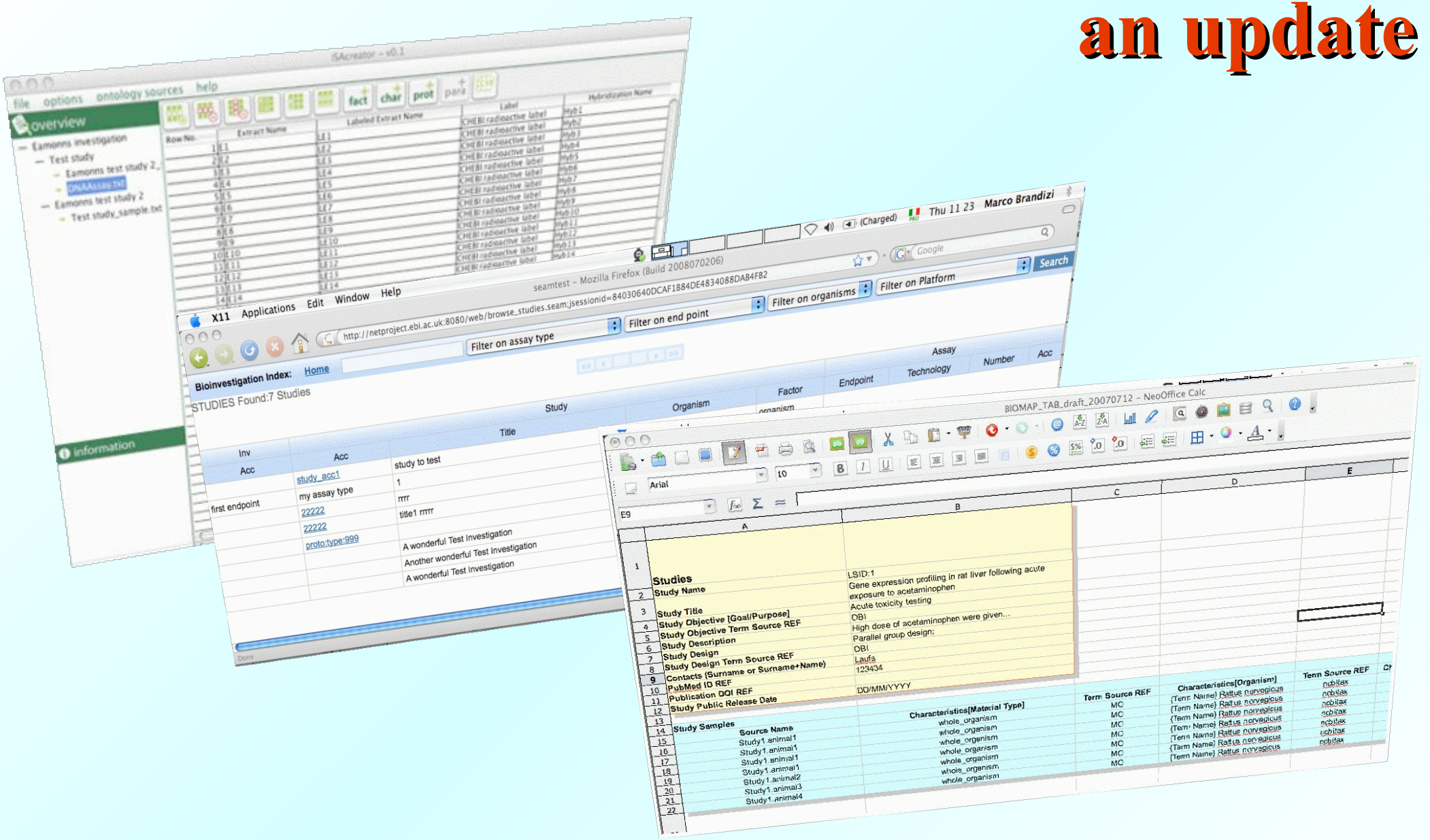
The BioInvestigation Index Project: an update



Marco Brandizi, 24 Jul 2008

EMBL-EBI European Bioinformatics

Microarray

Outline

The BII Project and its components

The Submission Tool

The Submission Tool Internals

Thoughts, Future

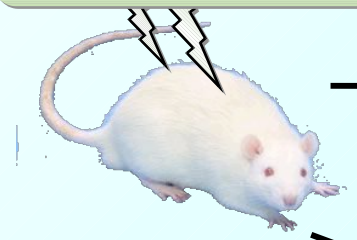
BII: Rationale

OMICS-BASED INVESTIGATION

(hypothesis, design, contact persons etc)

STUDY(1)

Compound X treatment



Euthanasia

Target organ, tissue

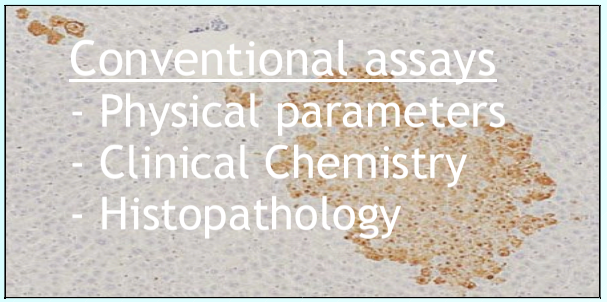
DESIGN

Number of animals, controls, doses, length etc

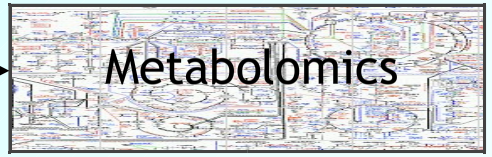
STUDY(2)

Compound X treatment

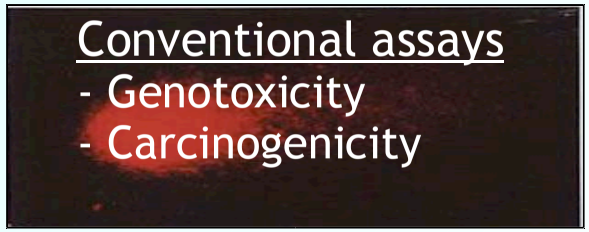
ASSAY(s) and DATA



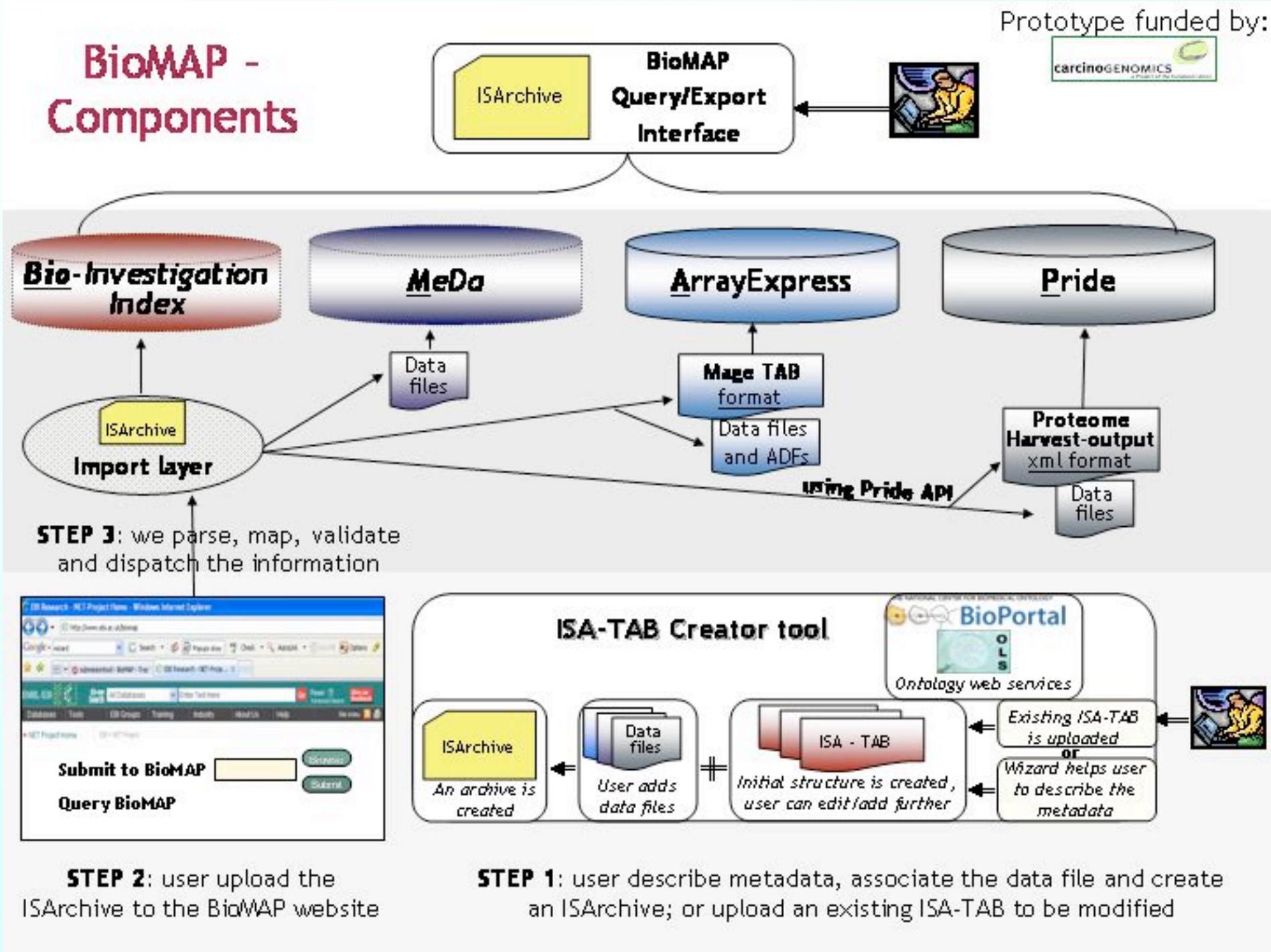
RNA →

Fluids, Tissues, Cells → 

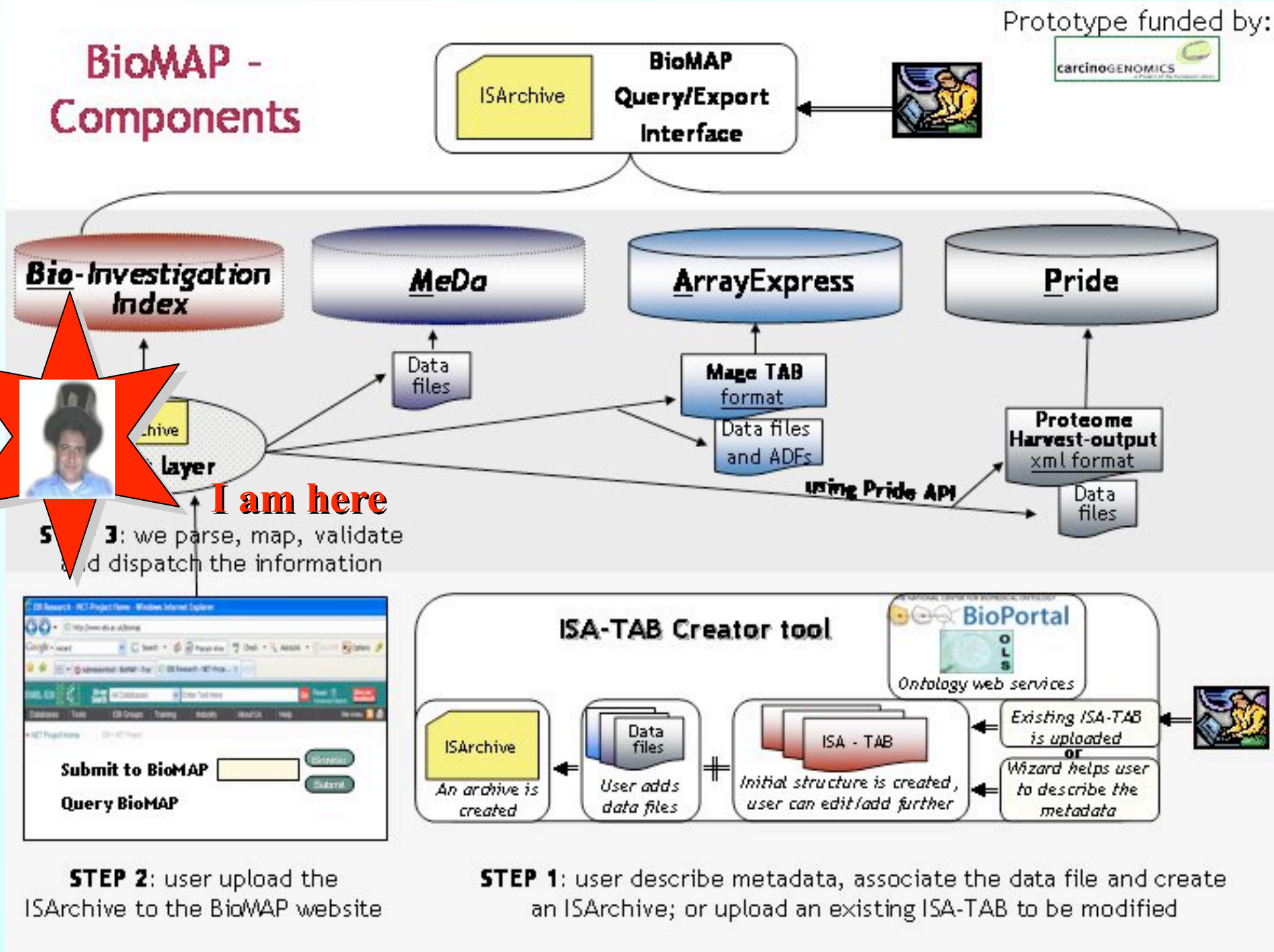
Proteins →



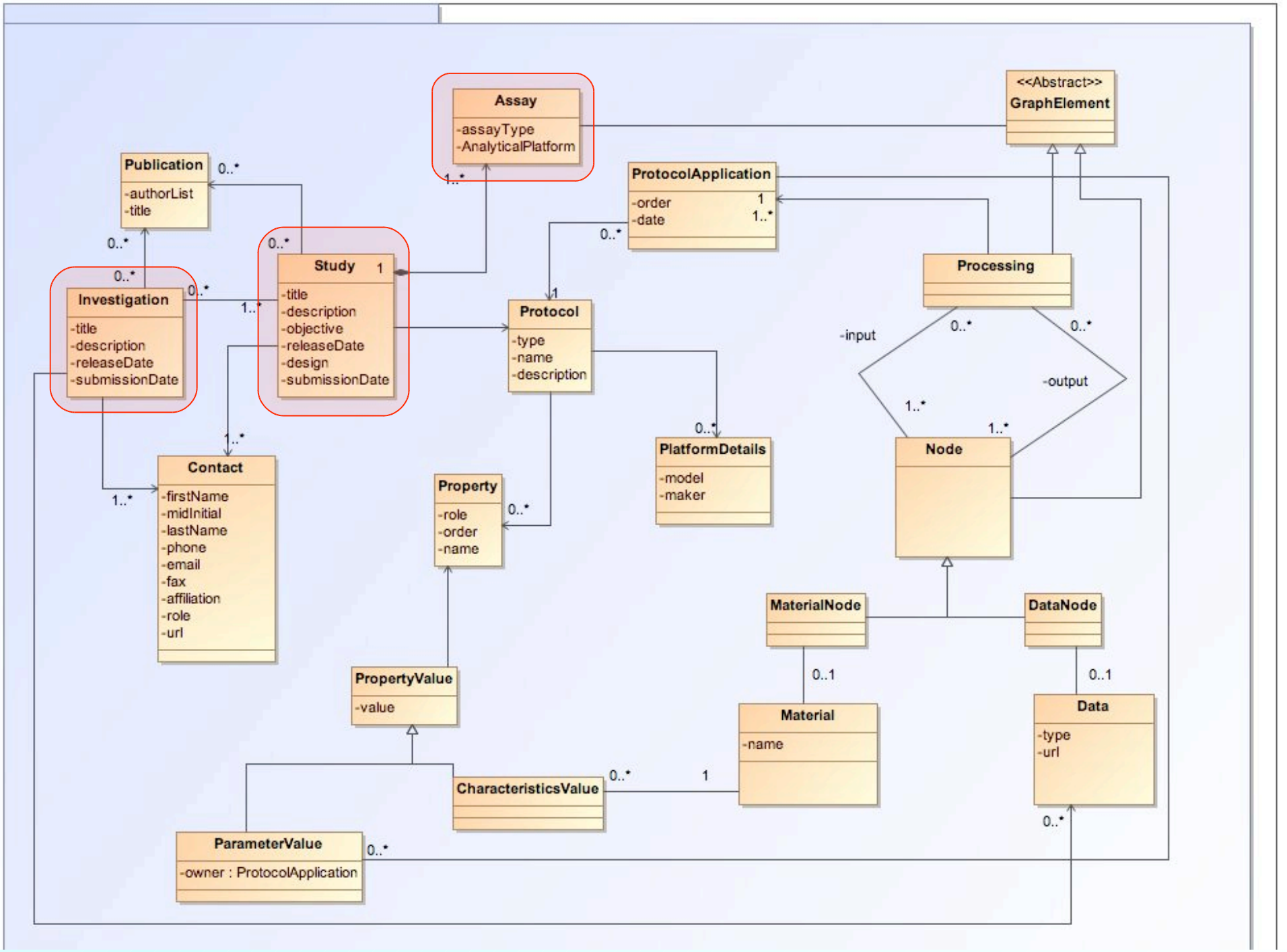
BII: General Architecture



BII: General Architecture



The BioInvestigationIndex Model



Outline

The BII Project and its components

The Submission Tool

The Submission Tool Internals

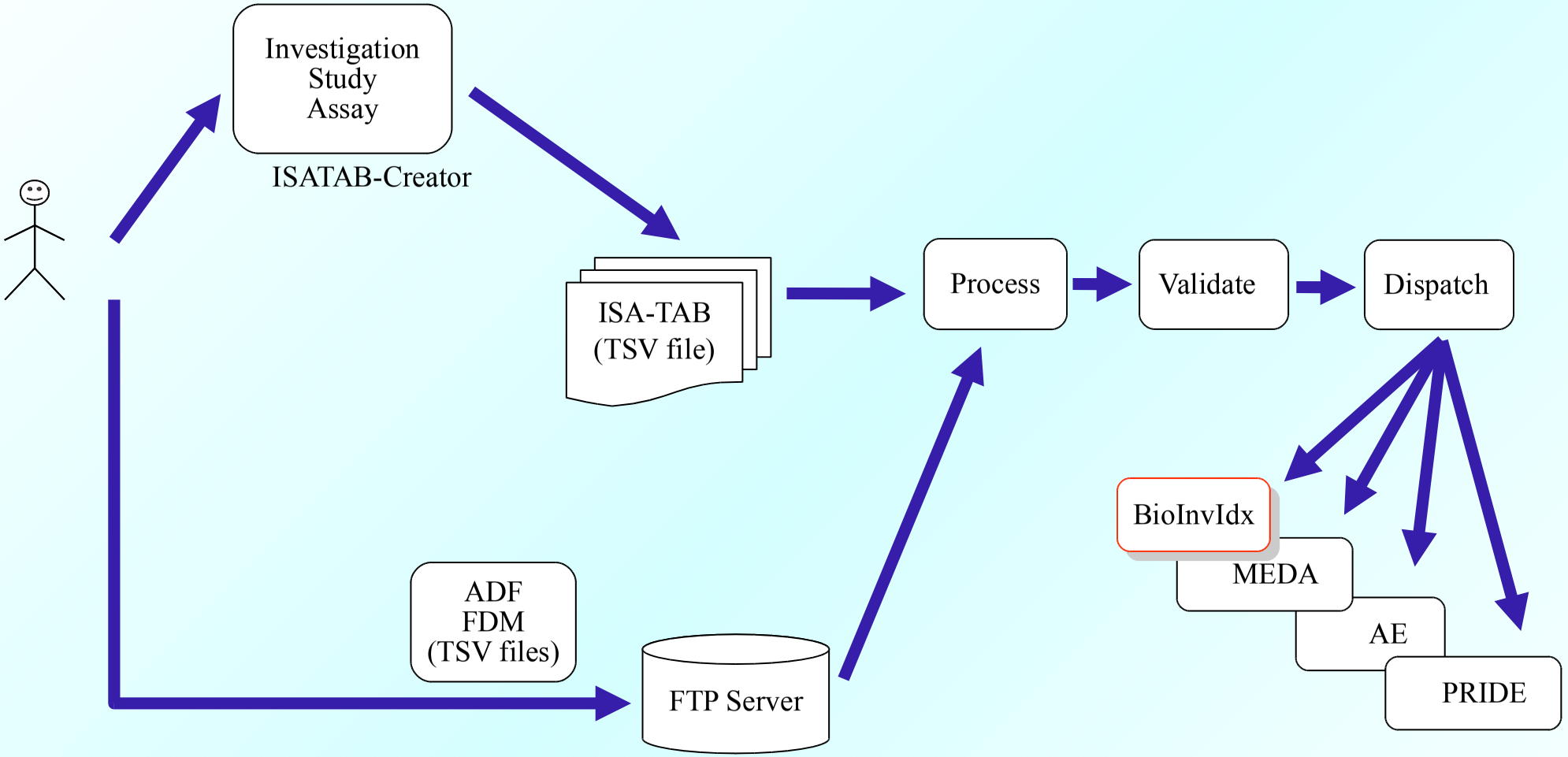
Thoughts, Future

The ISA-TAB format

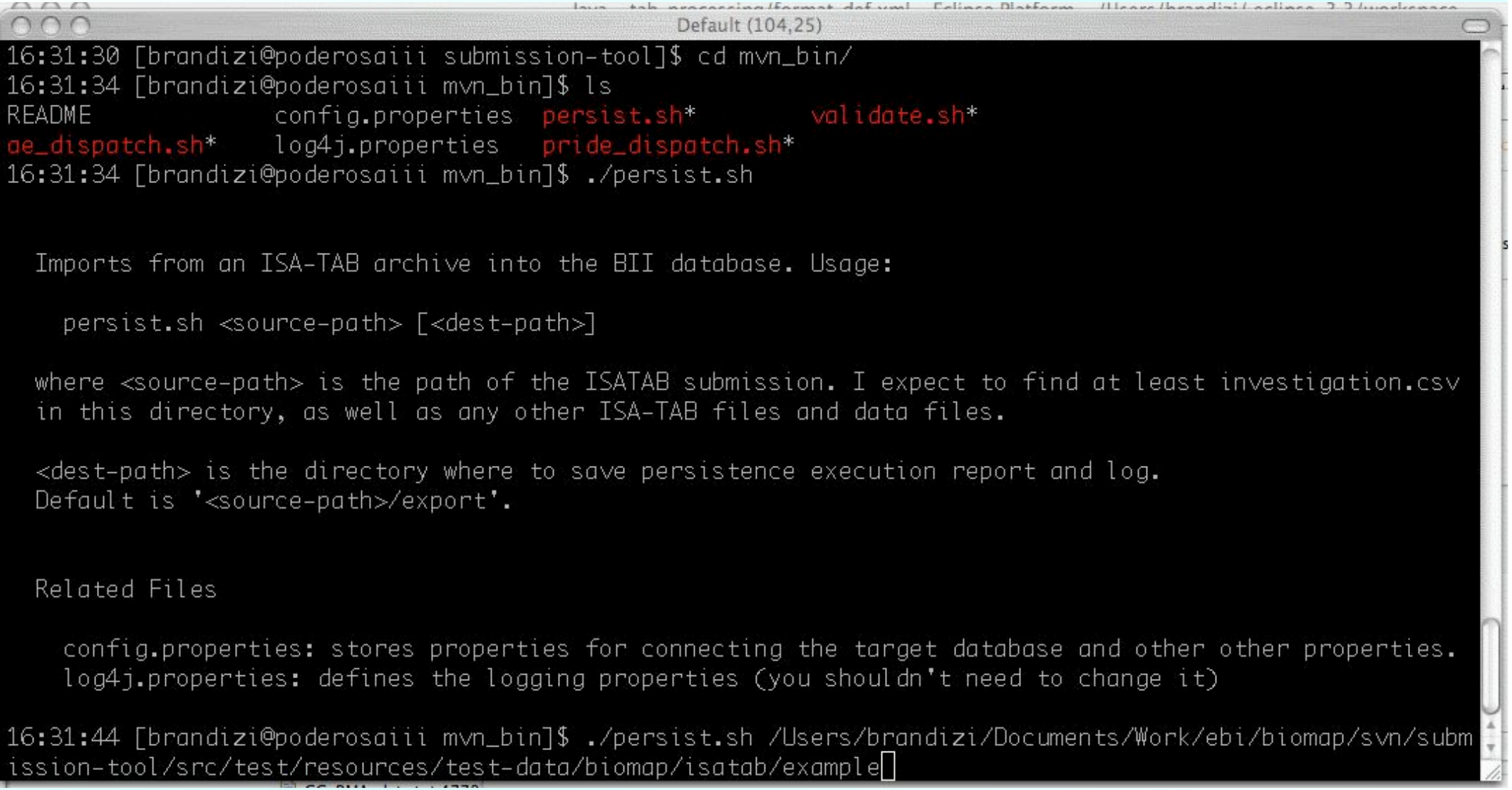
- A tabular format inspired to similar cases (MAGE-TAB, PRIDE-Harvest)
- Has a structure similar to the one of the Model
  - an IDF (Investigation Description Format)
    - General Info + Investigation + Studies
  - SDRF-like sections, with source/sample processing separated from the technology-specific processing

An example follows
(see http://isatab.sourceforge.net)

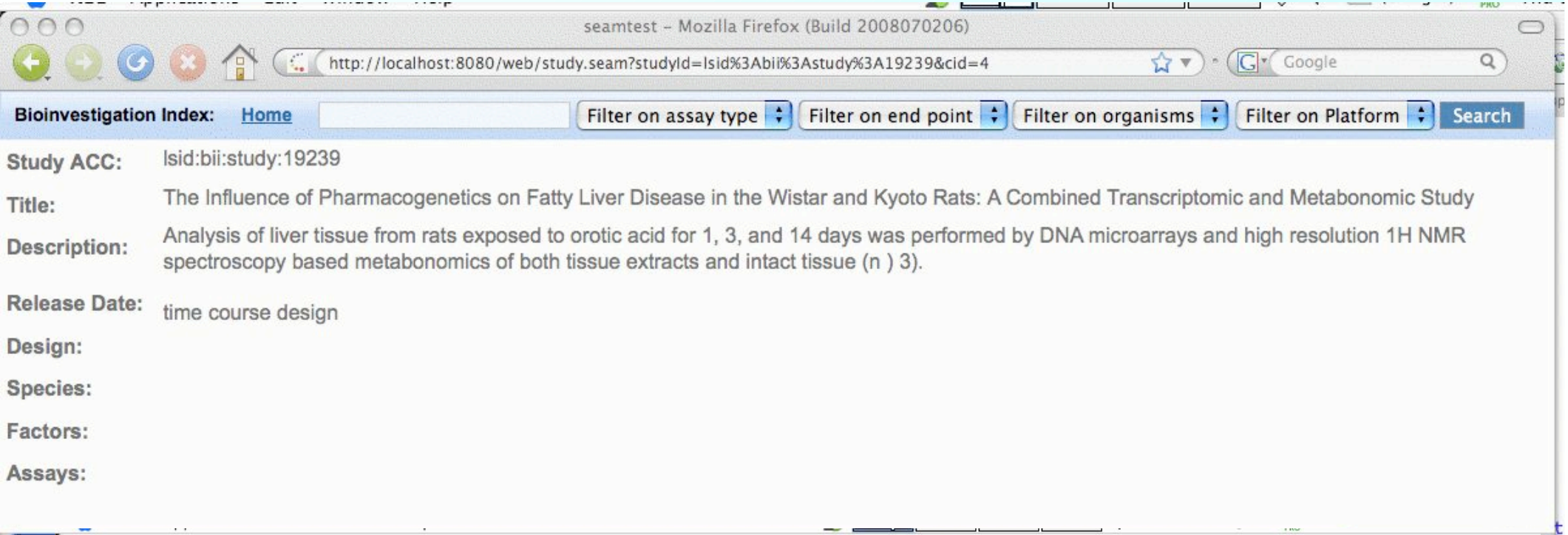
Submission Tool: General Workflow

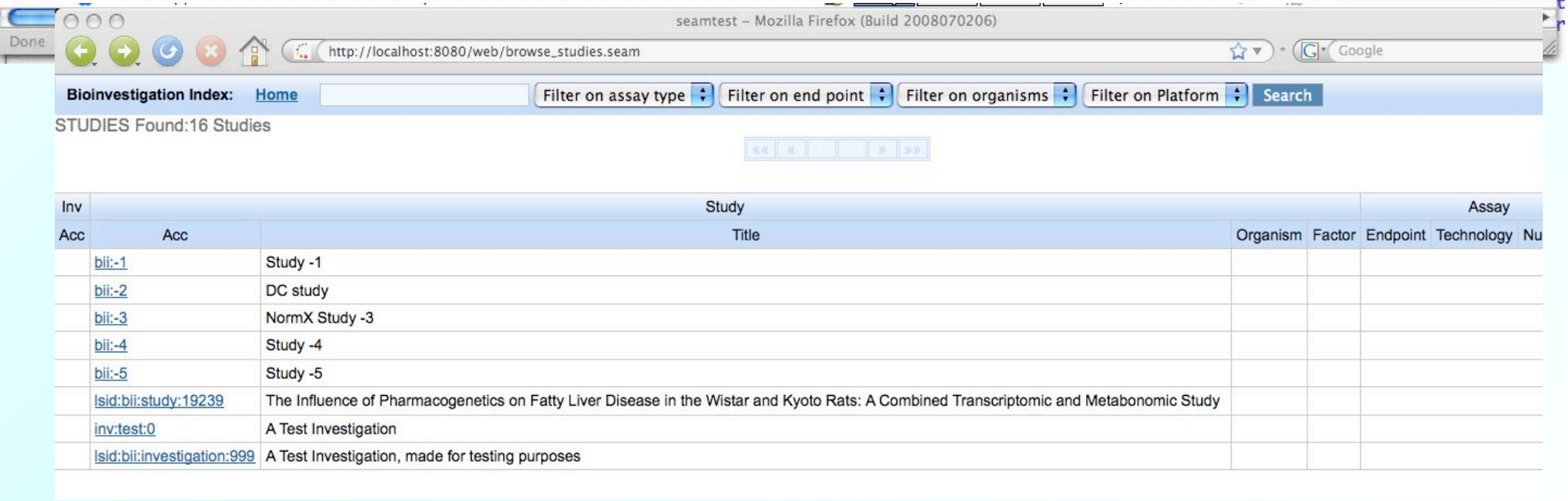


Submission Tool: Example



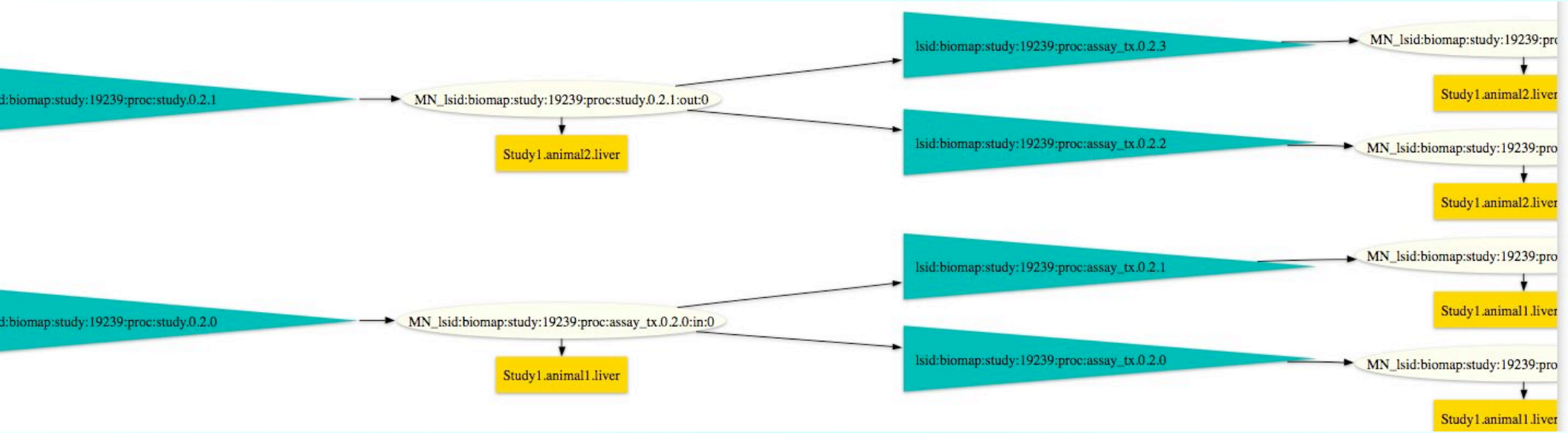
Submission Tool: Example

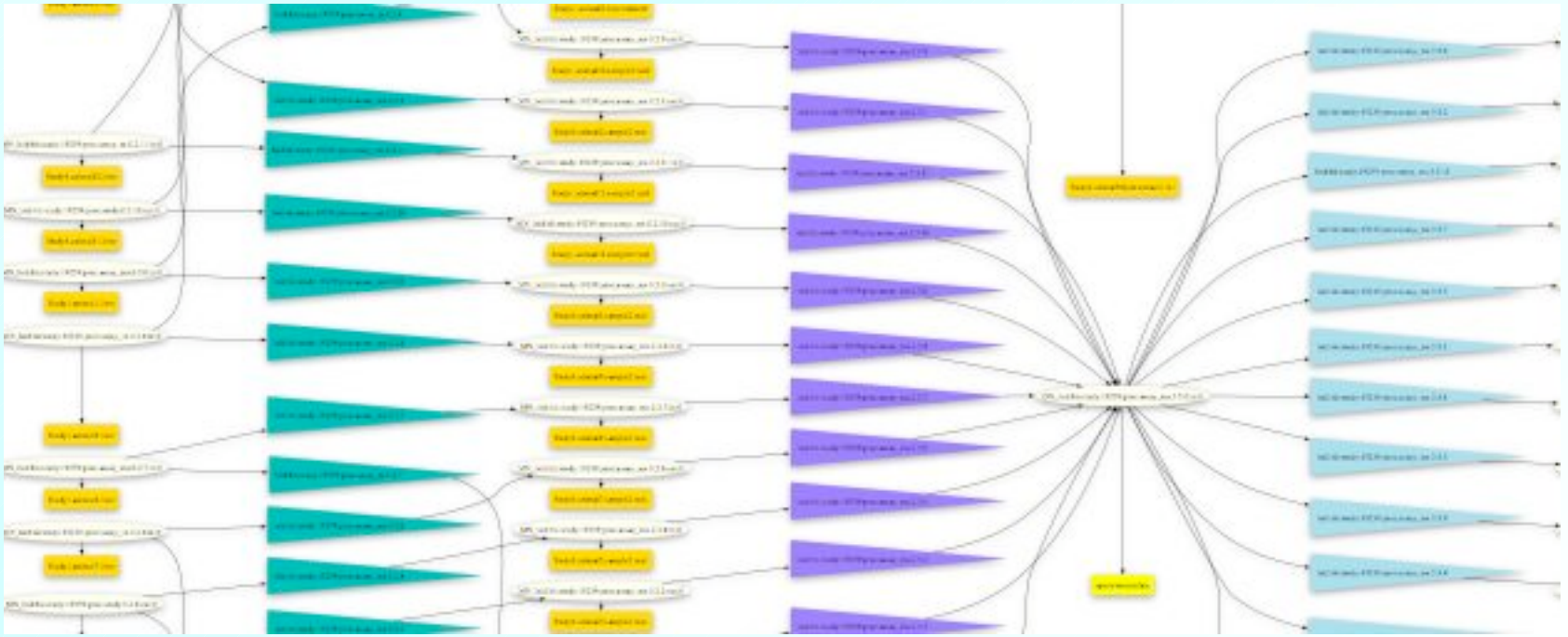




Inv	Study				Assay		
Acc	Acc	Title	Organism	Factor	Endpoint	Technology	Nu
	bii:-1	Study -1					
	bii:-2	DC study					
	bii:-3	NormX Study -3					
	bii:-4	Study -4					
	bii:-5	Study -5					
	lsid:bii:study:19239	The Influence of Pharmacogenetics on Fatty Liver Disease in the Wistar and Kyoto Rats: A Combined Transcriptomic and Metabonomic Study					
	inv:test:0	A Test Investigation					
	lsid:bii:investigation:999	A Test Investigation, made for testing purposes					

Submission Tool: Example





Outline

The BII Project and its components

The Submission Tool

The Submission Tool Internals

Thoughts, Future

Submission Tool Impl.: General Approach

ISA-TAB (TSV file)

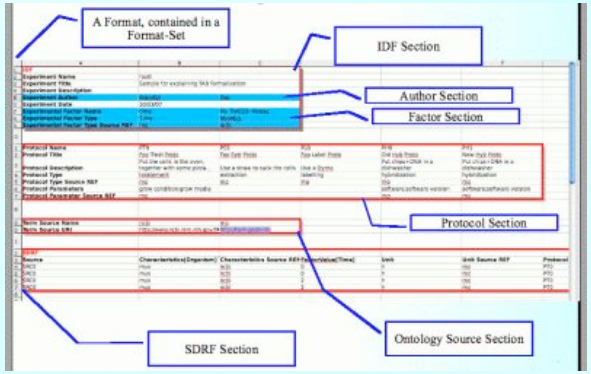
Ref Handler

Investigation Handler

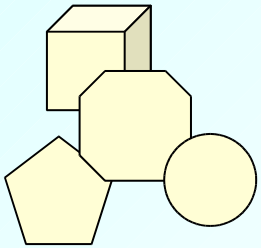
Study Handler

...

Parser/Loader



In-memory model of the TAB
Automatically done from TAB formal definition



Object Model (InvIndex + extensions)

InvIndex save()

MAGE-TAB files

PRIDE-Harvest files

InvIndex Dispatcher

MAGE-TAB Dispatcher

PRIDE Dispatcher

Dispatcher

TAB Modeling

A Format, contained in a Format-Set

IDF Section

Author Section

Factor Section

Protocol Section

Ontology Source Section

SDRF Section

	A	B	C	D	E	F	G
1	IDF						
2	Experiment Name	foo0					
3	Experiment Title	Sample for explaining TAB formalization					
4	Experiment Description						
5	Experiment Author	Brandizi	Foo				
6	Experiment Date	30/03/07					
7	Experimental Factor Name	time	My Th/CD3- Mouse				
8	Experimental Factor Type	Time	MusMus				
9	Experimental Factor Type Source REF	mo	ncbi				
10							
11	Protocol Name	PT0	PE0	PL0	PH0	PH1	
12	Protocol Title	Foo Treat Proto	Foo Extr Proto	Foo Label Proto	Old Hyb Proto	New Hyb Proto	
13	Protocol Description	Put the cells in the oven, together with some pizza...	Use a straw to suck the cells	Use a Dymo	Put chips+DNA in a dishwasher	Put chips+DNA in a dishwasher	
14	Protocol Type	treatement	extraction	labelling	hybridization	hybridization	
15	Protocol Type Source REF	mo	mo	mo	mo	mo	
16	Protocol Parameters	grow condition;grow media			software;software version	software;software version	
17	Protocol Parameter Source REF	mo			mo	mo	
18							
19	Term Source Name	ncbi	mo				
20	Term Source URI	http://www.ncbi.nlm.nih.gov/T	http://to/mged/onto				
21							
22	SDRF						
23	Source	Characteristics[Organism]	Characteristics Source REF	FactorValue[Time]	Unit	Unit Source REF	Protocol
24	SRC0	mus	ncbi	0	h	mo	PT0
25	SRC0	mus	ncbi	0	h	mo	PT0
26	SRC0	mus	ncbi	2	h	mo	PT0
27	SRC0	mus	ncbi	2	h	mo	PT0
28							

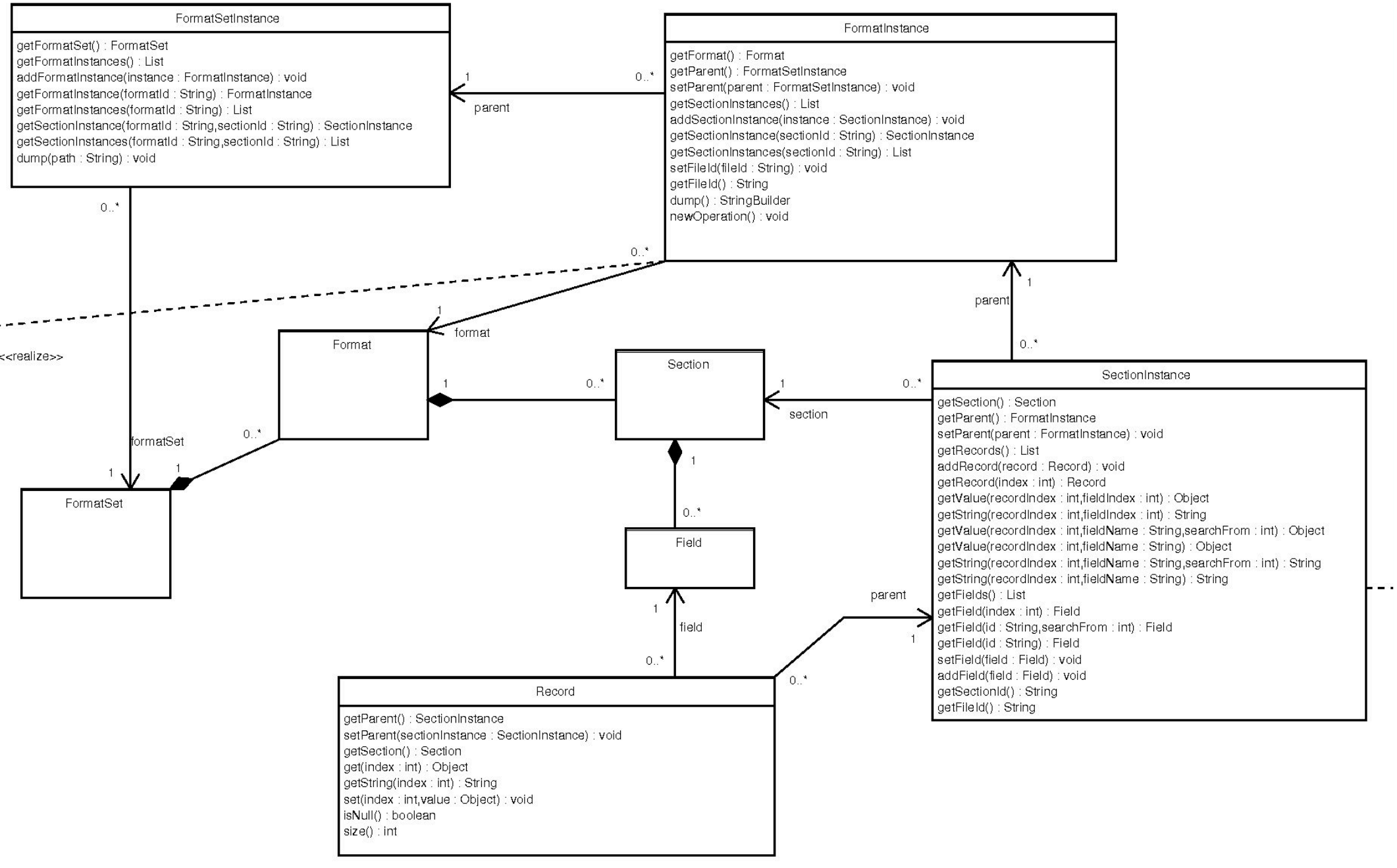
TAB Model Description


```
<format-set id = "isatab">

  <format id = "investigation">

    <section id = "ontoSources" type = "many-in-cols" header = "Ontology Source
             References">
       <field id = "Term Source Name" />
       <field id = "Term Source File" />
       <field id = "Term Source Version" />
       <field id = "Term Source Description" />
    </section>

    <section id = "investigation" type = "one-in-col" header = "Investigation">
      <field id = "Investigation Identifier" />
      <field id = "Investigation Title" />
      <field id = "Investigation Description" />
      <field id = "Date of Investigation Submission" />
      <field id = "Investigation Public Release Date" />
    </section>

    ...
```

TAB Model Description

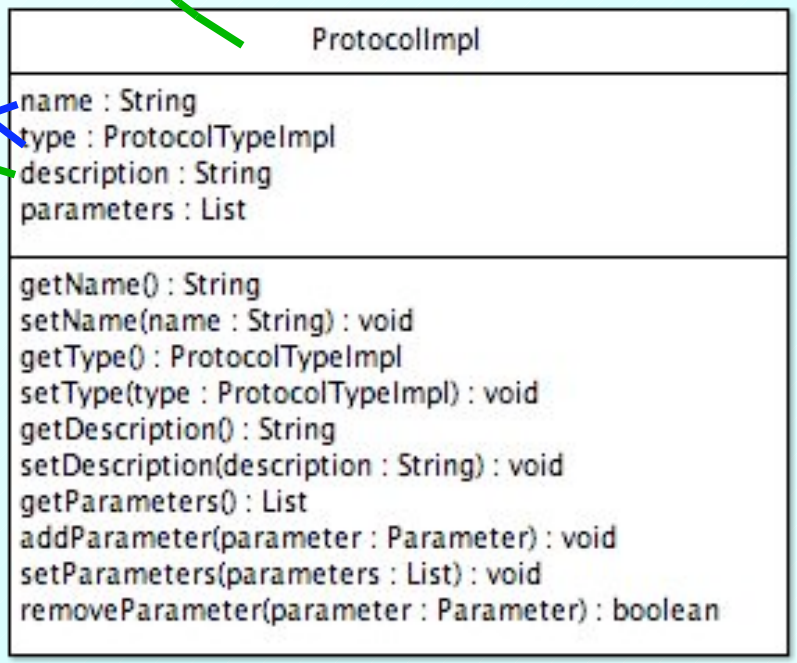

```
<format id = "transcriptomics_assay">
  <section id = "transcriptomics_pipeline" type = 'many-in-rows' header = "Assays">
    <field id = "Sample Name" />

    <field id = "Description" />
    <field id = "Characteristics" type = "typed(string)" />
    <field id = "Factor Value" type = "typed(string)" />
    <field id = "Unit" />
    <field id = "Term Source REF" />
    <field id = "Term Accession Number" />

    <field id = "Protocol REF"  />
    <field id = "Parameter" type = "typed(string)" />

    <field id = "Extract Name" />

    <field id = "Labeled Extract Name" />
    <field id = "Label" />

    <field id = "Hybridization Name" />

    <field id = "Scan Name" />
    <field id = "Array Data File" />
    <field id = "Array Data Matrix File" />

    <field id = "Normalization Name" />
    <field id = "Derived Array Data File" />
    <field id = "Derived Array Data Matrix File" />

  </section>
```

TAB Model Description



TAB Model Mapping to Object Model

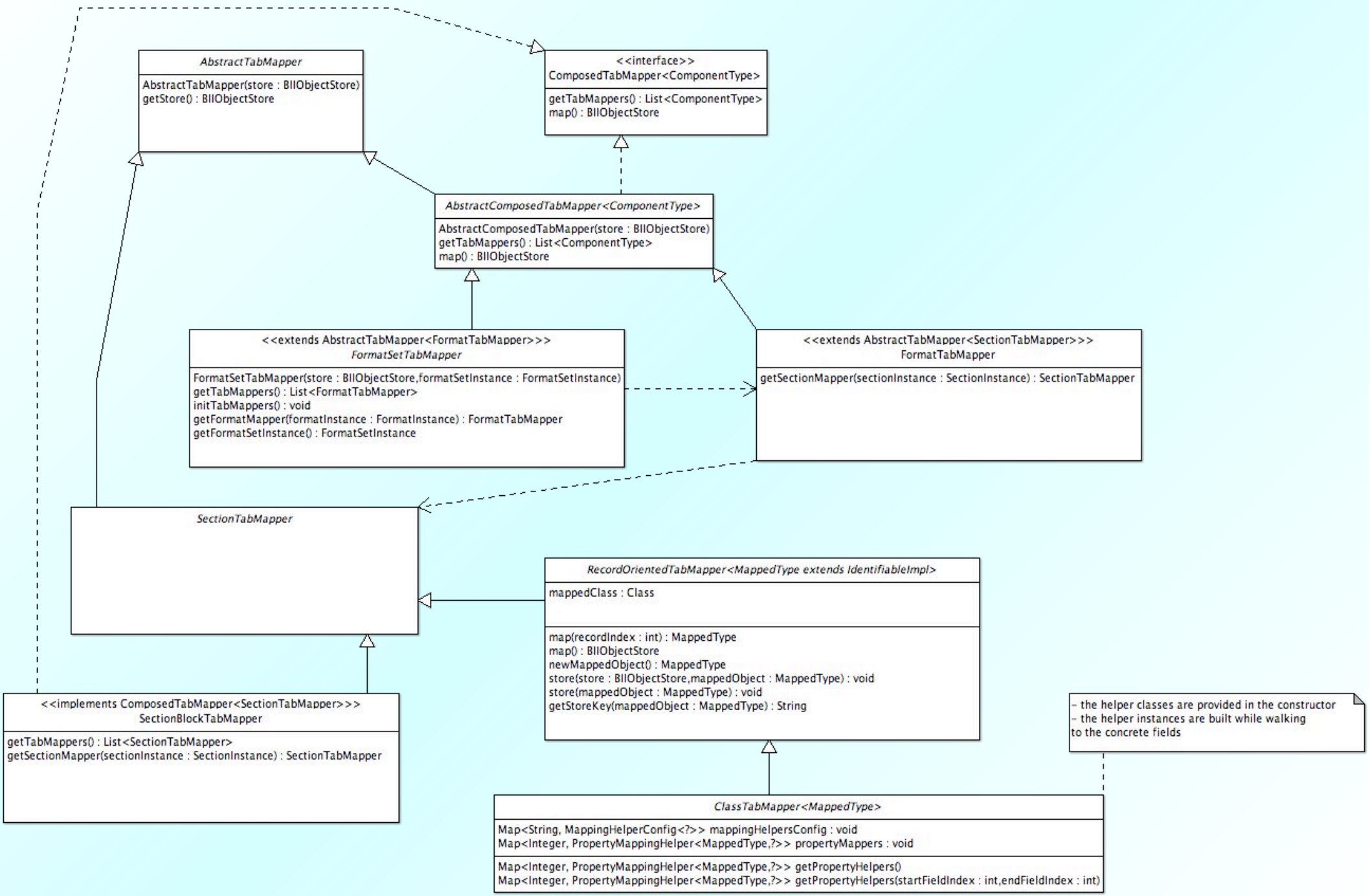
Protocol Name	PT0	PE0
Protocol Title	Foo Treat Proto	Foo Extr Proto
Protocol Description	Put the cells in the oven, together with some pizza...	Use a straw to suck the
Protocol Type	treatement	extraction
Protocol Type Source REF	mo	mo
Protocol Parameters	grow condition;grow media	
Protocol Parameter Source REF	mo	



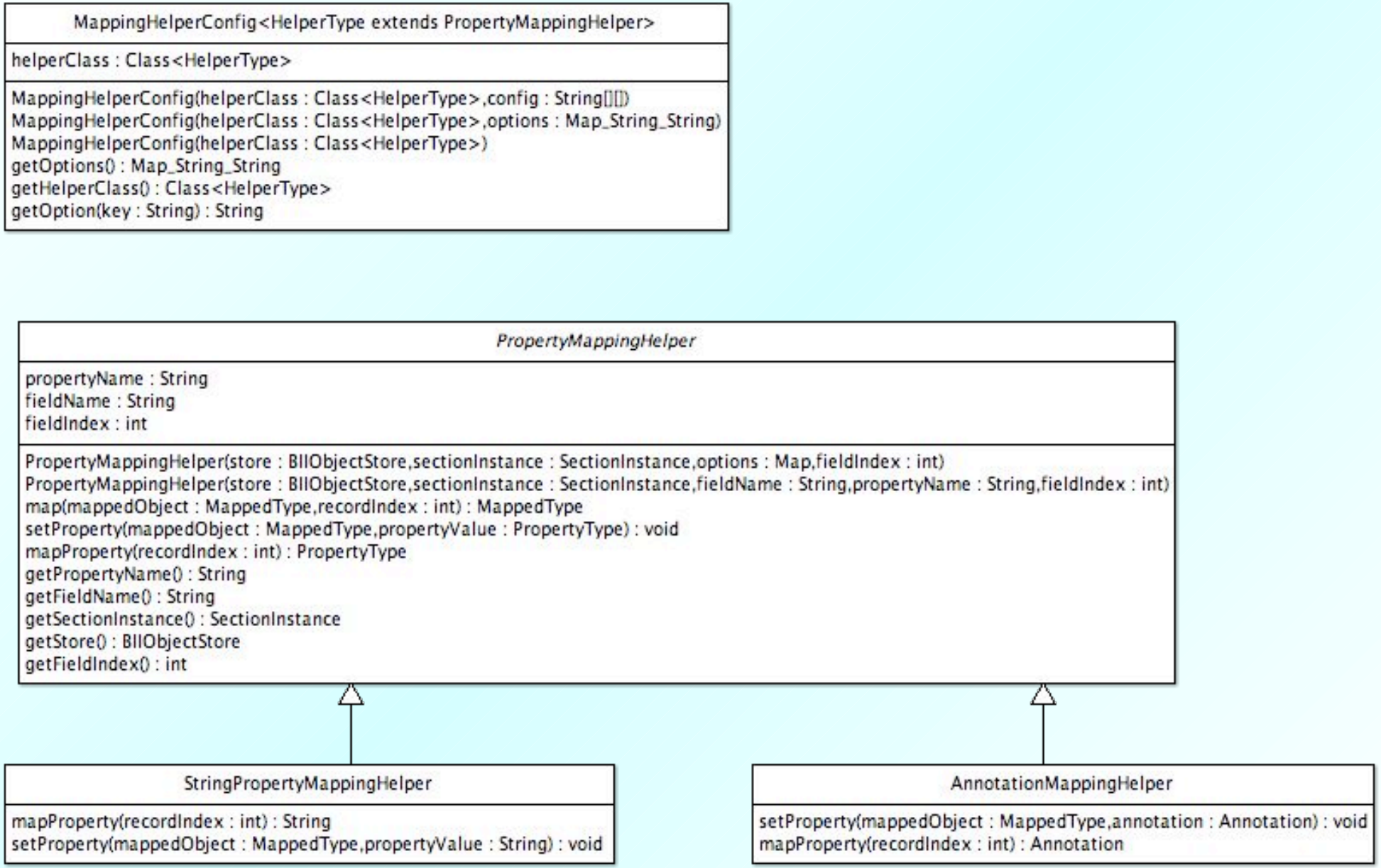
ProtocolImpl
name: PT0
title: Foo Treat Proto
description: Put the cells in the oven...
...

ProtocolImpl
name: PE0
title: Foo Extr Proto
description: Use a straw to suck the cells...
...

TAB Model Mapping



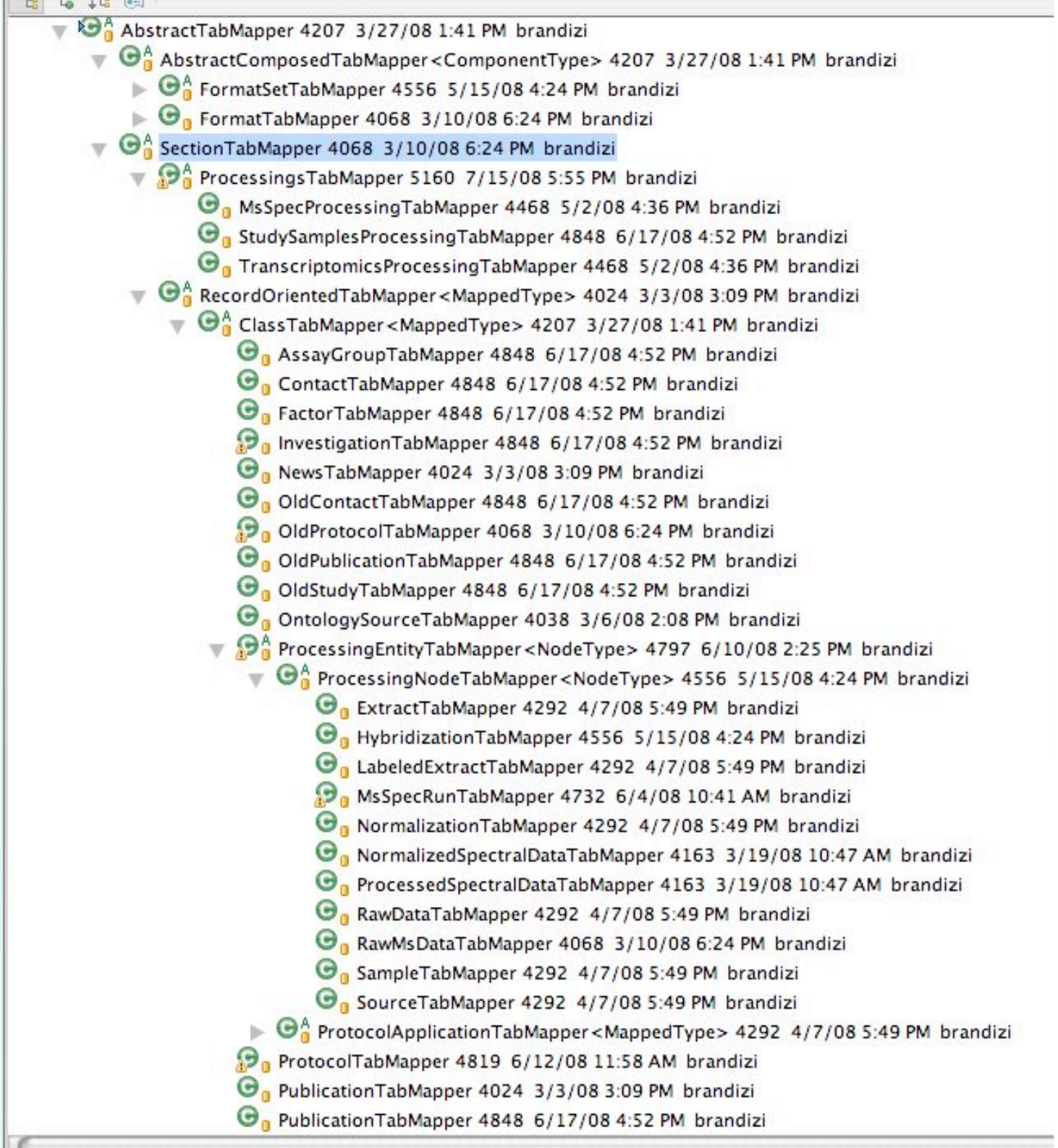
TAB Model Mapping



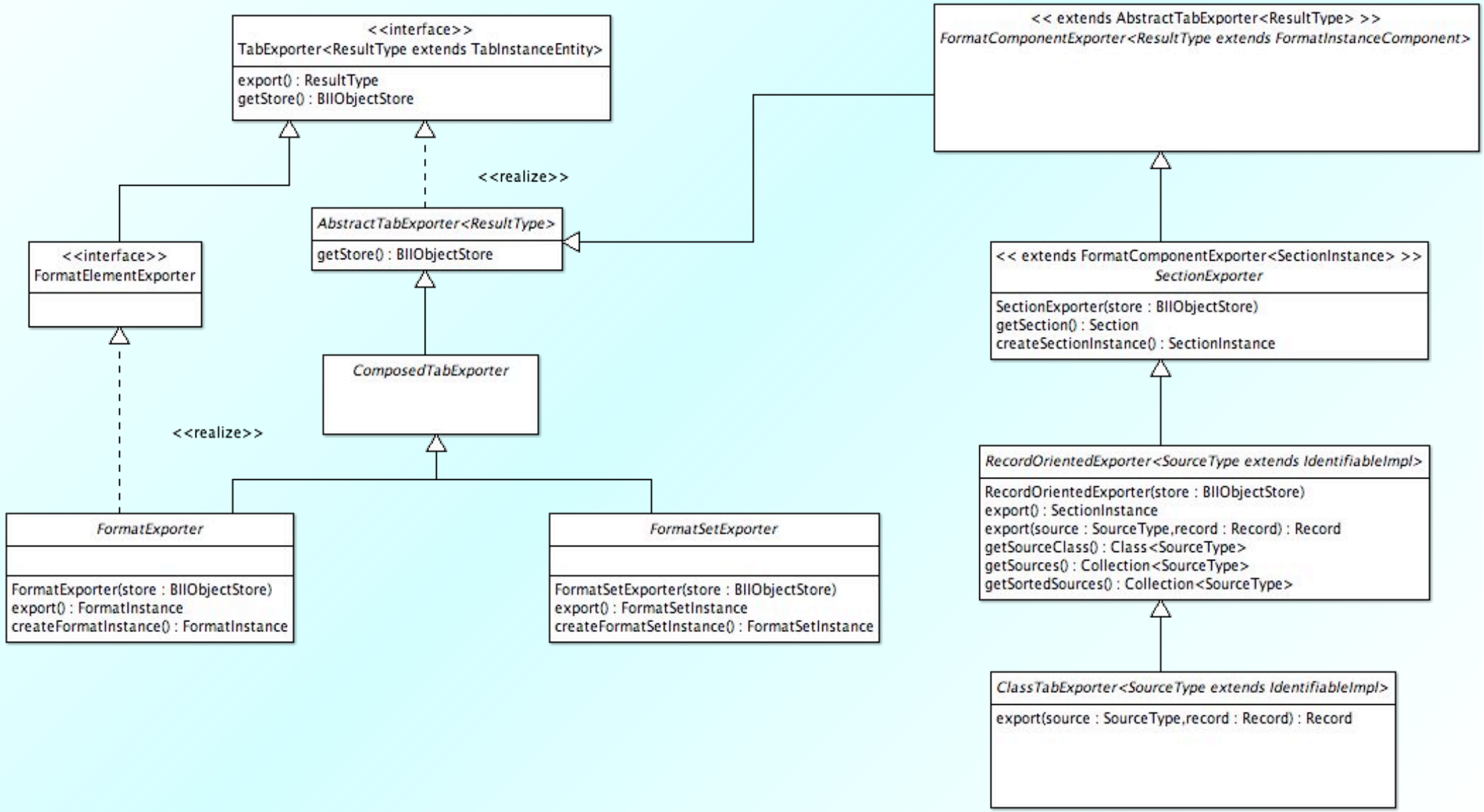
Mapping Example Code


```
public class ProtocolTabMapper extends ClassTabMapper<ProtocolImpl> implements
StudyComponentTabMapper
{
    private MappingUtils mappingUtils;
    private StudyImpl mappedStudy;

    @SuppressWarnings("unchecked")
    public ProtocolTabMapper ( BIIObjectStore store, SectionInstance sectionInstance )
    {
        super ( store, sectionInstance );

        mappingHelpersConfig.put ( "Protocol Name", new
          MappingHelperConfig<StringPropertyMappingHelper> (
            StringPropertyMappingHelper.class, new String[][] { { "propertyName", "name" } }
        ));

        mappingHelpersConfig.put ( "Protocol Description", new
          MappingHelperConfig<StringPropertyMappingHelper> (
            StringPropertyMappingHelper.class,
            new String[][] { { "propertyName", "description" } }
        ));

        mappingUtils = new MappingUtils ( store );
    }

    ...
```

ISA-TAB mappers



Object->TAB export



Outline

The BII Project and its components

The Submission Tool

The Submission Tool Internals

Thoughts, Future

Submission Tool: Other Components

- **MAGE-TAB dispatcher**:
  - Produces MAGE-TAB files which by using information in the ISA-TAB.
  - Uses the BII model + exporter classes + union of Study and Sample files
- **PRIDE-ML** dispatcher
  - Uses BII model + 3 identification spreadsheets (proteins, peptides, modifications) + mzData files
  - XML generation is realised by using PRIDE-API (i.e.: model-to-model mapping)
  - mzData is injected into the PRIDE-ML as-is (comes from hardware). TODO: XML file injection instead of file->object->file
  - spreadsheets are wrapped by iterators which read rows and output PRIDE-API objects (directly from the reader, keeping in memory only one object at time)

To-do and future

- Some items still not covered
  - Multiple designs, multiple contacts
  - Recent changes (e.g.: Protocol Components instead of Instruments and Software)
  - Pooling
- Related features to be implemented
  - Unloading from the BII database
  - Easy way for Accessions insertion (when feedback is received from AE or PRIDE)
- Metabolomics file repository (will be the early, very simple version of MeDA)
- Validation
  - I have only the very basics at the moment
  - To work with Eammon for checking the code written for the creator
- TAB description in the ISA-TAB creator, possible unification

Future: TAB-MDA

- Will the TAB format become popular?
- The mapping could be formally defined, either via an XML language or via Java annotations (pro and cons in both cases)
- From the mapping a number of things could be done automatically:
  - From TAB model to the Java classes + TAB parser + TAB exporter. Or from Java classes to TAB parser + TAB exporter
  - From TAB model + editing information to creation of an editor
  - From TAB model + rules to creation of validator
  - Close to us: we already have FUGE-ISATAB (made by Philippe via XSL). ISA-TAB=>FUGE possible by mapping ISA-TAB to FUGE-tk
    - already possible via programmatic approach

Thanks to...

- Susanna Sansone, project leader
- Philippe Rocca Serra, coordinator, ISA-TAB format
- Nataliya Sklyar, BII model and Web Application
- Eamonn Maguire, ISA-TAB Creator
- Phil Jones, PRIDE group, PRIDE support
- Chris Taylor, help with the model and the format
- Misha Kapushesky, Ibrahim Emam, Ugis Sarkans (and ArrayExpress people), help with unification of the model
- People from the ISA-TAB workshops

...and to you all!

By the way, sorry about...