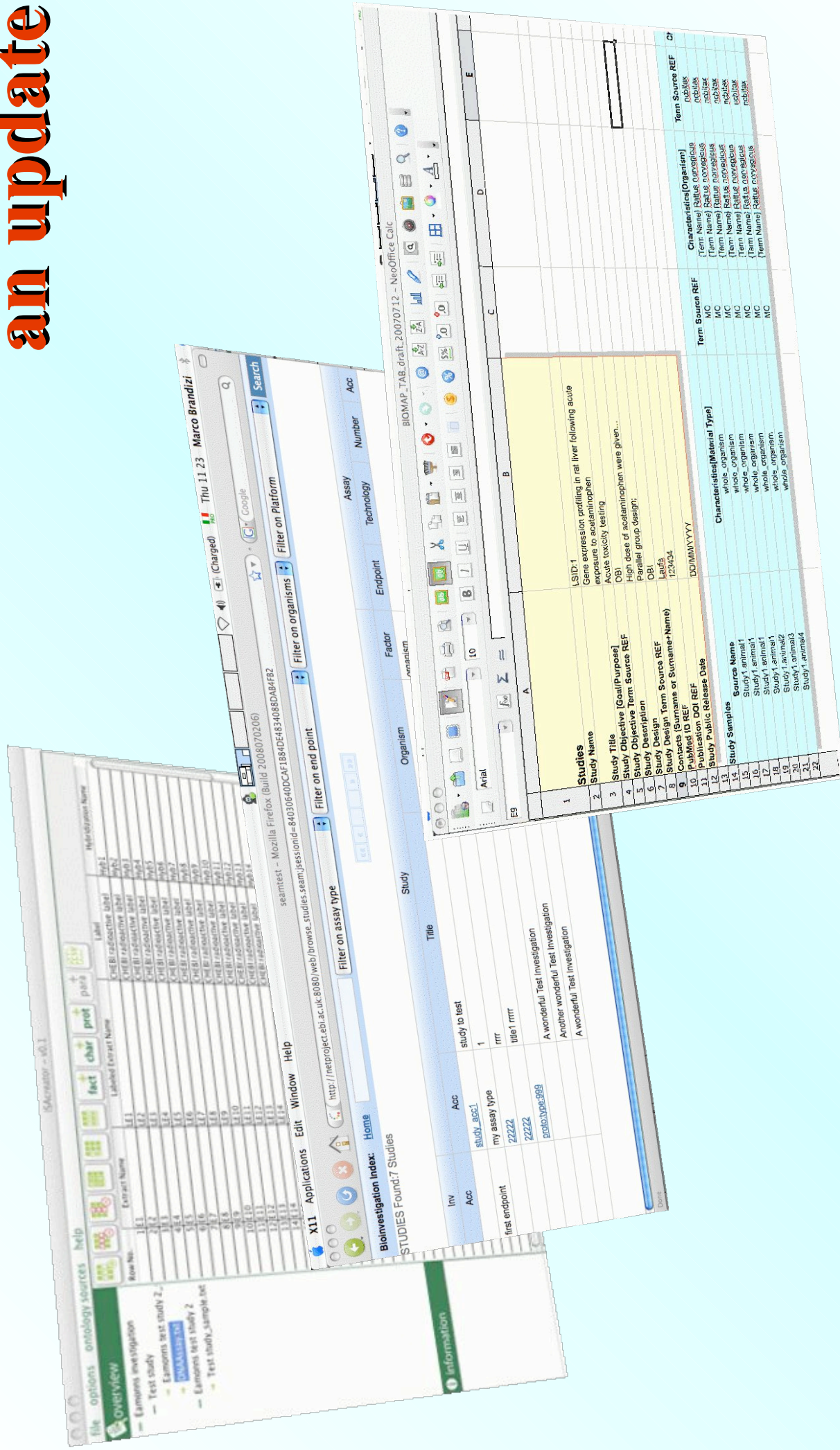


# The BioInvestigation Index Project: an update



Marco Brandizi, 24 Jul 2008



## **The BII Project and its components**

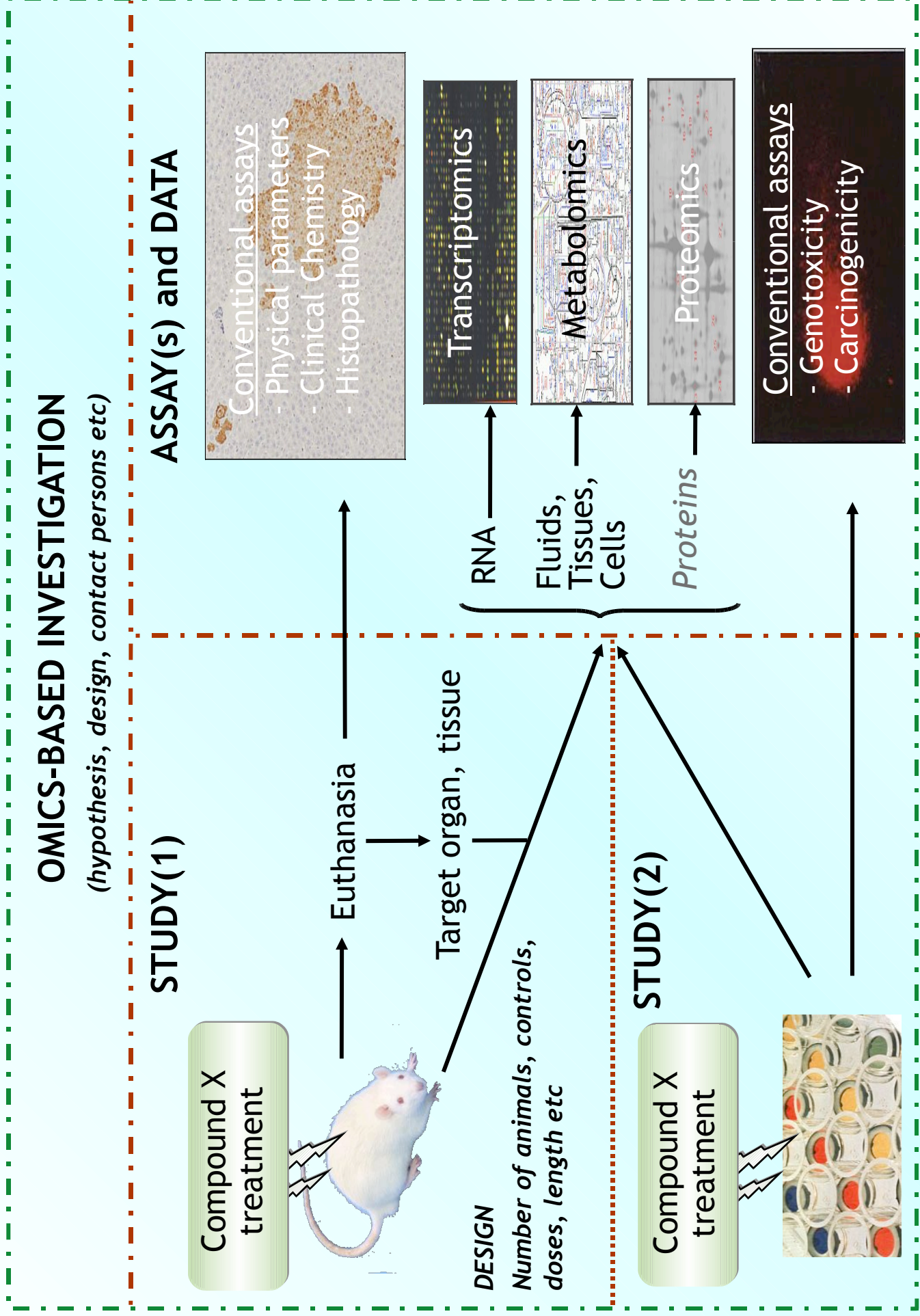
### **The Submission Tool**

### **The Submission Tool Internals**

### **Thoughts, Future**

# BII: Rationale

3





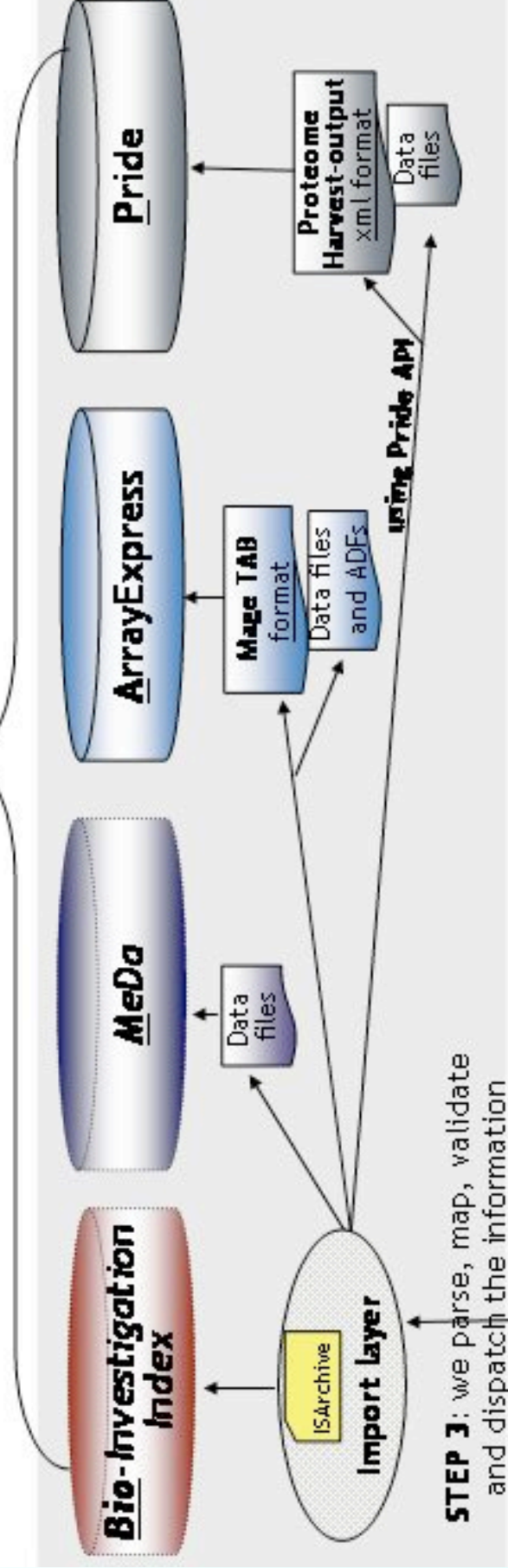
# BII: General Architecture

4

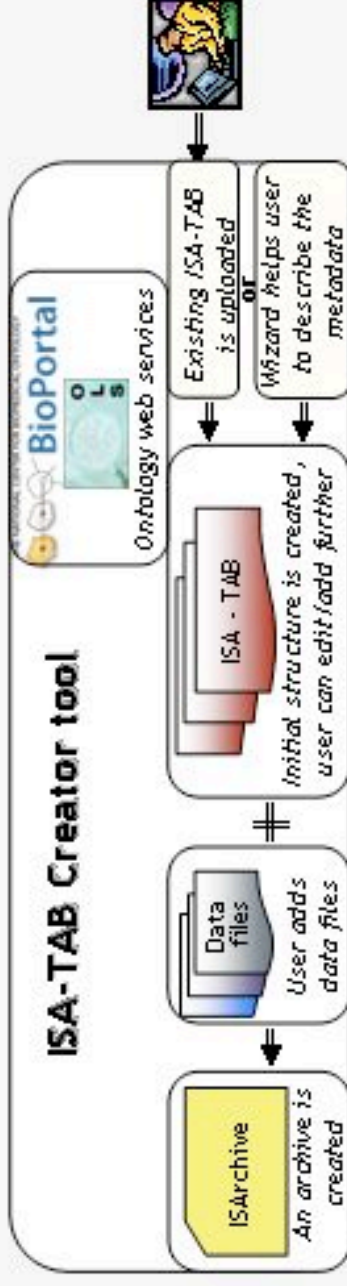
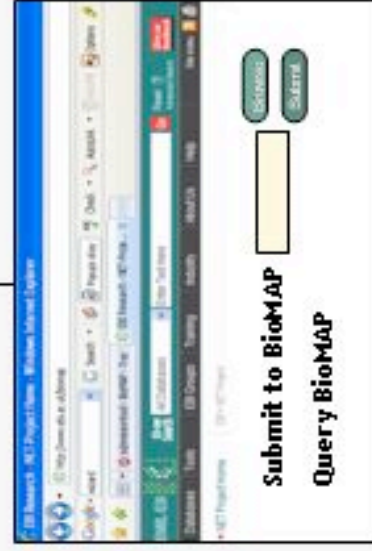
Prototype funded by:



## BioMAP - Components



**STEP 3:** we parse, map, validate and dispatch the information

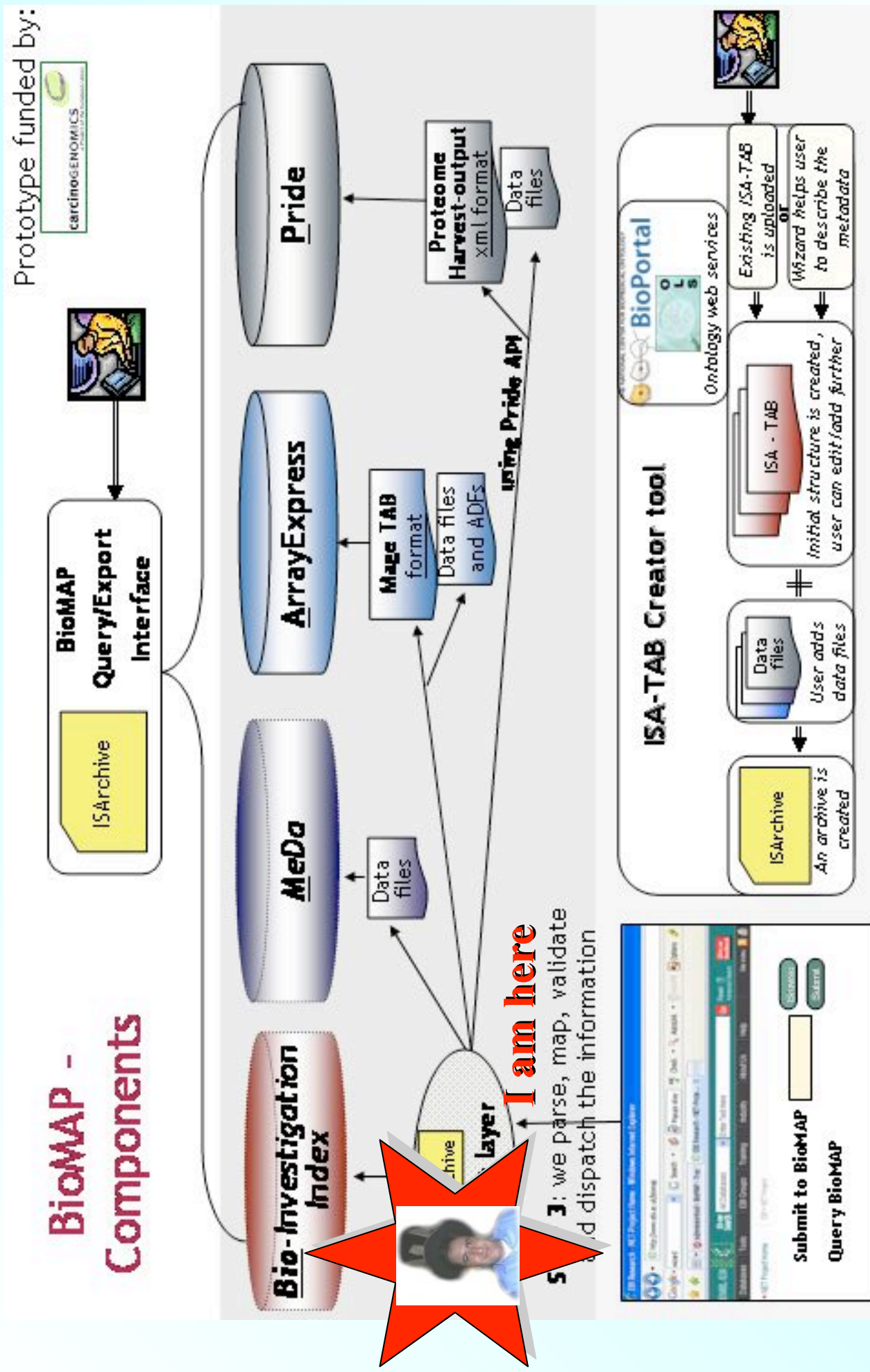


**STEP 2:** user upload the ISArchive to the BioMAP website

**STEP 1:** user describe metadata, associate the data file and create an ISArchive; or upload an existing ISA-TAB to be modified

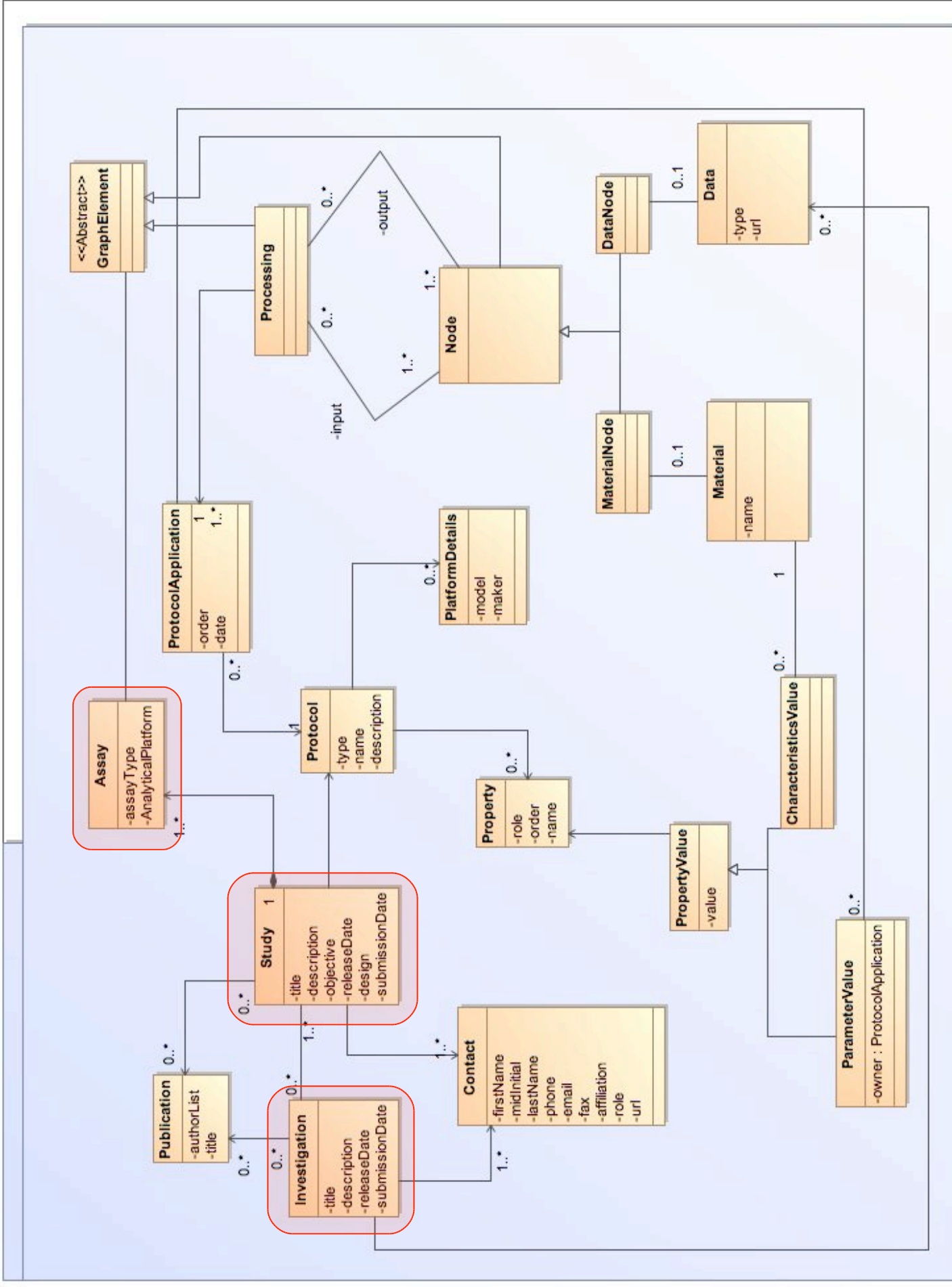
# BII: General Architecture

5





# The BioInvestigationIndex Model



**The BII Project and its components**

**The Submission Tool**

**The Submission Tool Internals**

**Thoughts, Future**

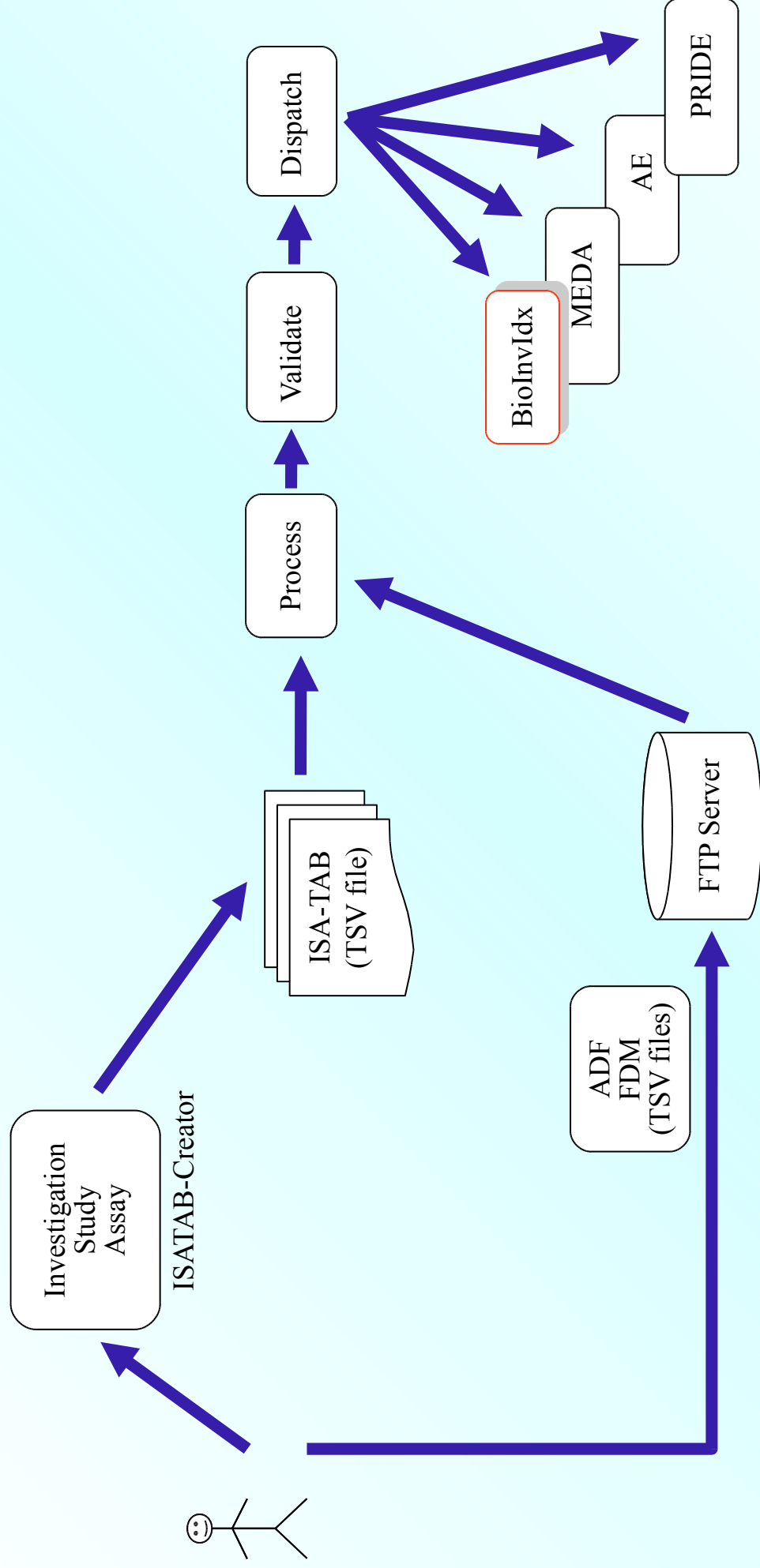
# The ISA-TAB format

- A tabular format inspired to similar cases (MAGETAB, 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

10

```
16:31:30 [brandizi@poderosaiii submission-tool]$ cd mvn_bin/
16:31:34 [brandizi@poderosaiii mvn_bin]$ ls
README               config.properties  persist.sh*      validate.sh*
ae_dispatch.sh*      log4j.properties  pride_dispatch.sh*
16:31:34 [brandizi@poderosaiii mvn_bin]$ ./persist.sh
```

Imports from an ISA-TAB archive into the BII database. Usage:

```
persist.sh <source-path> [<dest-path>]
```

where <source-path> is the path of the ISATAB submission. I expect to find at least investigation.csv in this directory, as well as any other ISA-TAB files and data files.

<dest-path> is the directory where to save persistence execution report and log.  
Default is '`<source-path>/export`'.

## Related Files

- `config.properties`: stores properties for connecting the target database and other other properties.
- `log4j.properties`: defines the logging properties (you shouldn't need to change it)

```
16:31:44 [brandizi@poderosaiii mvn_bin]$ ./persist.sh /Users/brandizi/Documents/Work/ebi/biomap/svn/submission-tool/src/test/resources/test-data/biomap/isatab/example[]
```

# Submission Tool: Example

seamtest - Mozilla Firefox (Build 2008070206)

http://localhost:8080/web/study.seam?studyid=Isid%3A19239&cid=4

☆

Google

Filter on assay type

Filter on end point

Filter on organisms

Filter on Platform

Search

Study ACC:

Isid:bii:study:19239

Title:

The Influence of Pharmacogenetics on Fatty Liver Disease in the Wistar and Kyoto Rats: A Combined Transcriptomic and Metabonomic Study

Description:

Analysis of liver tissue from rats exposed to orotic acid for 1, 3, and 14 days was performed by DNA microarrays and high resolution 1H NMR spectroscopy based metabonomics of both tissue extracts and intact tissue (n ) 3).

Release Date:

time course design

Design:

Species:

Factors:

Assays:

seamtest - Mozilla Firefox (Build 2008070206)

http://localhost:8080/web/browse\_studies.seam

☆

Google

BioInvestigation Index: Home

Filter on assay type

Filter on end point

Filter on organisms

Filter on Platform

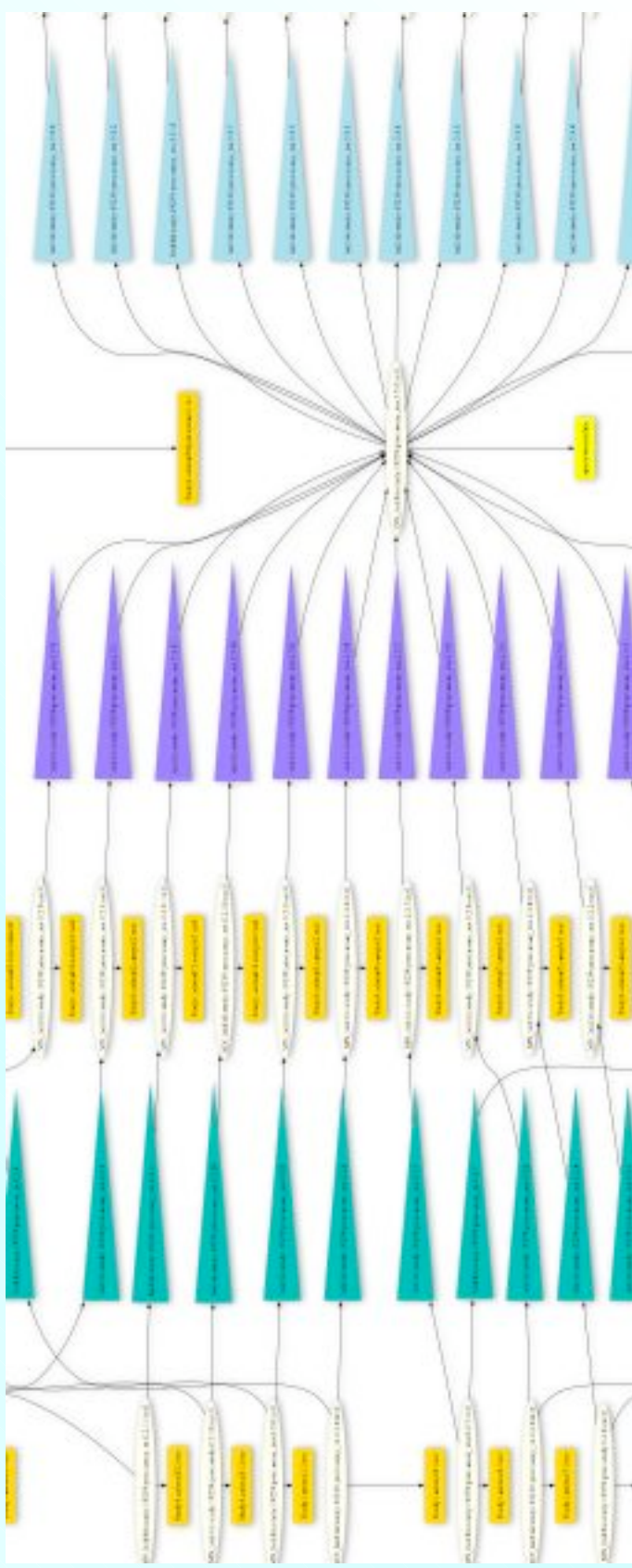
Search

STUDIES Found:16 Studies

Inv	Acc	Study	Assay
		Title	Endpoint Technology Nu
bii:-1	Study -1		
bii:-2	DC study		
bii:-3	NormX Study -3		
bii:-4	Study -4		
bii:-5	Study -5		
Isid: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		
Isid:bii:investigation:999	A Test Investigation, made for testing purposes		



## 12



**The BII Project and its components**

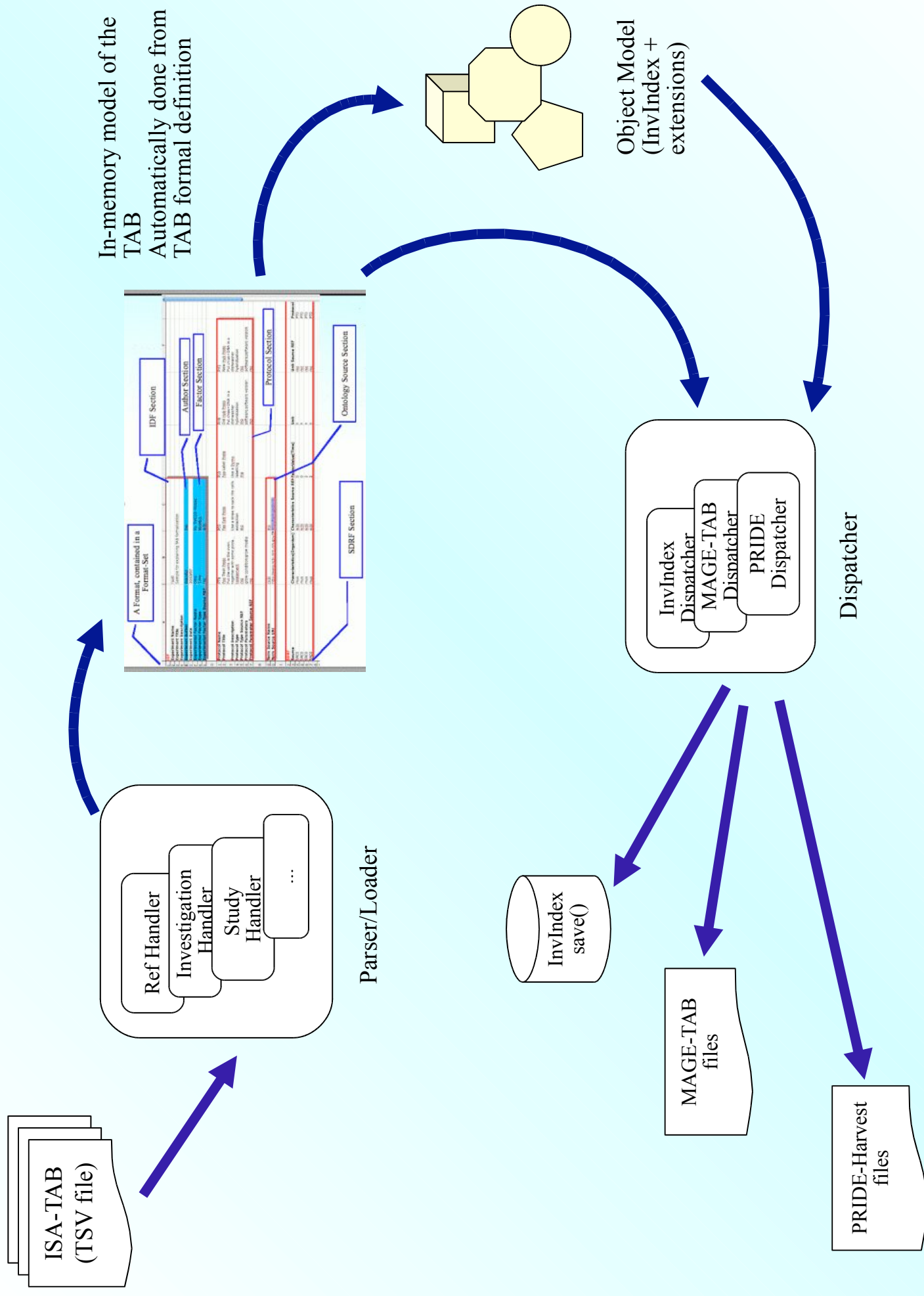
**The Submission Tool**

**The Submission Tool Internals**

**Thoughts, Future**

# Submission Tool Impl.: General Approach

14





# 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	F
1	<b>IDF</b>			
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	
0				
1	Protocol Name	PT0	PE0	PH1
2	Protocol Title	Foo Treat Proto	Foo Extr Proto	Old Hyb Proto
3	Protocol Description	Put the cells in the oven, together with some pizza...	Use a straw to suck the cells	Put chips+DNA in a dishwasher
4	Protocol Type	treatment	extraction	hybridization
5	Protocol Type Source REF	mo	mo	mo
6	Protocol Parameters	grow condition;grow media		software;software version
7	Protocol Parameter Source REF	mo		mo
8				
9	Term Source Name	ncbi	mo	
0	Term Source URI	<a href="http://www.ncbi.nlm.nih.gov/Taxonomy/ncbi/mged/onto">http://www.ncbi.nlm.nih.gov/Taxonomy/ncbi/mged/onto</a>		
1				
2	<b>SDRF</b>			
3	Source	Characteristics[Organism]	Characteristics Source REF	Unit
4	SRC0	mus	ncbi	h
5	SRC0	mus	ncbi	h
6	SRC0	mus	ncbi	h
7	SRC0	mus	ncbi	h
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7				
8				
9				
0				
1				
2				
3				
4				
5				
6				
7</				

# 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" />

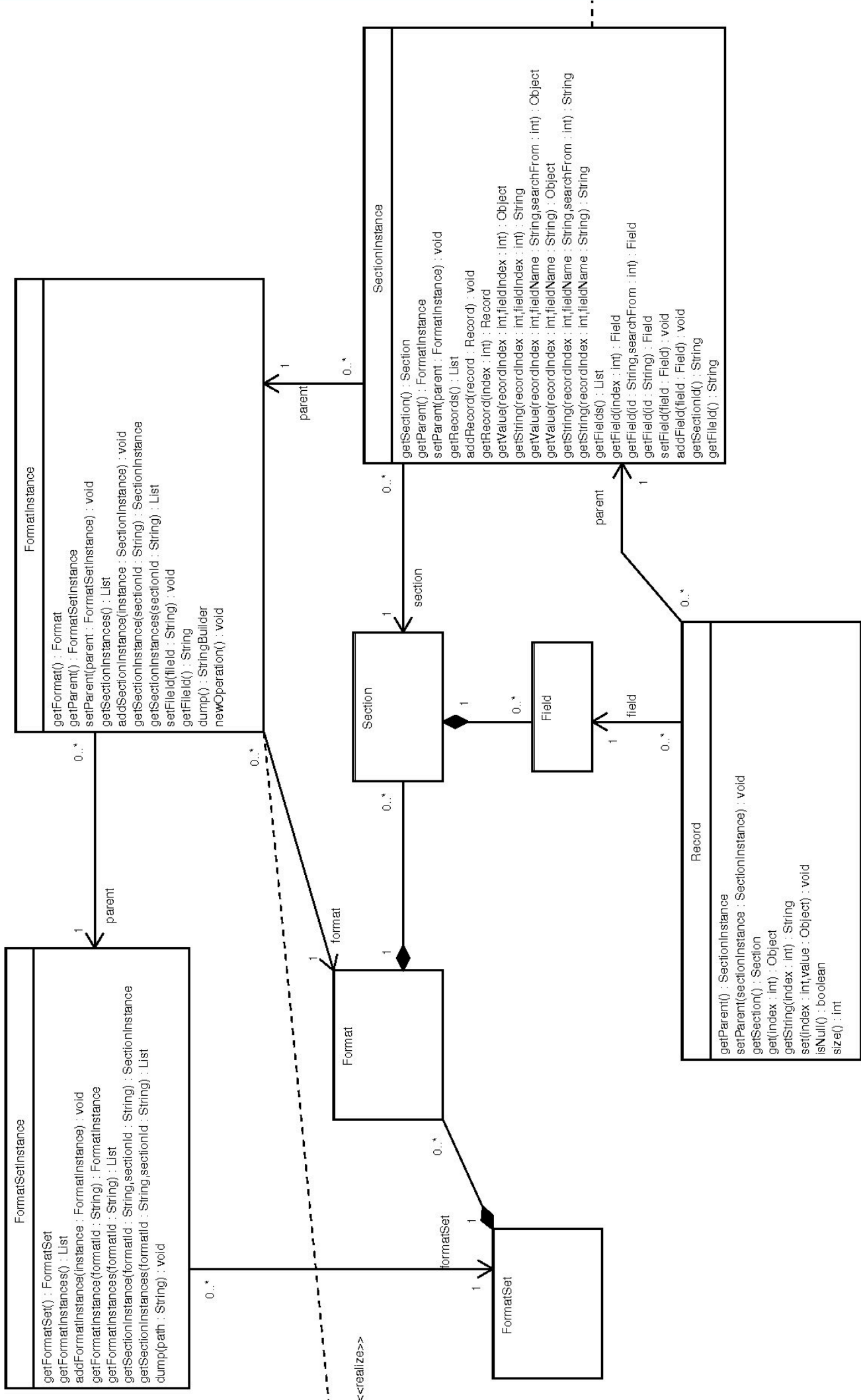
    <field id = "Hybridization Name" />

    <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 Mapping to Object Model

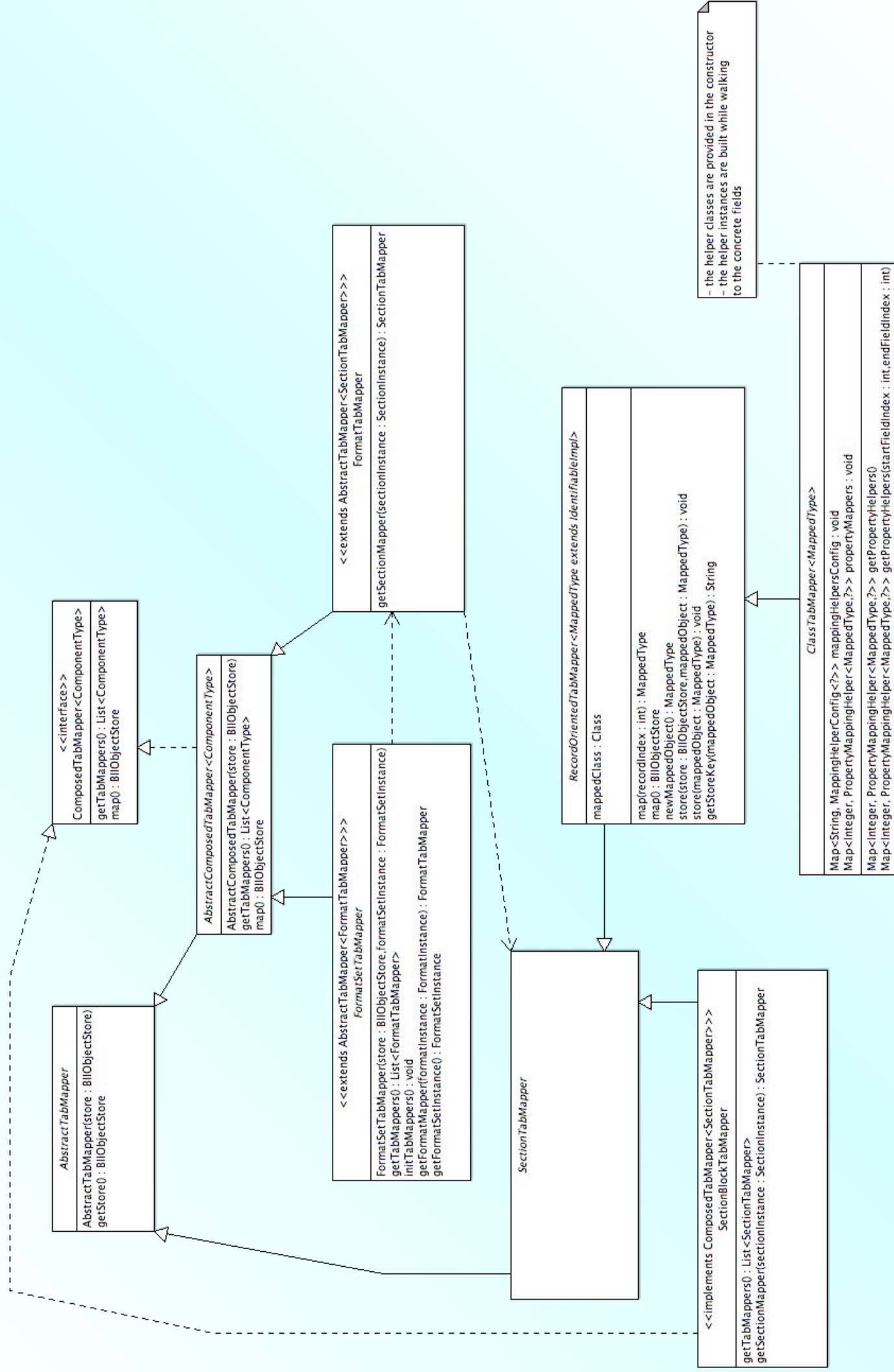
<b>Protocol Name</b>	PT0	PE0
<b>Protocol Title</b>	Foo Treat Proto	Foo Extr Proto
<b>Protocol Description</b>	Put the cells in the oven, together with some pizza...	Use a straw to suck the
<b>Protocol Type</b>	<u>treatment</u>	extraction
<b>Protocol Type Source REF</b>	mo	mo
<b>Protocol Parameters</b>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

ProtocolImpl
name : String type : ProtocolTypeImpl description : String parameters : List  getName() : String setName(name : String) : void getType() : ProtocolTypeImpl setType(type : ProtocolTypeImpl) : void getDescription() : String setDescription(description : String) : void getParameters() : List addParameter(parameter : Parameter) : void setParameters(parameters : List) : void removeParameter(parameter : Parameter) : boolean

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

MappingHelperConfig<HelperType extends PropertyMappingHelper>
helperClass : Class<HelperType>
MappingHelperConfig(helperClass : Class<HelperType>, config : String[[[]]) MappingHelperConfig(helperClass : Class<HelperType>, options : Map_String_String) MappingHelperConfig(helperClass : Class<HelperType>) getOptions() : Map_String_String getHelperClass() : Class<HelperType> getOption(key : String) : String

PropertyMappingHelper
propertyName : String fieldName : String fieldIndex : int  PropertyMappingHelper(store : BIIObjectStore, sectionInstance : SectionInstance, options : Map, fieldIndex : int) PropertyMappingHelper(store : BIIObjectStore, sectionInstance : SectionInstance, fieldName : String, fieldIndex : int) map(mappedObject : MappedType, recordIndex : int) : MappedType setProperty(mappedObject : MappedType, propertyValue : PropertyType) : void mapProperty(recordIndex : int) : PropertyType getPropertyName() : String getFieldName() : String getSectionInstance() : SectionInstance getStore() : BIIObjectStore getFieldIndex() : int

StringPropertyMappingHelper
mapProperty(recordIndex : int) : String setProperty(mappedObject : MappedType, propertyValue : String) : void

AnnotationMappingHelper
setProperty(mappedObject : MappedType, annotation : Annotation) : void mapProperty(recordIndex : int) : Annotation

# 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" } }
            ));

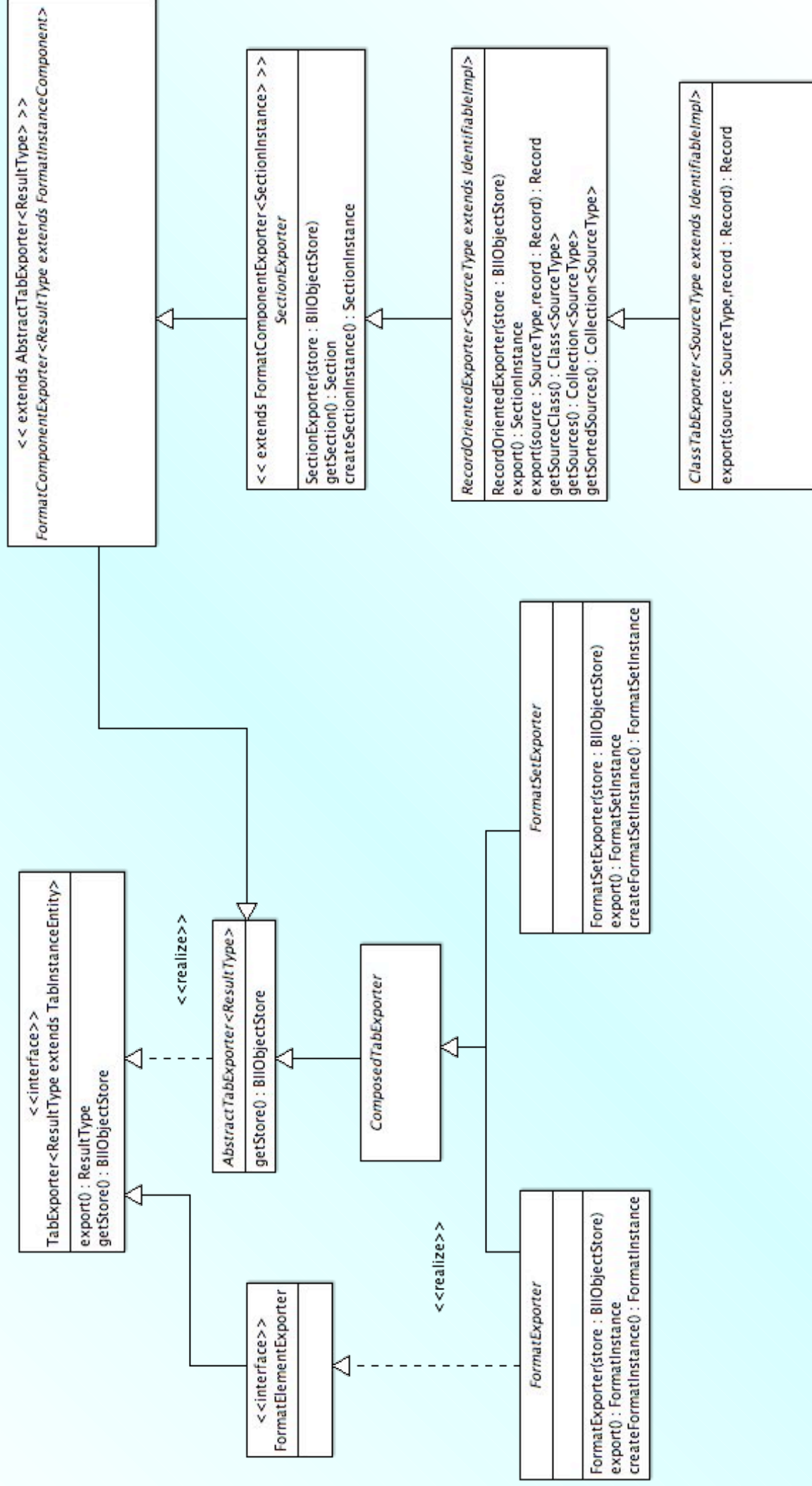
        mappingUtils = new MappingUtils ( store );
    }
    ...
}
```

# ISA-TAB mappers

▼	AbstractTabMapper	4207	3/27/08 1:41 PM	brandizi
▼	AbstractComposedTabMapper<ComponentType>	4207	3/27/08 1:41 PM	brandizi
▶	FormatSetTabMapper	4556	5/15/08 4:24 PM	brandizi
▶	FormatTabMapper	4068	3/10/08 6:24 PM	brandizi
▼	SectionTabMapper	4068	3/10/08 6:24 PM	brandizi
▼	ProcessingsTabMapper	5160	7/15/08 5:55 PM	brandizi
▶	MsSpecProcessingTabMapper	4468	5/2/08 4:36 PM	brandizi
▶	StudySamplesProcessingTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	TranscriptomicsProcessingTabMapper	4468	5/2/08 4:36 PM	brandizi
▼	RecordOrientedTabMapper<MappedType>	4024	3/3/08 3:09 PM	brandizi
▼	ClassTabMapper<MappedType>	4207	3/27/08 1:41 PM	brandizi
▶	AssayGroupTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	ContactTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	FactorTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	InvestigationTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	NewsTabMapper	4024	3/3/08 3:09 PM	brandizi
▶	OldContactTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	OldProtocolTabMapper	4068	3/10/08 6:24 PM	brandizi
▶	OldPublicationTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	OldStudyTabMapper	4848	6/17/08 4:52 PM	brandizi
▶	OntologySourceTabMapper	4038	3/6/08 2:08 PM	brandizi
▼	ProcessingEntityTabMapper<NodeType>	4797	6/10/08 2:25 PM	brandizi
▼	ProcessingNodeTabMapper<NodeType>	4556	5/15/08 4:24 PM	brandizi
▶	ExtractTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	HybridizationTabMapper	4556	5/15/08 4:24 PM	brandizi
▶	LabeledExtractTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	MsSpecRunTabMapper	4732	6/4/08 10:41 AM	brandizi
▶	NormalizationTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	NormalizedSpectralDataTabMapper	4163	3/19/08 10:47 AM	brandizi
▶	ProcessedSpectralDataTabMapper	4163	3/19/08 10:47 AM	brandizi
▶	RawDataTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	RawMsDataTabMapper	4068	3/10/08 6:24 PM	brandizi
▶	SampleTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	SourceTabMapper	4292	4/7/08 5:49 PM	brandizi
▶	ProtocolApplicationTabMapper<MappedType>	4292	4/7/08 5:49 PM	brandizi
▶	ProtocolTabMapper	4819	6/12/08 11:58 AM	brandizi
▶	PublicationTabMapper	4024	3/3/08 3:09 PM	brandizi
▶	PublicationTabMapper	4848	6/17/08 4:52 PM	brandizi



# Object->TAB export





**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

- 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...**

