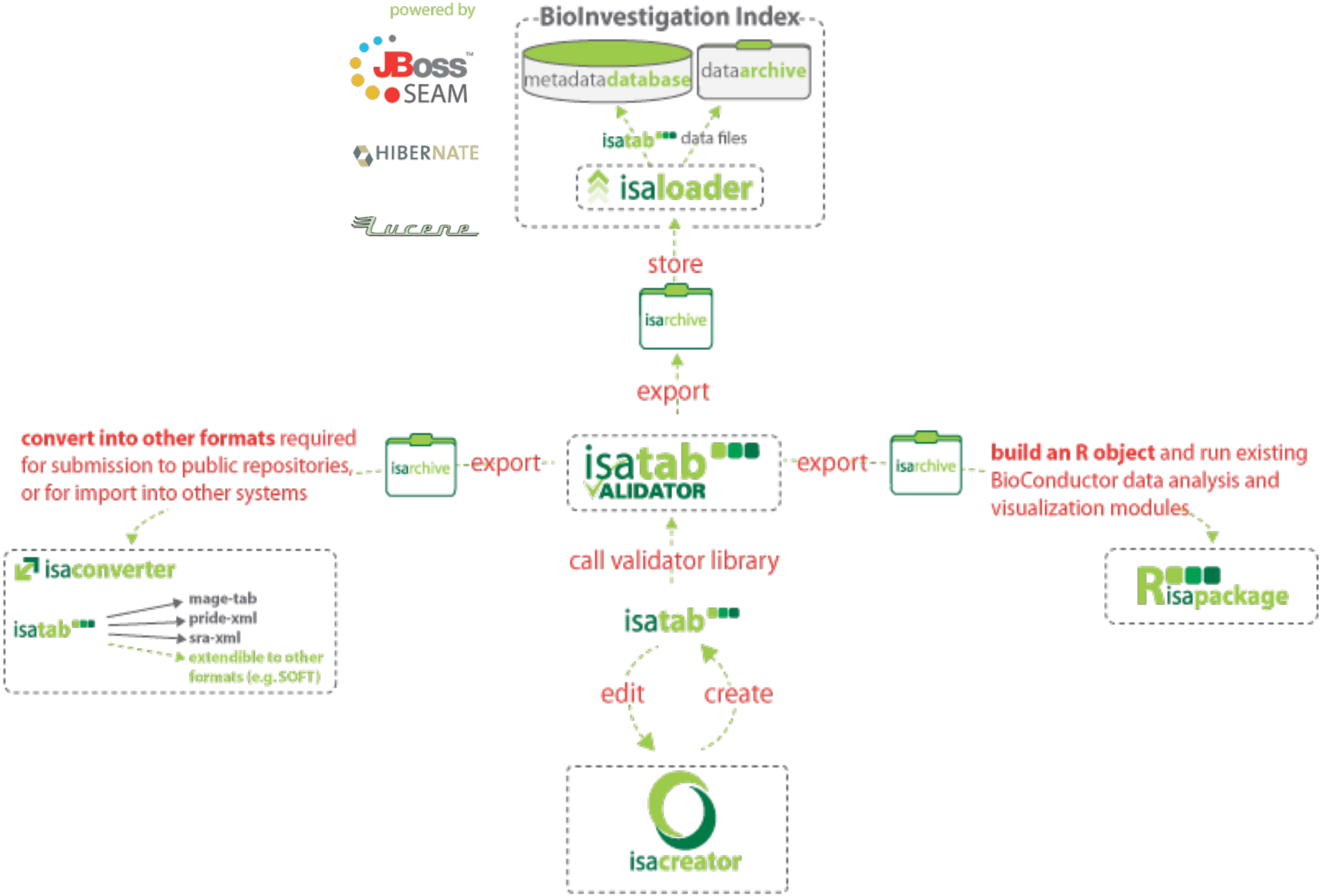


Component 3: BioInvestigation Index database

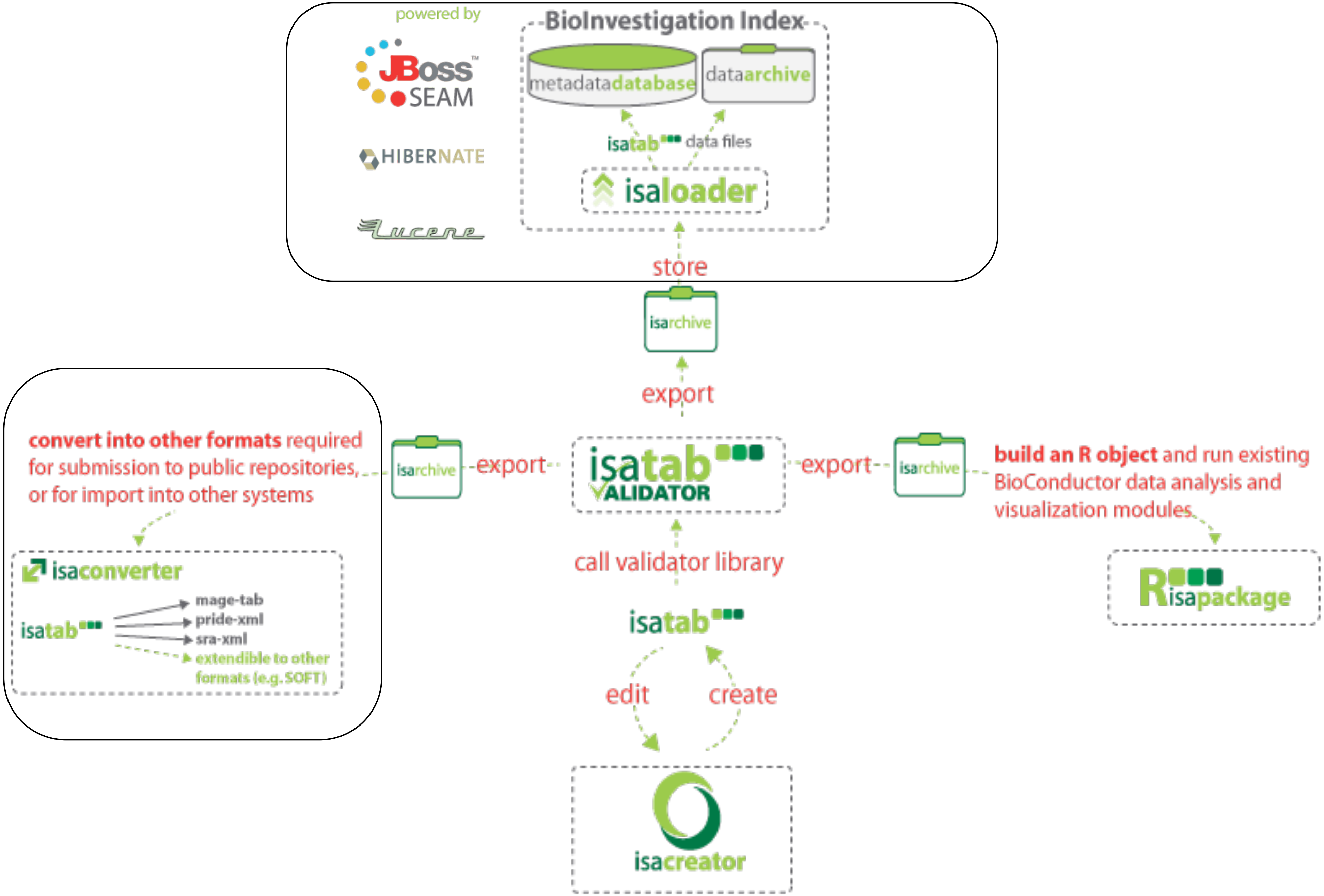
Component 4: ISAconverter



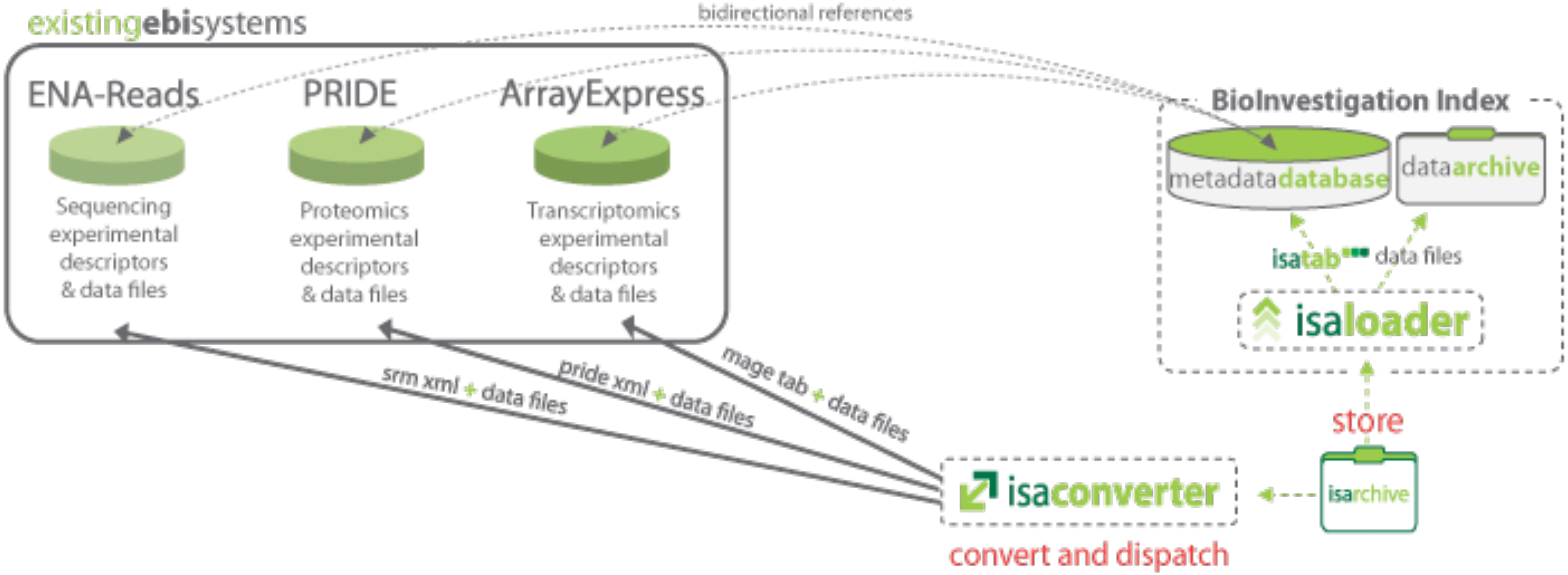
Component 5: R package for ISA-TAB (ongoing)



Instance deployed at EBI, as prototype



Instance deployed at EBI, as prototype



http//www.ebi.ac.uk/bioinvindex

Public studies are visible and searchable

bii bio investigation index

browse
submit
contact
credits

search

freetext
organism: Filter on organisms
measurement: Filter on measurement
technology: Filter on technology
platform: Filter on Platform

clearfields searchindex

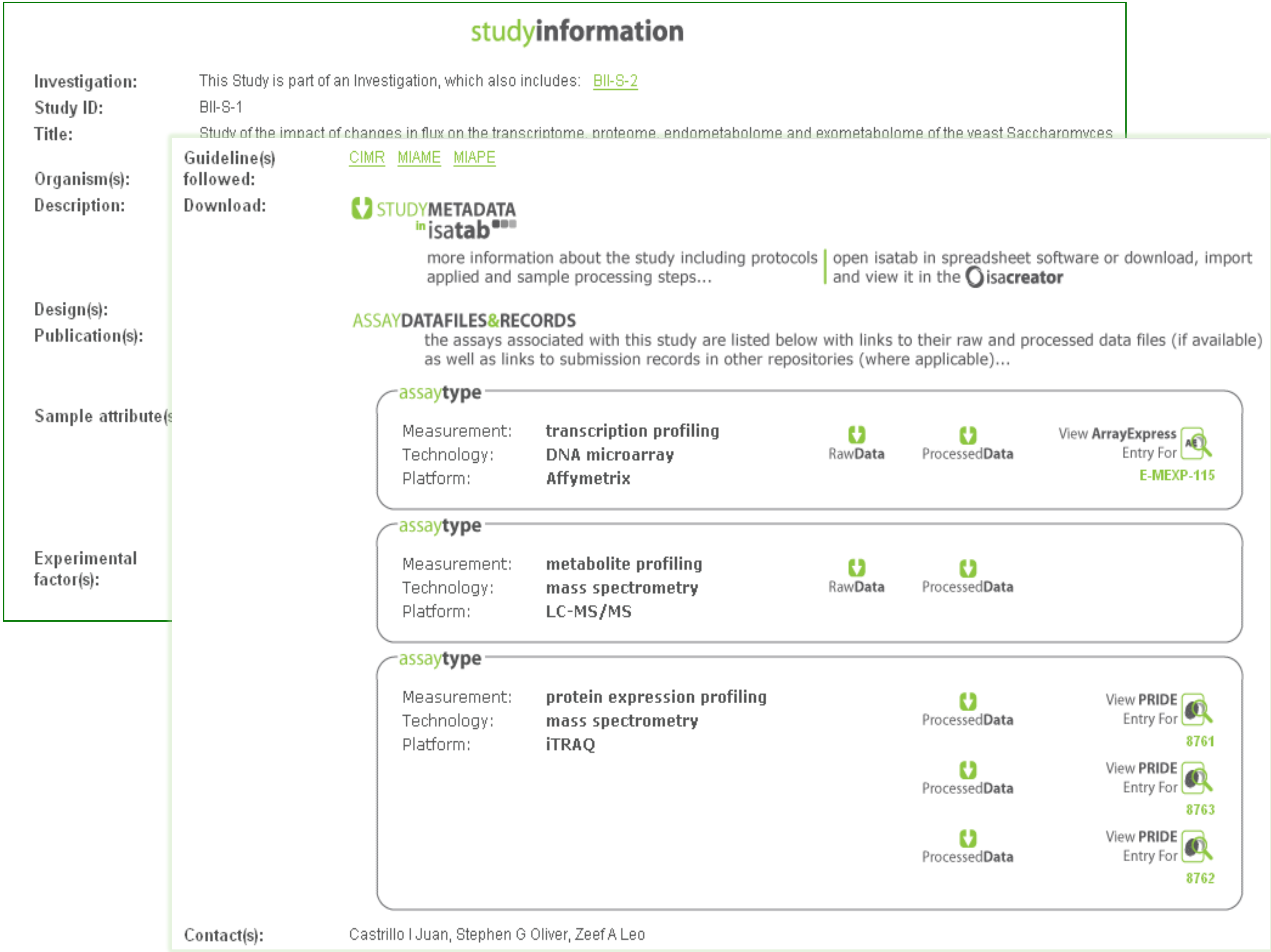
views

COMPOUND-CENTRIC VIEW

browsestudies

6 public studies containing 282 assays

Investigation	Study				Assay		
Acc	Acc	Title	Organism	Factor	Measurement	Technology	#
	BII-S-5	Determination of the complete genome sequence of Salmonella paratyphi A str. AKU_12601	Salmonella enterica subsp. enterica serovar Paratyphi A str. AKU_12601		genome sequencing	nucleotide sequencing	1
	BII-S-6	The Influence of Pharmacogenetics on Fatty Liver Disease in the Wistar and Kyoto Rats: A Combined Transcriptomic and Metabonomic	Rattus norvegicus (Rat)	time, strain, compound	transcription profiling metabolite profiling	DNA microarray NMR spectroscopy	17 79
	BII-S-3	Metagenomes and Metatranscriptomes of phytoplankton blooms from an ocean acidification mesocosm experiment	marine metagenome	dose, collection time, compound	metagenome sequencing transcription profiling	nucleotide sequencing nucleotide sequencing	4 4
	BII-S-4	An initial characterisation of the Fasciola hepatica transcriptome using 454-FLX sequencing	Fasciola hepatica (Liver fluke)		transcription profiling	nucleotide sequencing	1
BII-I-1	BII-S-2	A time course analysis of transcription response in yeast treated with rapamycin, a specific inhibitor of the TORC1 complex: impact on yeast growth	Saccharomyces cerevisiae (Baker's yeast)	exposure time, dose, compound	transcription profiling	DNA microarray	14
BII-I-1	BII-S-1	Study of the impact of changes in flux on the transcriptome, proteome, endometabolome and exometabolome of the yeast Saccharomyces cerevisiae under different nutrient limitations	Saccharomyces cerevisiae (Baker's yeast)	limiting nutrient, rate	protein expression profiling transcription profiling metabolite profiling	mass spectrometry DNA microarray mass spectrometry	3 48 111


studyinformation
Investigation: This Study is part of an Investigation, which also includes: BII-S-2
Study ID: BII-S-1
Title: Study of the impact of changes in flux on the transcriptome, proteome, endometabolome and exometabolome of the yeast Saccharomyces
Organism(s):
Description:
Design(s):
Publication(s):
Sample attribute(s
Experimental factor(s):
Guideline(s) followed: CIMR MIAME MIAPE
Download: STUDYMETADATA in isatab
more information about the study including protocols applied and sample processing steps...
open isatab in spreadsheet software or download, import and view it in the isacreator
ASSAYDATAFILES&RECORDS
the assays associated with this study are listed below with links to their raw and processed data files (if available) as well as links to submission records in other repositories (where applicable)...
assaytype
Measurement: transcription profiling
Technology: DNA microarray
Platform: Affymetrix
RawData
ProcessedData
View ArrayExpress Entry For
E-MEXP-115
assaytype
Measurement: metabolite profiling
Technology: mass spectrometry
Platform: LC-MS/MS
RawData
ProcessedData
assaytype
Measurement: protein expression profiling
Technology: mass spectrometry
Platform: iTRAQ
ProcessedData
View PRIDE Entry For
8761
ProcessedData
View PRIDE Entry For
8763
ProcessedData
View PRIDE Entry For
8762
Contact(s): Castrillo I Juan, Stephen G Oliver, Zeef A Leo

Acknowledgements and references

Marco Brandizi (*Software Engineer*)
Eamonn Maguire (*Software Engineer*)
Nataliya Sklyar (*Software Engineer*)
Chris Taylor (*Bioinformatician*)

Manon Delahaye (*Trainees -Software Engineer*)
Richard Evans (*Trainees -Software Engineer*)

Technical Coordinator: Philippe Rocca-Serra
Coordinator: Susanna-Assunta Sansone

Open source codes,
soon: http://isatab.sf.net

Posters: F4, E27, E3

carcinoGENOMICS a Project of the European Union

NEBC

NERC Environmental Bioinformatics Centre

NATURAL ENVIRONMENT RESEARCH COUNCIL

NuGO

European Nutrigenomics Organisation

bbsrc biotechnology and biological sciences research council

EMBL-EBI

FDA FOOD AND DRUG ADMINISTRATION

The National Center for Toxicological Research (NCTR) Center for Toxicoinformatics