Seedpod - A Model Driven Laboratory Information Management System*

Maya Li
MITRE TIE E549
March 19, 2008

* PhD thesis project, Department of Biomedical informatics, University of Washington. Thesis adviser is Dr. Jim Brinkley.

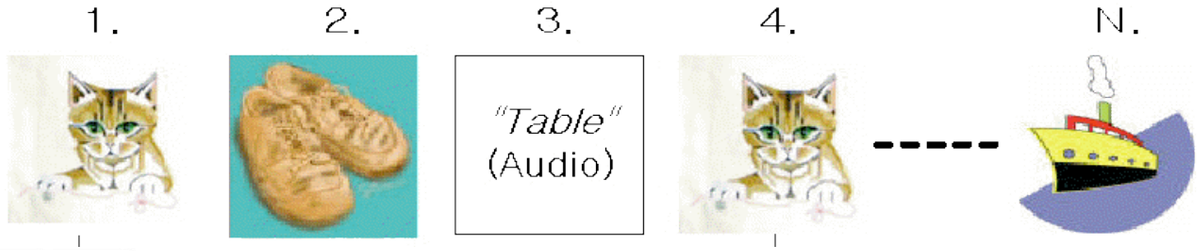
Problem Statement

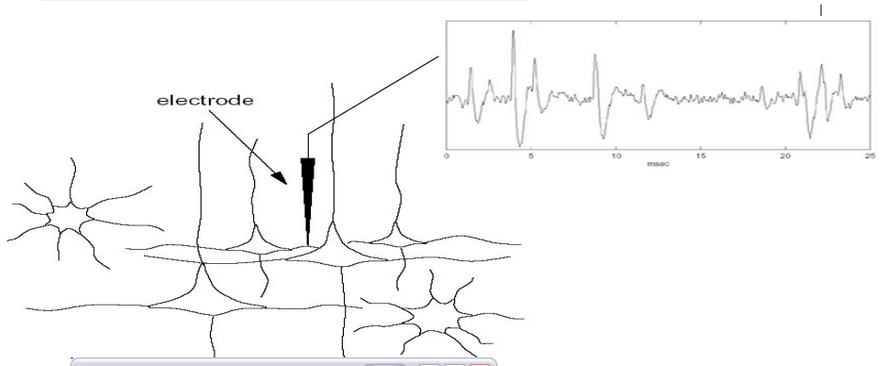
- Biomedical researchers need data management tools that are more robust than excel spreadsheets. However, the rate that biomedical informaticists have been developing laboratory information management systems (LIMS) can’t keep up with the demand for LIMS and the rate that experiment protocols change.

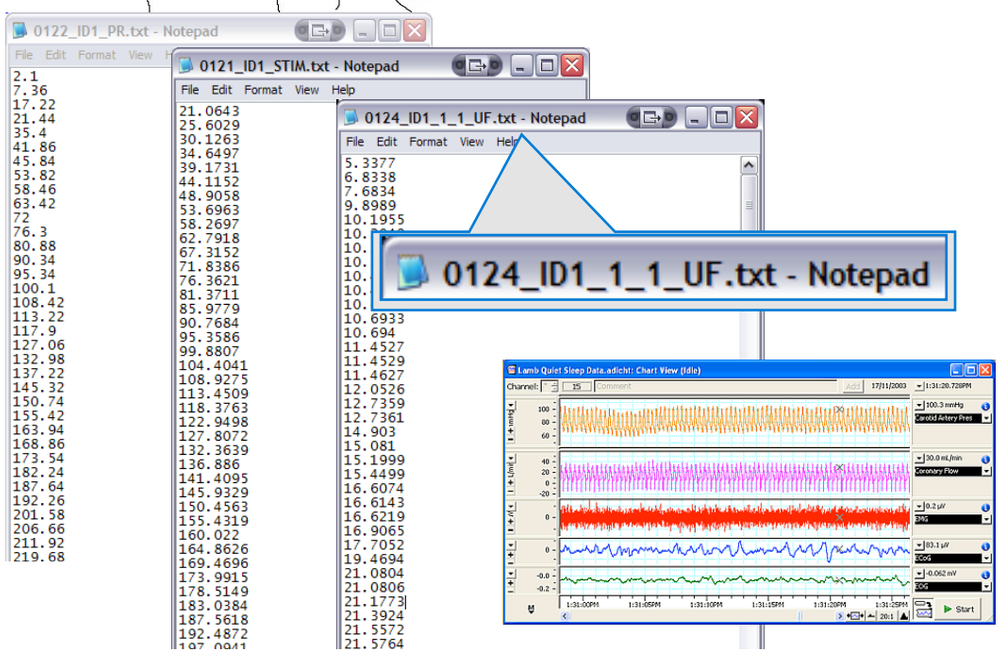
Background: Biomedical Research

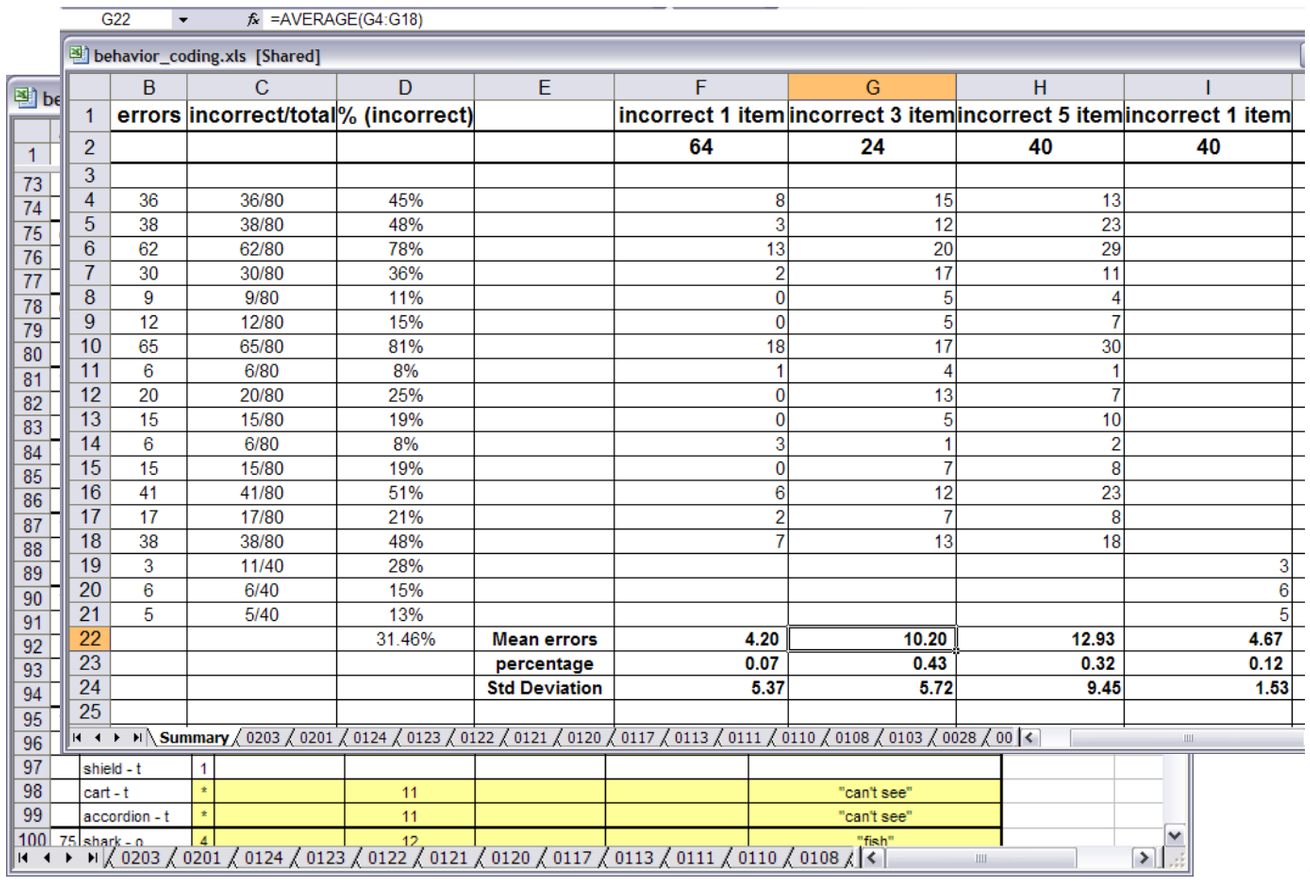
- UW research, single unit recording during neurosurgery

Experiment trials









G22 =AVERAGE(G4:G18)

behavior_coding.xls [Shared]

	B	C	D	E	F	G	H	I
1	errors	incorrect/total	% (incorrect)		incorrect 1 item	incorrect 3 item	incorrect 5 item	incorrect 1 item
2					64	24	40	40
3								
4	36	36/80	45%		8	15	13	
5	38	38/80	48%		3	12	23	
6	62	62/80	78%		13	20	29	
7	30	30/80	36%		2	17	11	
8	9	9/80	11%		0	5	4	
9	12	12/80	15%		0	5	7	
10	65	65/80	81%		18	17	30	
11	6	6/80	8%		1	4	1	
12	20	20/80	25%		0	13	7	
13	15	15/80	19%		0	5	10	
14	6	6/80	8%		3	1	2	
15	15	15/80	19%		0	7	8	
16	41	41/80	51%		6	12	23	
17	17	17/80	21%		2	7	8	
18	38	38/80	48%		7	13	18	
19	3	11/40	28%					3
20	6	6/40	15%					6
21	5	5/40	13%					5
22			31.46%	Mean errors	4.20	10.20	12.93	4.67
23				percentage	0.07	0.43	0.32	0.12
24				Std Deviation	5.37	5.72	9.45	1.53
25								

Summary | 0203 | 0201 | 0124 | 0123 | 0122 | 0121 | 0120 | 0117 | 0113 | 0111 | 0110 | 0108 | 0103 | 0028 | 00

97	shield - t	1			
98	cart - t	*	11		"can't see"
99	accordion - t	*	11		"can't see"
100	75 shark - p	4	12		"fish"

0203 | 0201 | 0124 | 0123 | 0122 | 0121 | 0120 | 0117 | 0113 | 0111 | 0110 | 0108

Background: Biomedical Data Management

- Needs from the biomedical researchers*
  - Data handling problem has become a barrier to progress of research
  - Available computational solutions are prohibitively expensive
  - Available solutions are too complex for their needs
  - No solution to their specific problems

*N.R. Anderson, ES Lee, JS Brockenbrough, ME Minie, S Fuller, J Brinkley, P Tarczy-Hornoch. Issues in Biomedical Research Data Management and Analysis: Needs and Barriers. J Am Med Inform Assoc. 2007;14:478-488. DOI 10.1197/jamia.M2114

Background: Biomedical Informatics

- LIMS development is challenging
  - Diversity in biomedical research makes LIMS solutions lab-specific.
  - Research protocol change frequently, every 4 months to a year.

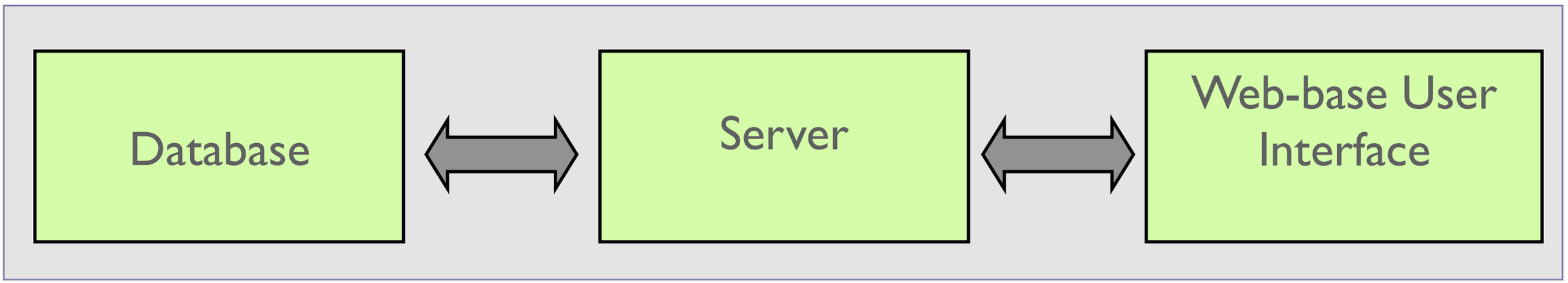
Biomedical Researchers' Wish List

- **Data management** (Seedpod)
- Data archiving
- Data analysis
- Protocol management
- Workflow management
- Data sharing

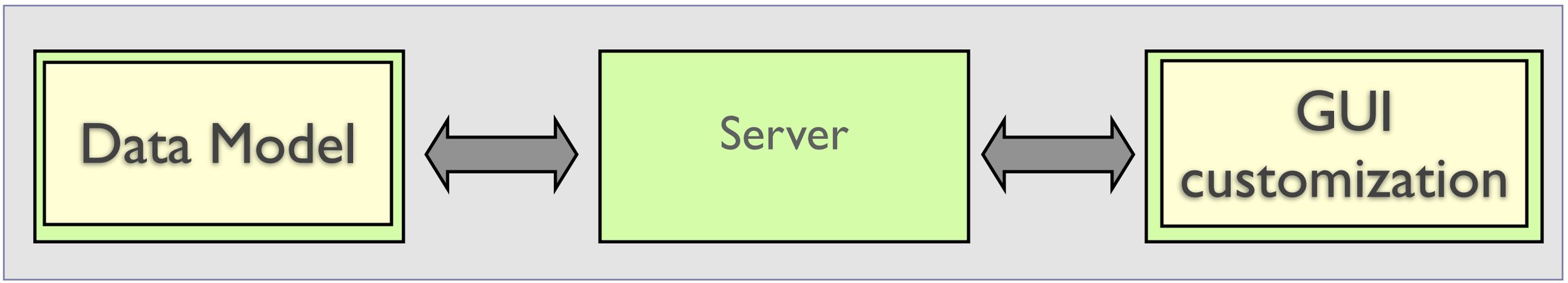
Aims

- **Data management** for researchers
  - Organize and store large heterogeneous data set
  - Enter and browse data through a user friendly interface
- **LIMS development** for informaticists
  - Cost-efficient development
  - Generalizable solution

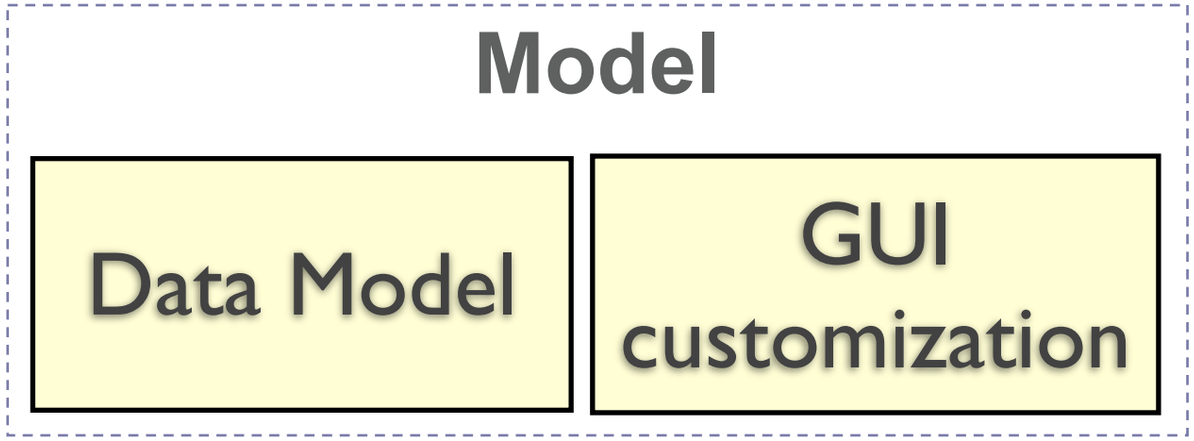
Seedpod Takes Model Driven Approach

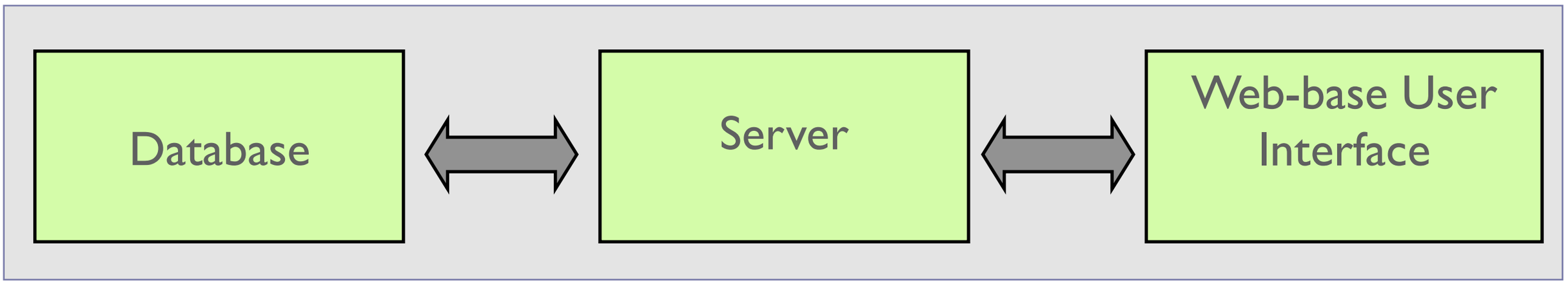


Seedpod Takes Model Driven Approach

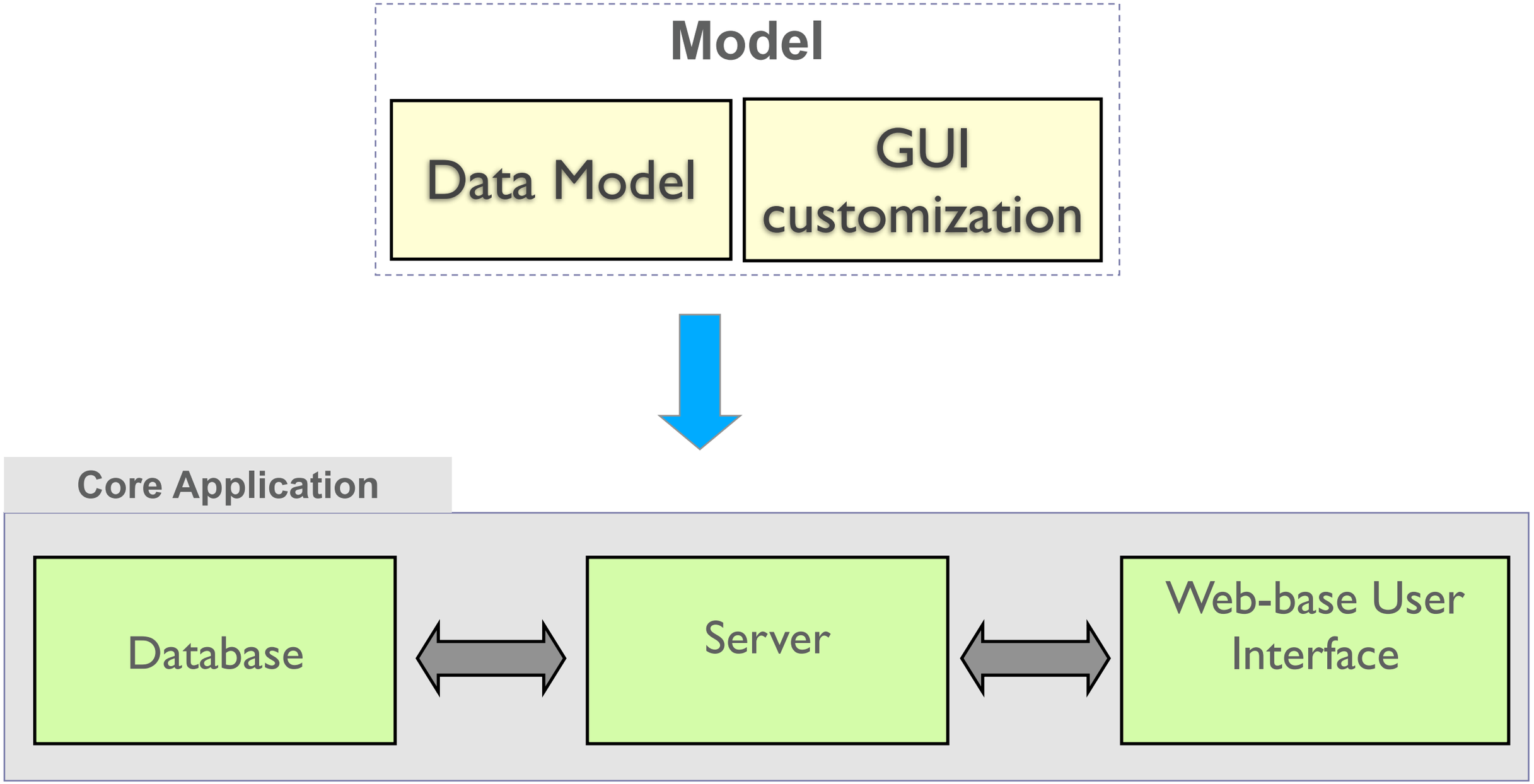


Seedpod Takes Model Driven Approach

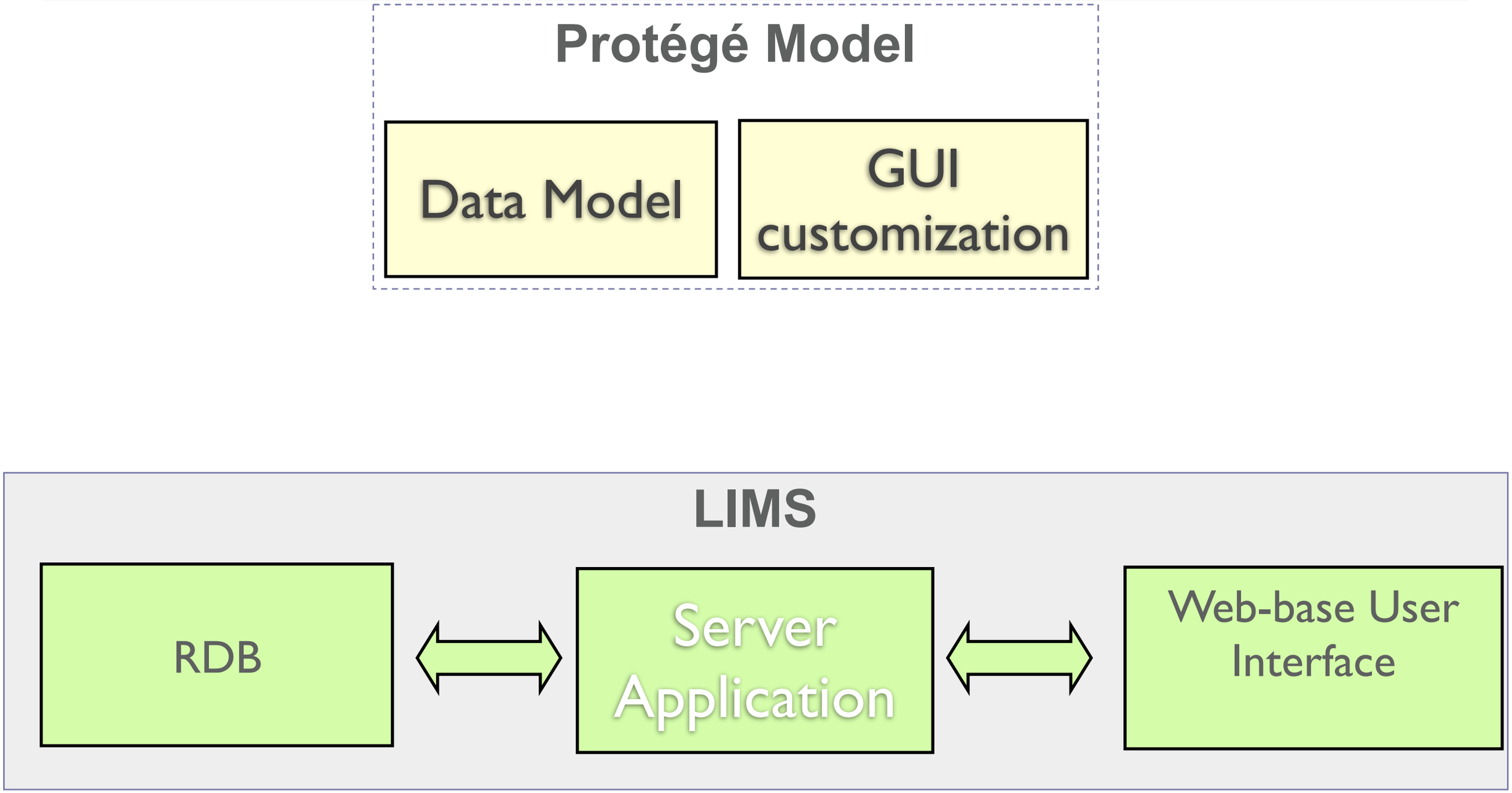




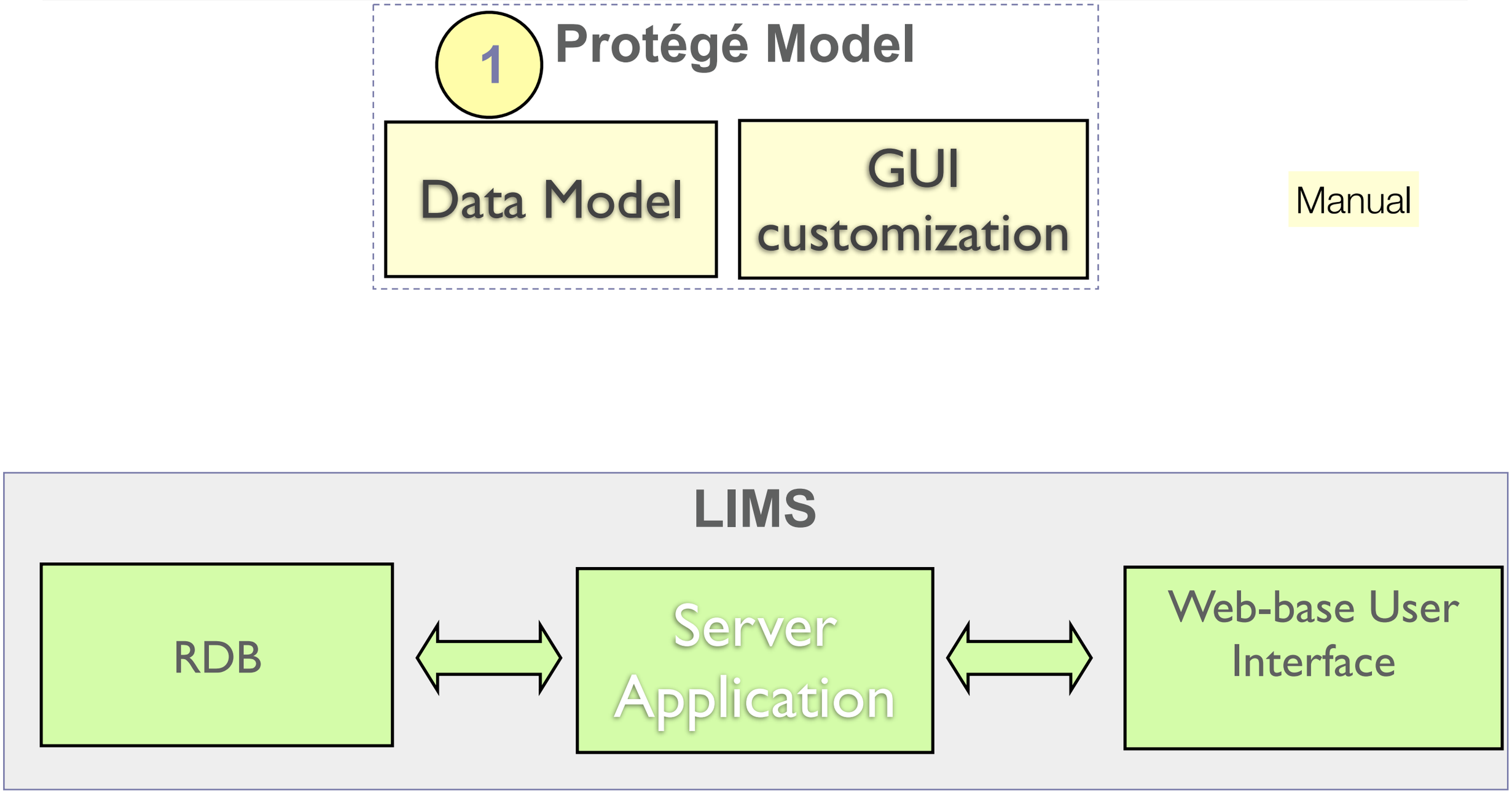
Seedpod Takes Model Driven Approach



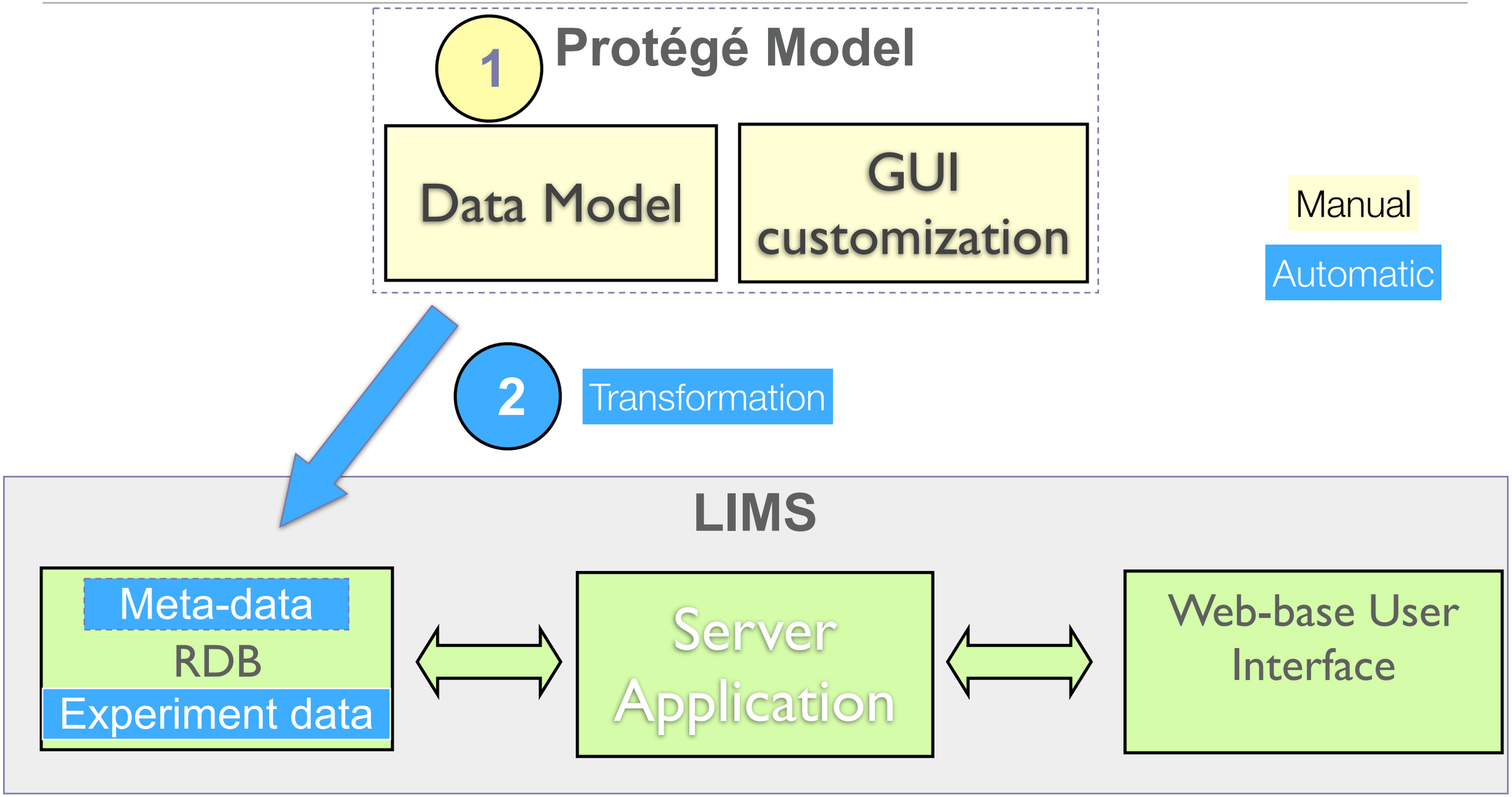
The Components



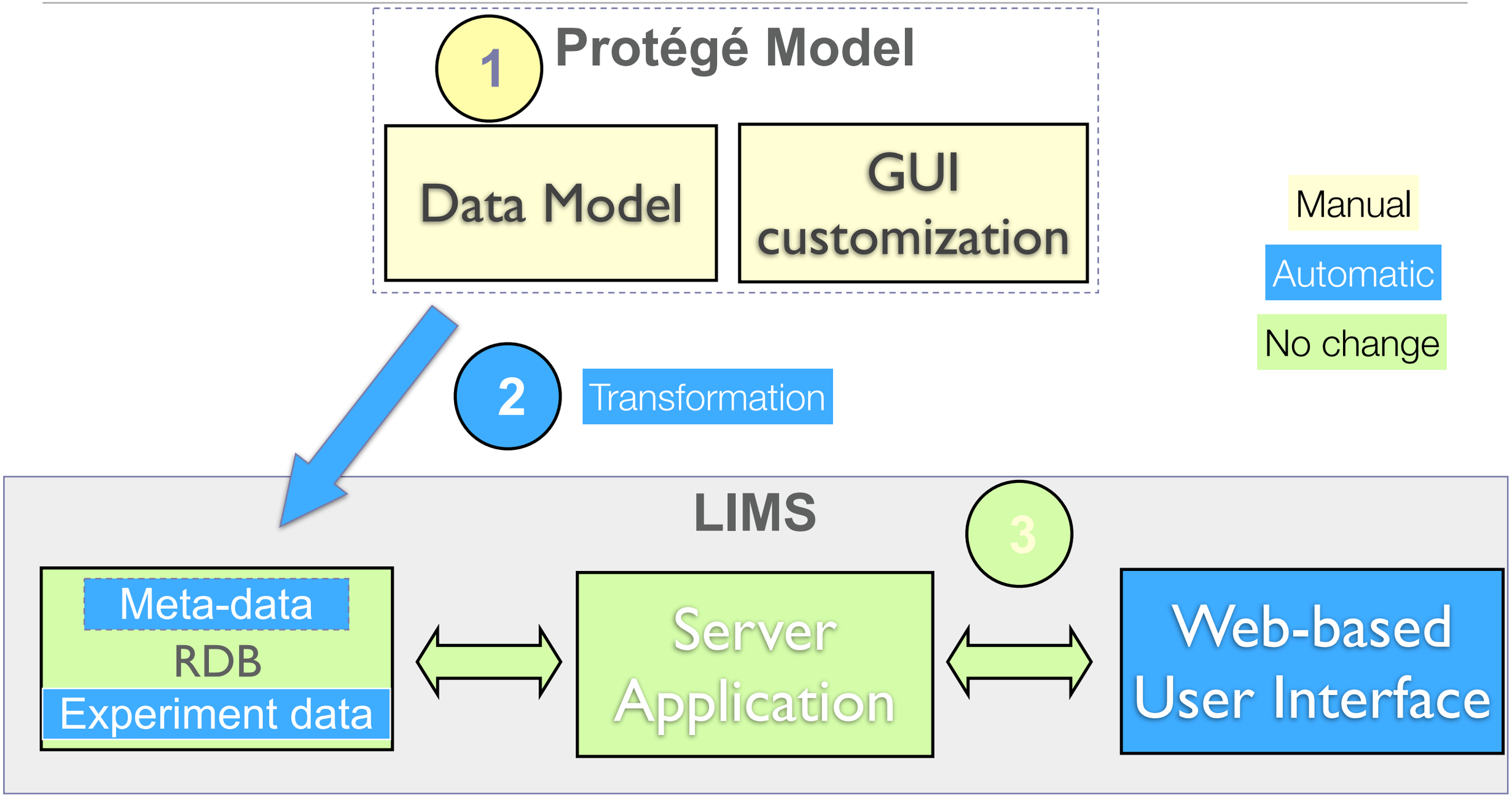
The Components



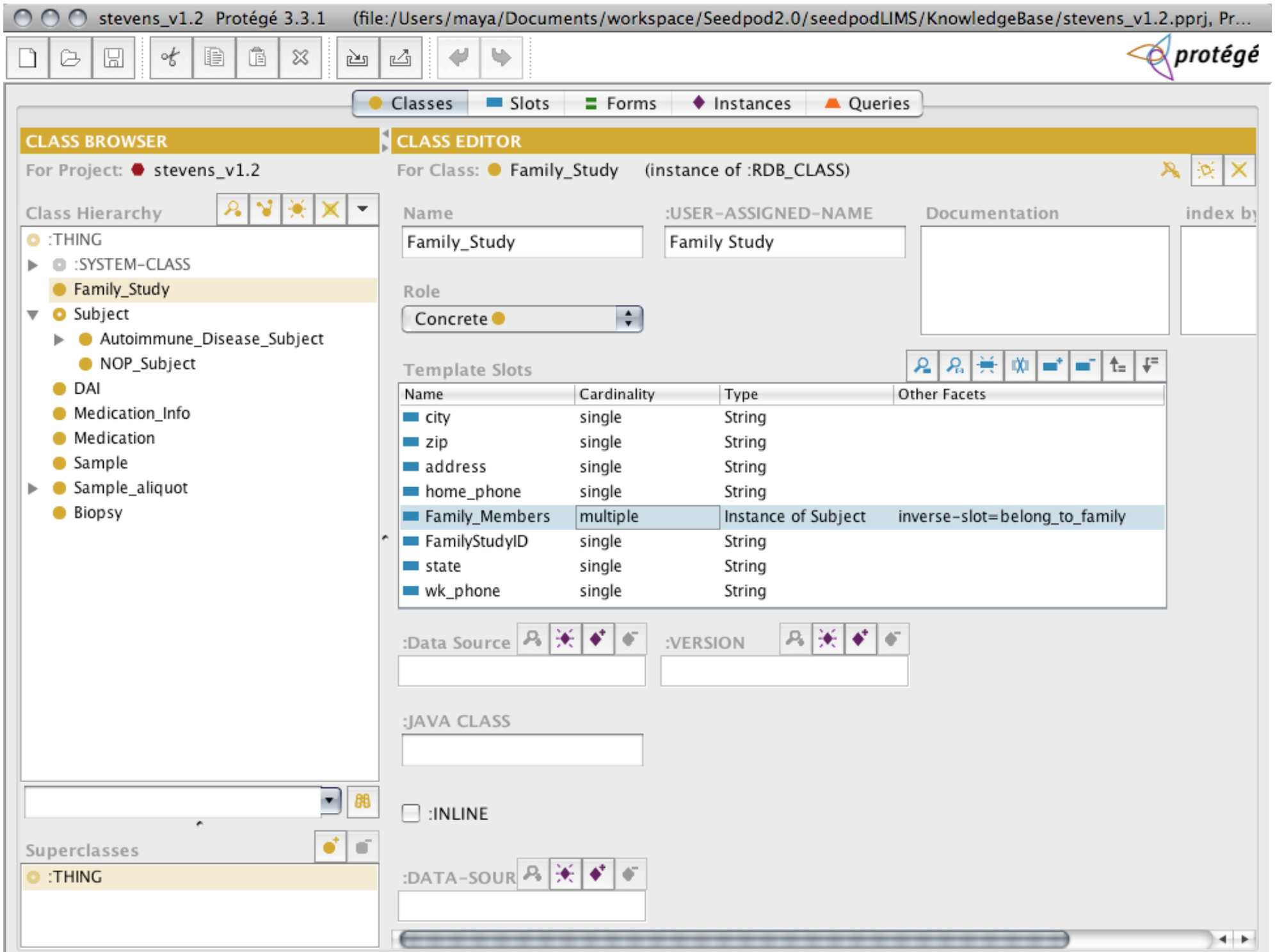
The Components



The Components

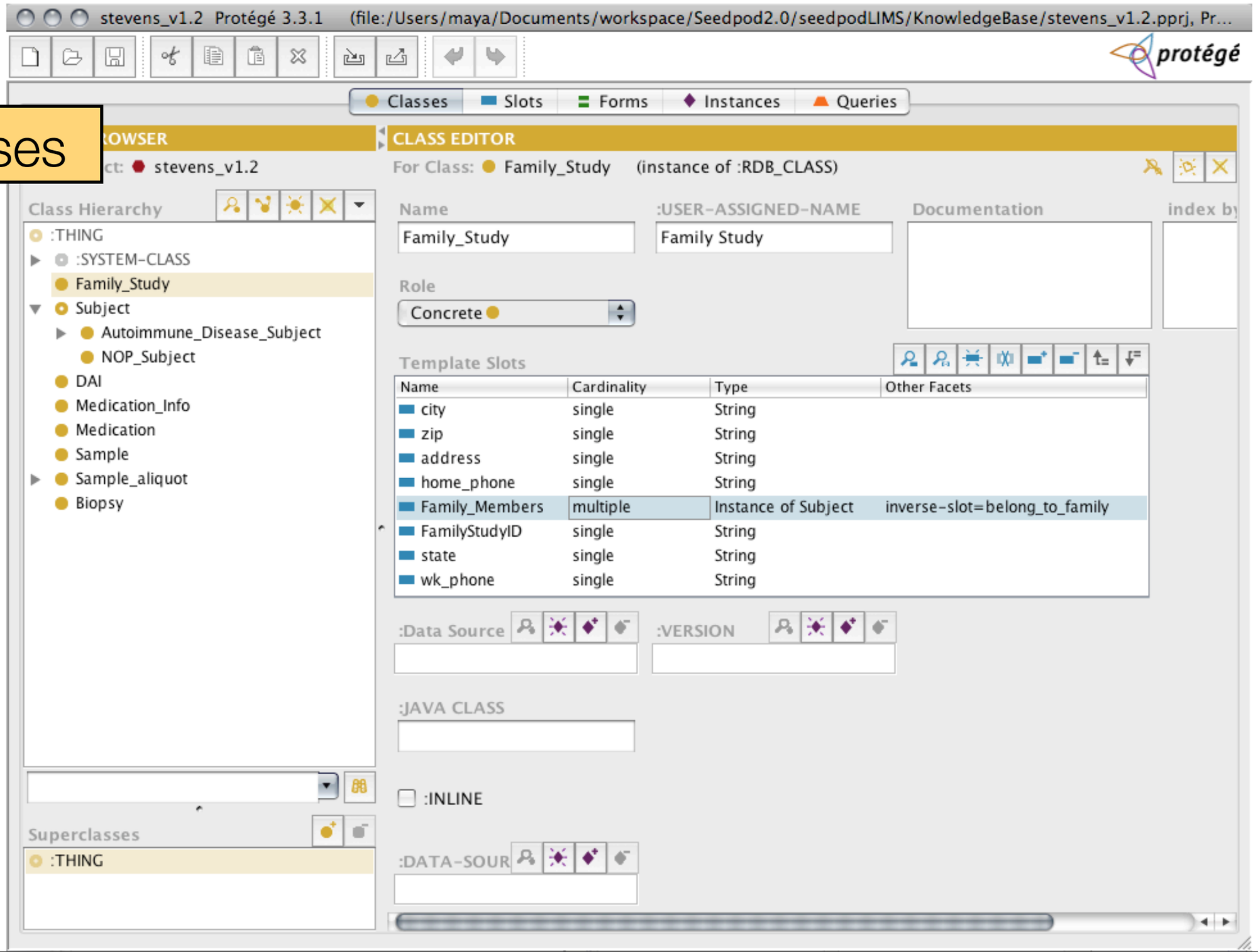


Part 1: Create a Protege model of LIMS

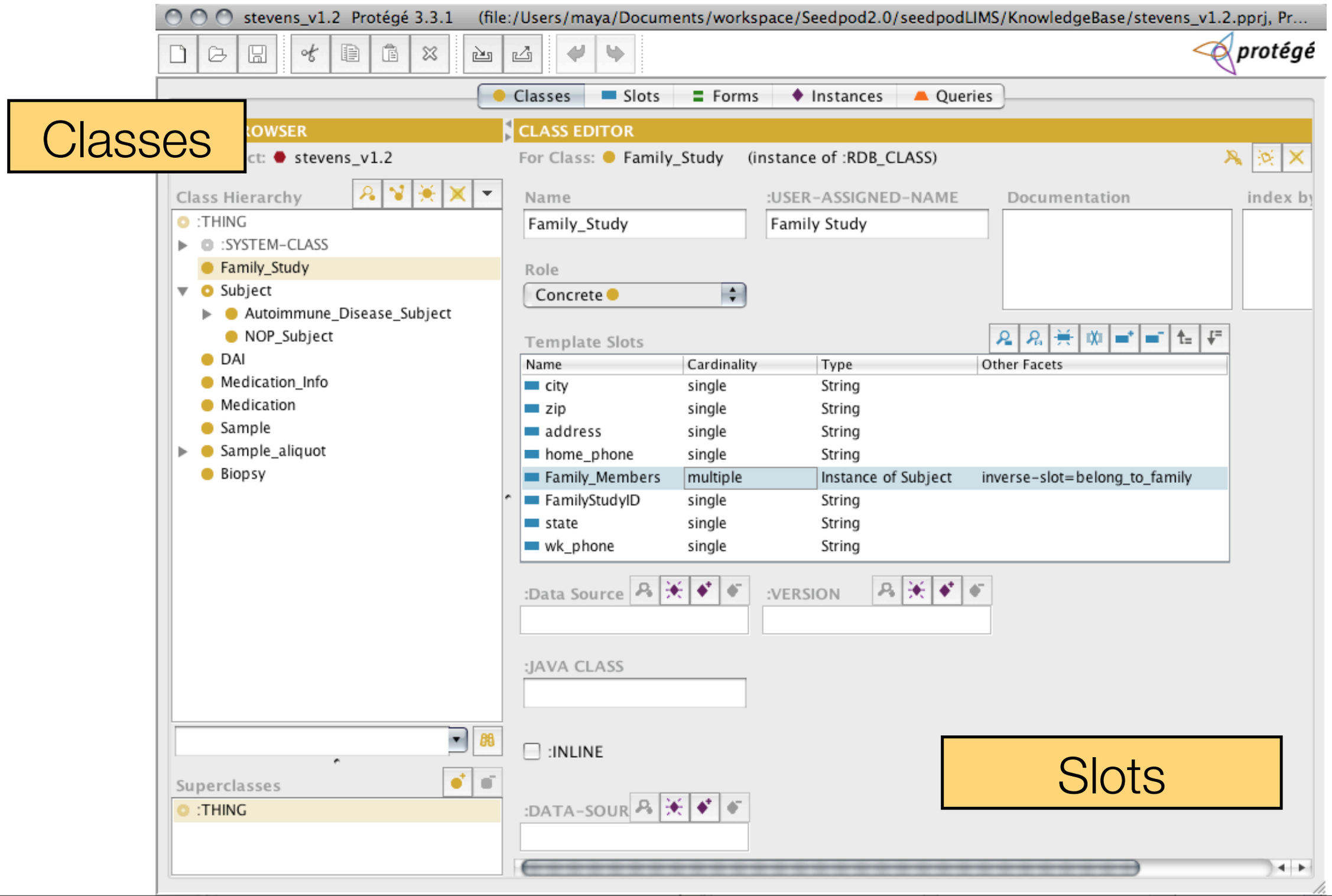


Part 1: Create a Protege model of LIMS

Classes



Part 1: Create a Protege model of LIMS



stevens_v1.2 Protégé 3.3.1 (file:/Users/maya/Documents/workspace/Seedpod2.0/seedpodLIMS/KnowledgeBase/stevens_v1.2.pprj, OWL, RDF)

protégé

Classes | Slots | Forms | Instances | Queries

CLASS BROWSER

For Project: stevens_v1.2

Class Hierarchy

- :THING
  - :SYSTEM-CLASS
  - Family_Study
  - Subject
    - Autoimmune_Disease_Subject
    - NOP_Subject
  - DAI
  - Medication_Info
  - Medication
  - Sample
  - Sample_aliquot
  - Biopsy

Superclasses

- :THING

CLASS EDITOR

For Class: Family_Study (instance of :RDB_CLASS)

Name: Family_Study

:USER-ASSIGNED-NAME: Family Study

Documentation

index by

Role: Concrete

Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	

:Data Source

:VERSION

:JAVA CLASS

:INLINE

:DATA-SOUR

stevens_v1.2 Protégé 3.3.1 (file:/Users/maya/Documents/workspace/Seedpod2.0/seedpodLIMS/KnowledgeBase/stevens_v1.2.pprj, Protégé Files)

protégé

Classes | Slots | Forms | Instances | Queries

CLASS BROWSER

For Project: stevens_v1.2

Class Hierarchy

- :THING
  - :SYSTEM-CLASS
  - Family_Study
  - Subject
    - Autoimmune_Disease_Subject
    - NOP_Subject
  - DAI
  - Medication_Info
  - Medication
  - Sample
  - Sample_aliquot
  - Biopsy

Superclasses

- :THING

CLASS EDITOR

For Class: Family_Study (instance of :RDB_CLASS)

Name: Family_Study

:USER-ASSIGNED-NAME: Family Study

Documentation

index by

Role: Concrete

Template Slots

Name	Cardinality	Type
city	single	String
zip	single	String
address	single	String
home_phone	single	String
Family_Members	multiple	Instan
FamilyStudyID	single	String
state	single	String
wk_phone	single	String

:Data Source

:VERSION

:JAVA CLASS

:INLINE

:DATA-SOUR

Family_Members (instance of :RDB_ATTRIBUTE)

Name: Family_Members

:USER-ASSIGNED-NAME

Documentation

Value Type: Instance

Unique

DB in...

Default Values

Cardinality

required at least

multiple at most

Allowed Classes

- Subject

Minimum

Maximum

SLOT-INVERSE

- belong_to_family

DB type

Database Type Parameter

Domain

- Family_Study

View widget: OBJEC...

View widget parameter

Form widget: OBJEC...

Form widget parameter

Permission

Compound Value Attribute

Auto Increm...

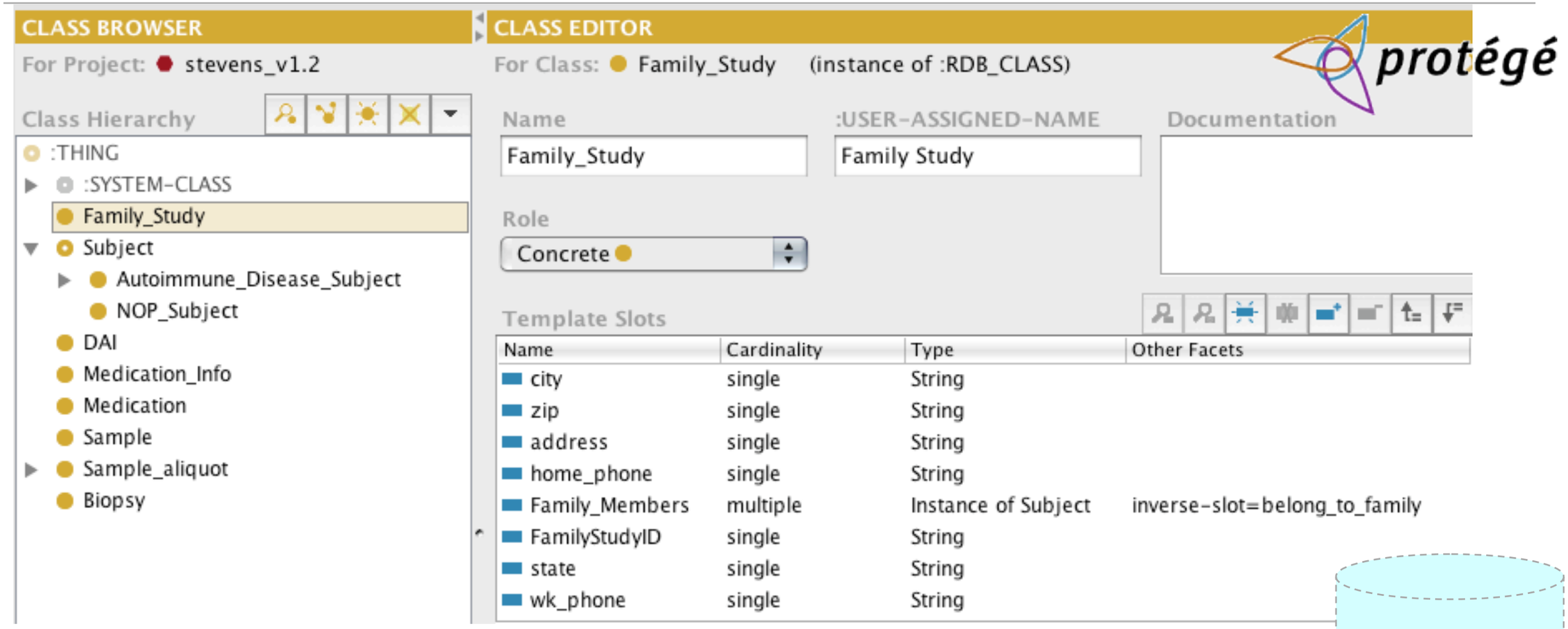
:VALUE-EXPRESSION

:UNIT

Part 2: Automatically Transform Protege Model to Relational Schema

```
CREATE TABLE Family_Study (
    ID INTEGER PRIMARY KEY,
    city  VARCHAR (50),
    zip  VARCHAR (5),
    address VARCHAR(100),
     family_study_ID INTEGER);


CREATE TABLE Subject (
    ID INTEGER PRIMARY KEY,
    dob DATE,
    subjectID INTEGER,
    last_name VARCHAR(50), ...
```

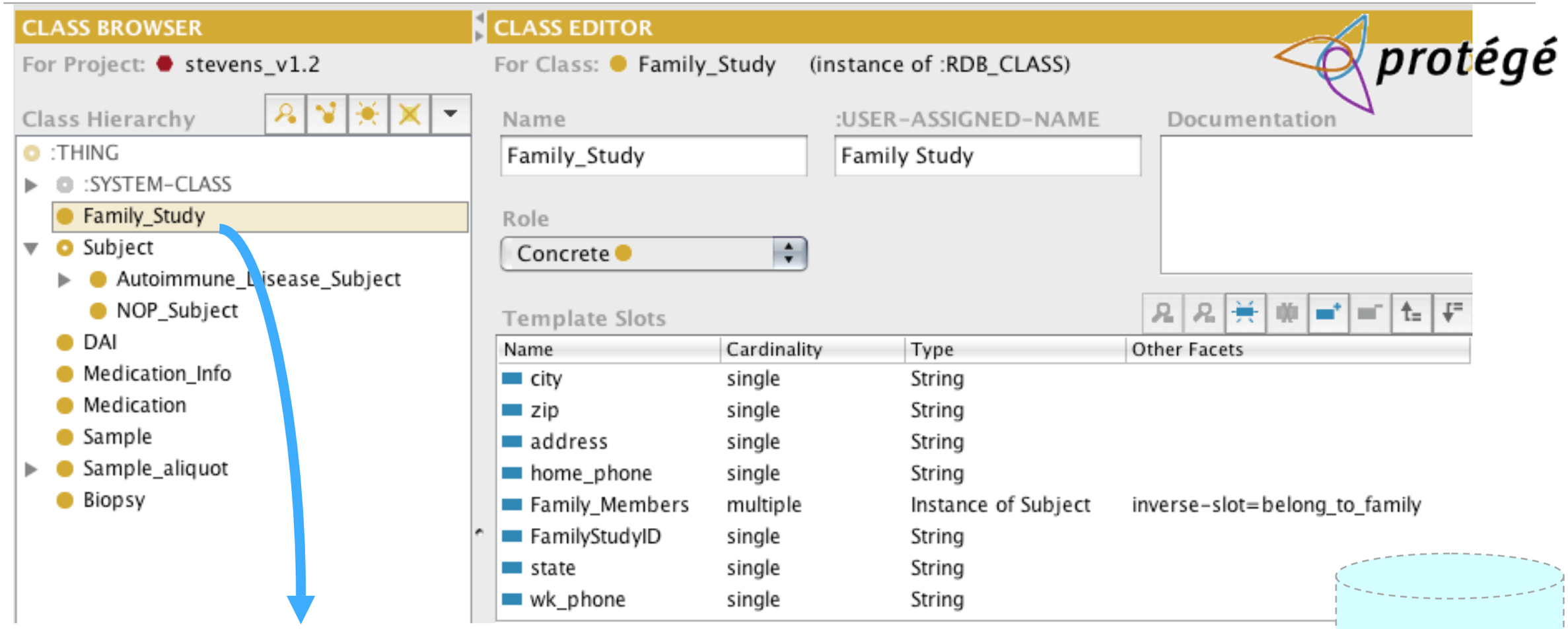
Transform Class ➔ Table




```
CREATE TABLE Neuron (

        …

);
```


Logo courtesy of http://protege.stanford.edu/

Transform Class ➔ Table

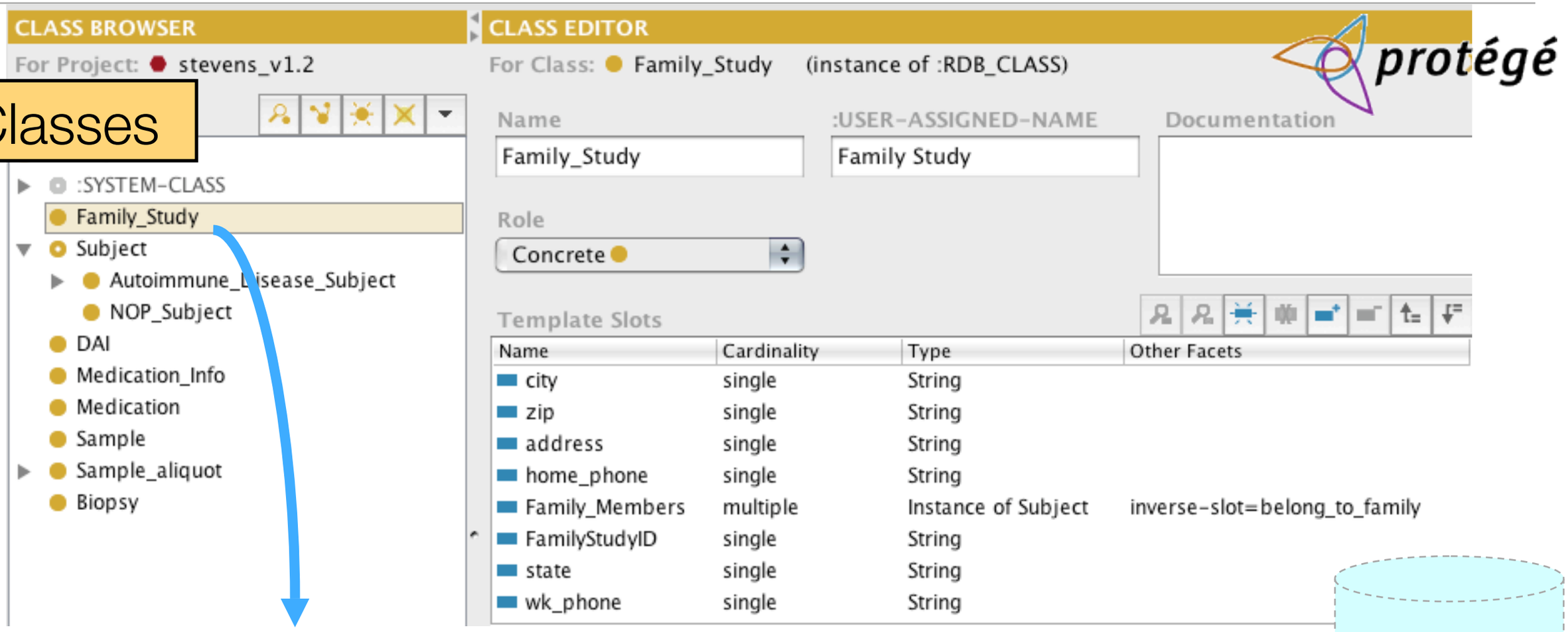



```
CREATE TABLE Neuron (

        ...

);
```


Logo courtesy of http://protege.stanford.edu/

Transform Class ➔ Table

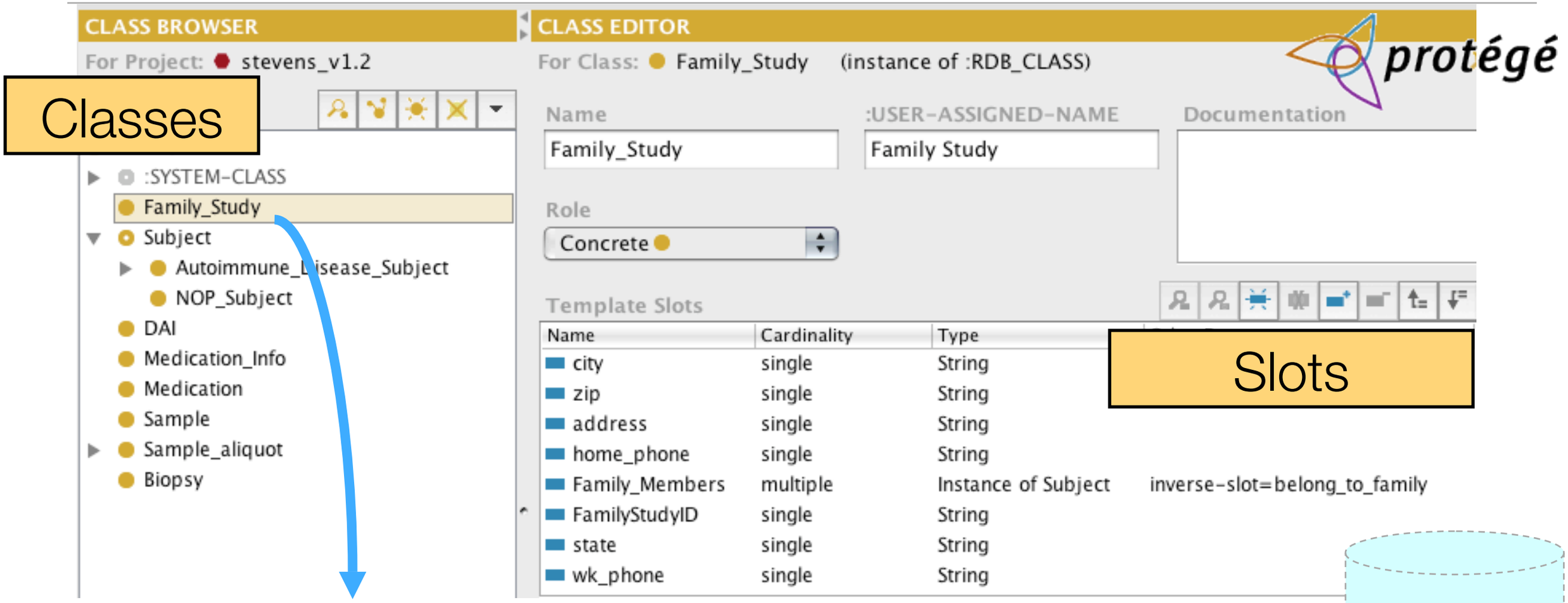
Classes




```
CREATE TABLE Neuron (

        ...

);
```


Logo courtesy of http://protege.stanford.edu/

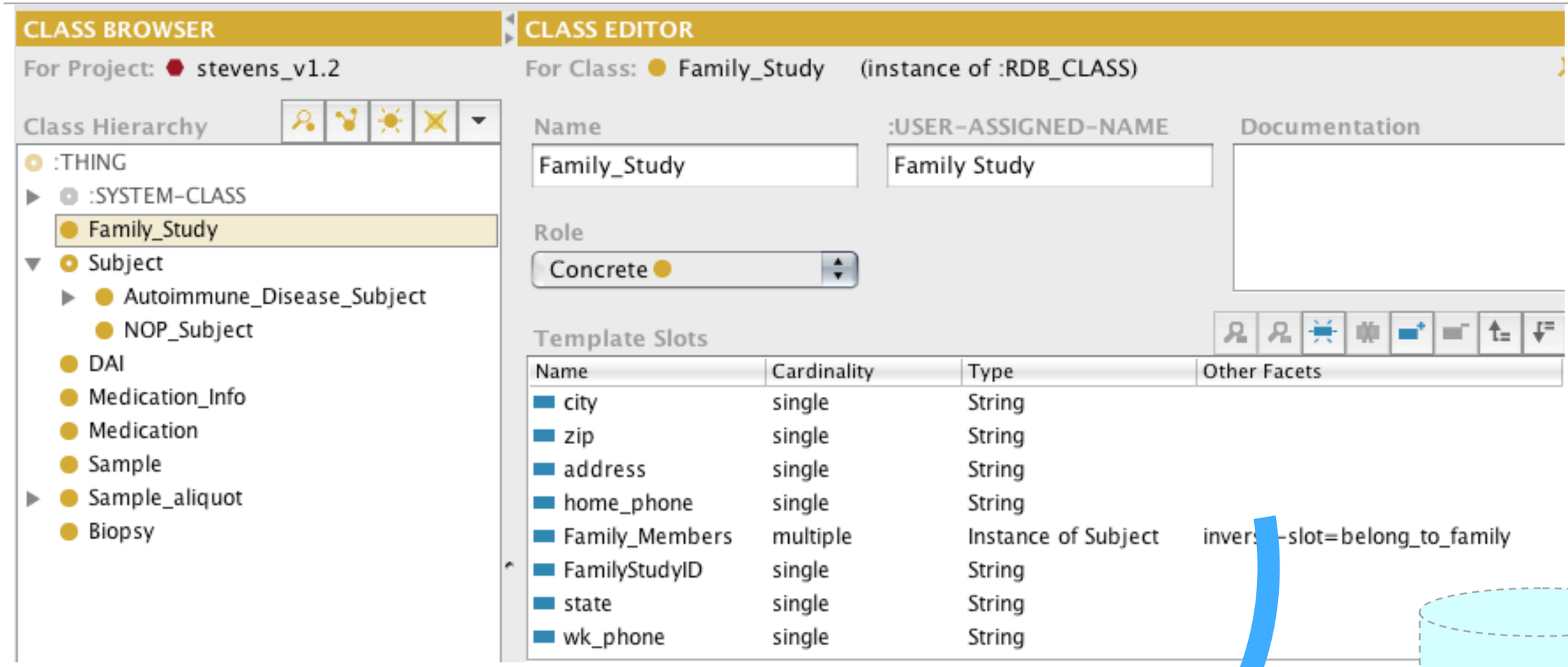
Transform Class ➔ Table




```
CREATE TABLE Neuron (

        ...

);
```


Logo courtesy of http://protege.stanford.edu/

Transform Slot ➔ Attribute, Foreign Key




```
CREATE TABLE Neuron (
    ID INTEGER PRIMARY KEY,
     city  VARCHAR (50),
     zip  VARCHAR (5),
     address VARCHAR(50),
     family_study_ID INTEGER);
```

Attribute, foreign key

Transform Slot ➔ Attribute, Foreign Key

CLASS BROWSER

For Project: stevens_v1.2

Class Hierarchy

- :THING
  - :SYSTEM-CLASS
  - Family_Study
  - Subject
    - Autoimmune_Disease_Subject
    - NOP_Subject
  - DAI
  - Medication_Info
  - Medication
  - Sample
  - Sample_aliquot
  - Biopsy

CLASS EDITOR

For Class: Family_Study (instance of :RDB_CLASS)

Name: Family_Study

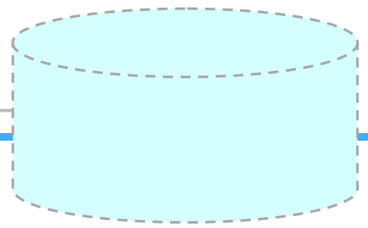
:USER-ASSIGNED-NAME: Family Study

Documentation:

Role: Concrete

Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	



```
CREATE TABLE Neuron (
    ID INTEGER PRIMARY KEY,
    city  VARCHAR (50),
    zip  VARCHAR (5),
    address VARCHAR(50),
    family_study_ID INTEGER);
```

Attribute, foreign key

Transform Slot ➔ Attribute, Foreign Key

Classes

CLASS EDITOR

For Class: Family_Study (instance of :RDB_CLASS)

Class Hierarchy

- :THING
  - :SYSTEM-CLASS
  - Family_Study
  - Subject
    - Autoimmune_Disease_Subject
    - NOP_Subject
  - DAI
  - Medication_Info
  - Medication
  - Sample
  - Sample_aliquot
  - Biopsy

Name: Family_Study

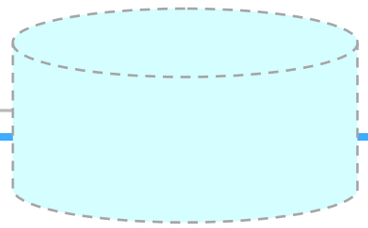
:USER-ASSIGNED-NAME: Family Study

Documentation:

Role: Concrete

Template Slots

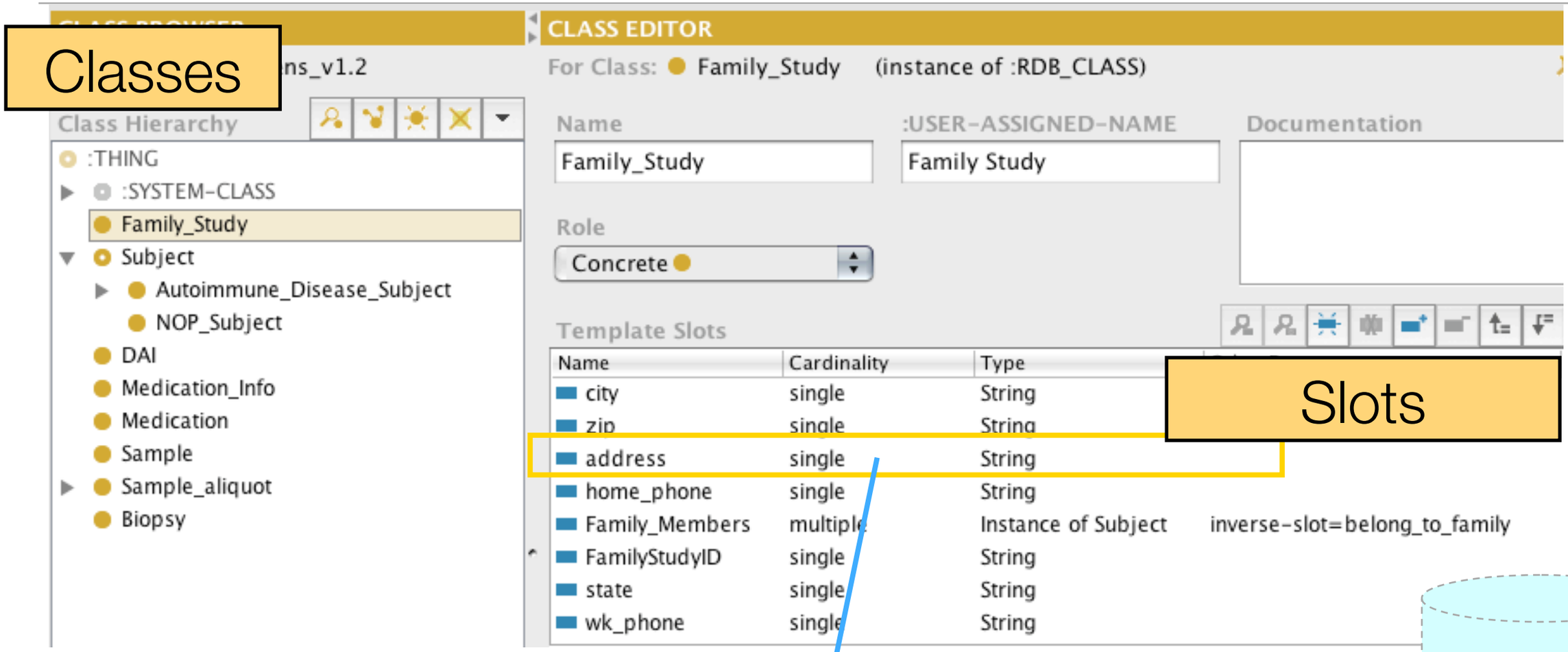
Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	



```
CREATE TABLE Neuron (
    ID INTEGER PRIMARY KEY,
    city  VARCHAR (50),
    zip  VARCHAR (5),
    address VARCHAR(50),
    family_study_ID INTEGER);
```

Attribute, foreign key

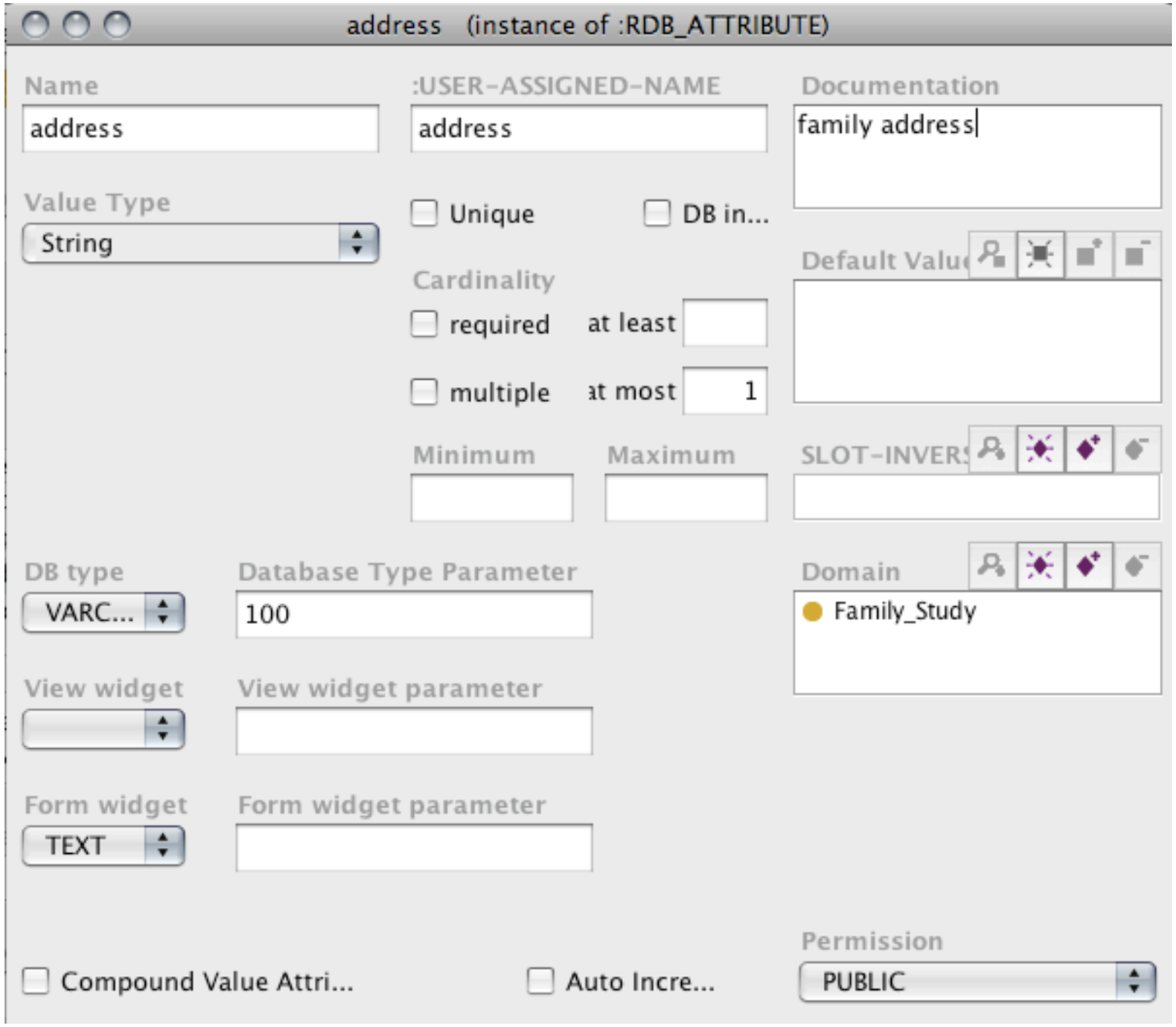
Transform Slot ➔ Attribute, Foreign Key



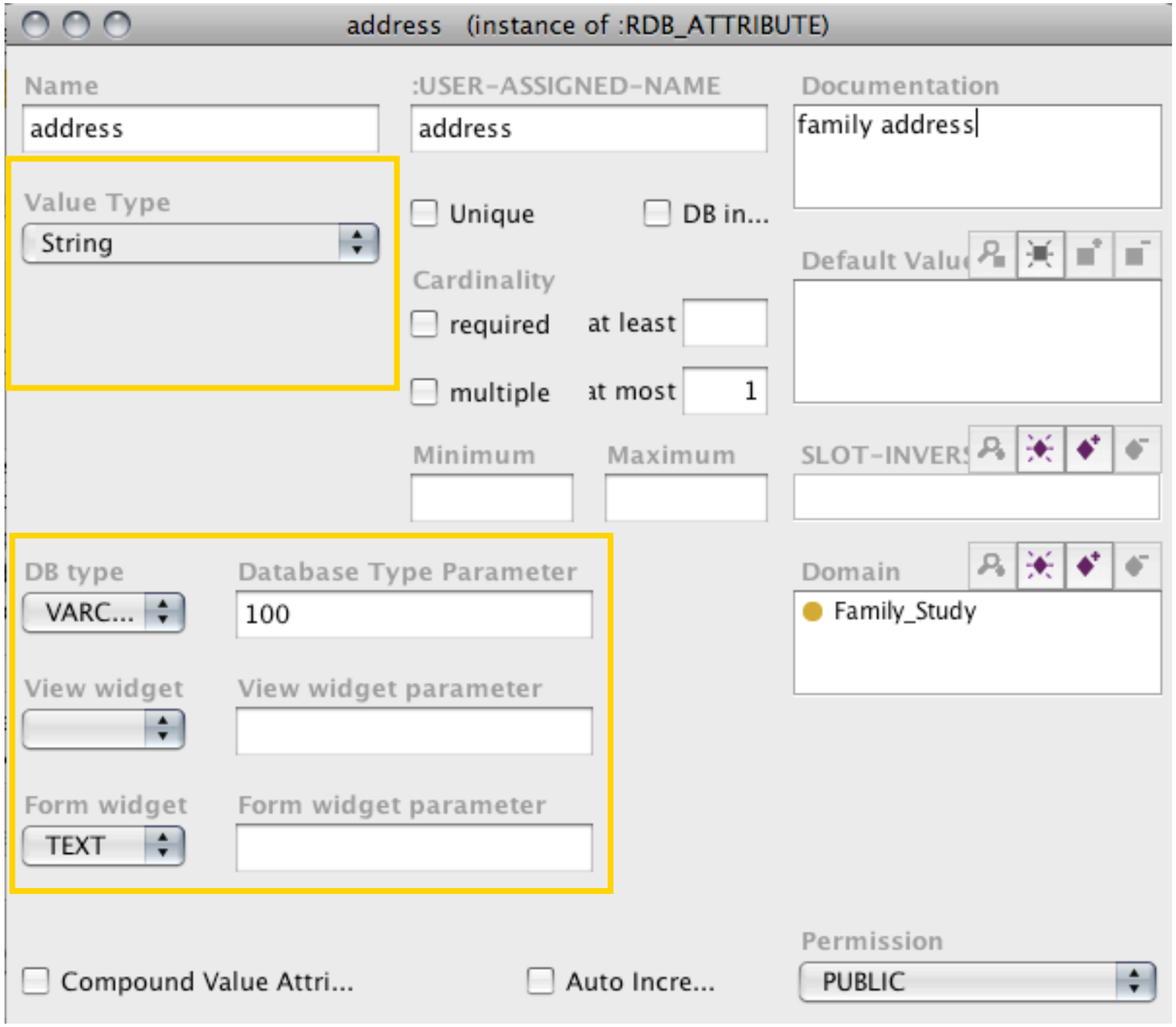

```
CREATE TABLE Neuron (
    ID INTEGER PRIMARY KEY,
    city  VARCHAR (50),
    zip  VARCHAR (5),
    address VARCHAR(50),
    family_study_ID INTEGER);
```

Attribute, foreign key

Extended Slot Definition To Bridge Model Differences Between Protege and Relational Model



Extended Slot Definition To Bridge Model Differences Between Protege and Relational Model

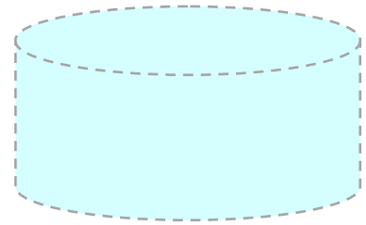


Transforming 1-to-many Relationship

Family_Study Class

Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	



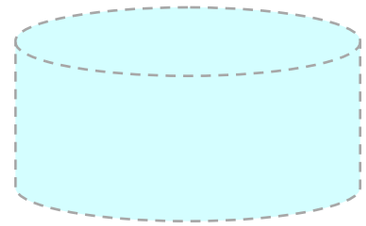
Logo courtesy of http://protege.stanford.edu/

Transforming 1-to-many Relationship

Family_Study Class

Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	



Family_Study

ID

FamilyID

...

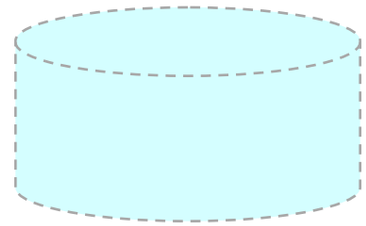
Logo courtesy of http://protege.stanford.edu/

Transforming 1-to-many Relationship

Family_Study Class

Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	



Family_Study

ID

FamilyID

...

Subject

ID

dob

...

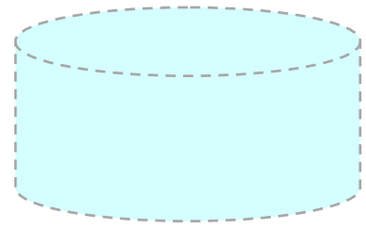
Logo courtesy of http://protege.stanford.edu/

Transforming 1-to-many Relationship

Family_Study Class

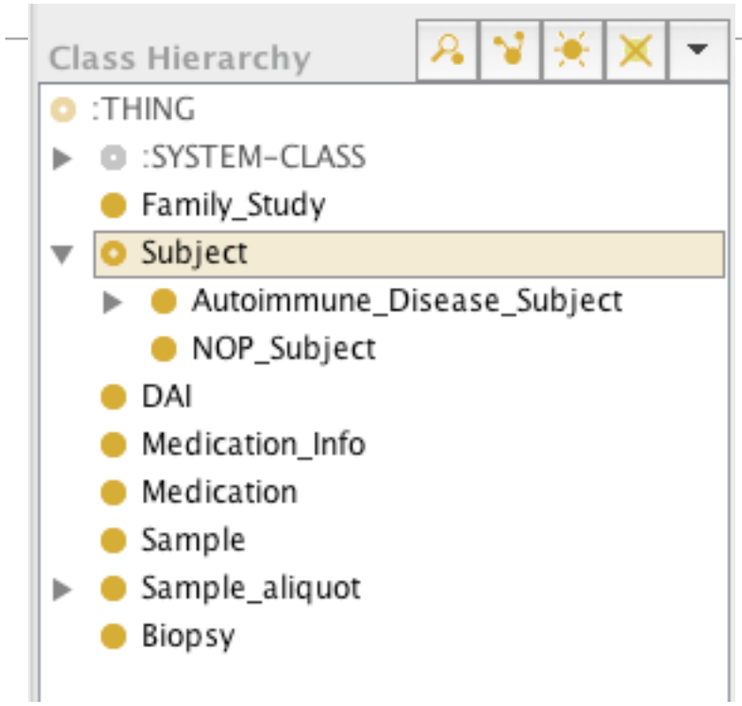
Template Slots

Name	Cardinality	Type	Other Facets
city	single	String	
zip	single	String	
address	single	String	
home_phone	single	String	
Family_Members	multiple	Instance of Subject	inverse-slot=belong_to_family
FamilyStudyID	single	String	
state	single	String	
wk_phone	single	String	

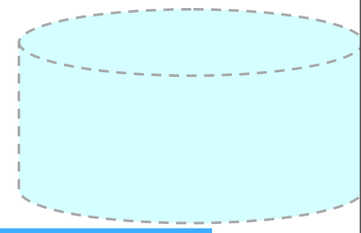


Logo courtesy of http://protege.stanford.edu/

Transforming Hierarchy



protégé

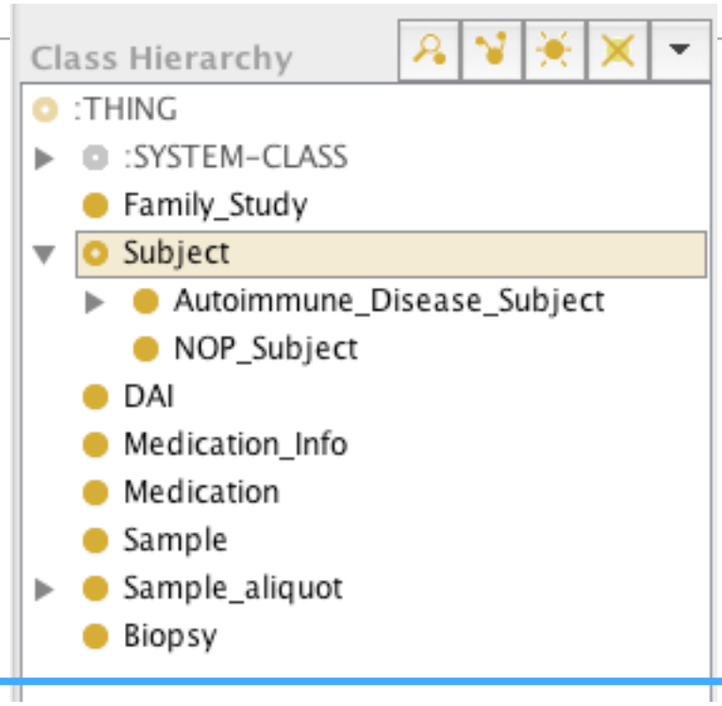


Autoimmune_ Disease_Subject
ID dob* ref_physician

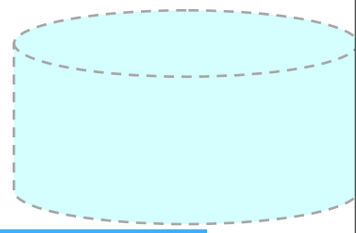
NOP_Subject
ID dob*

Logo courtesy of http://protege.stanford.edu/

Transforming Hierarchy



protégé



Autoimmune_ Disease_Subject

ID
dob*
ref_physician

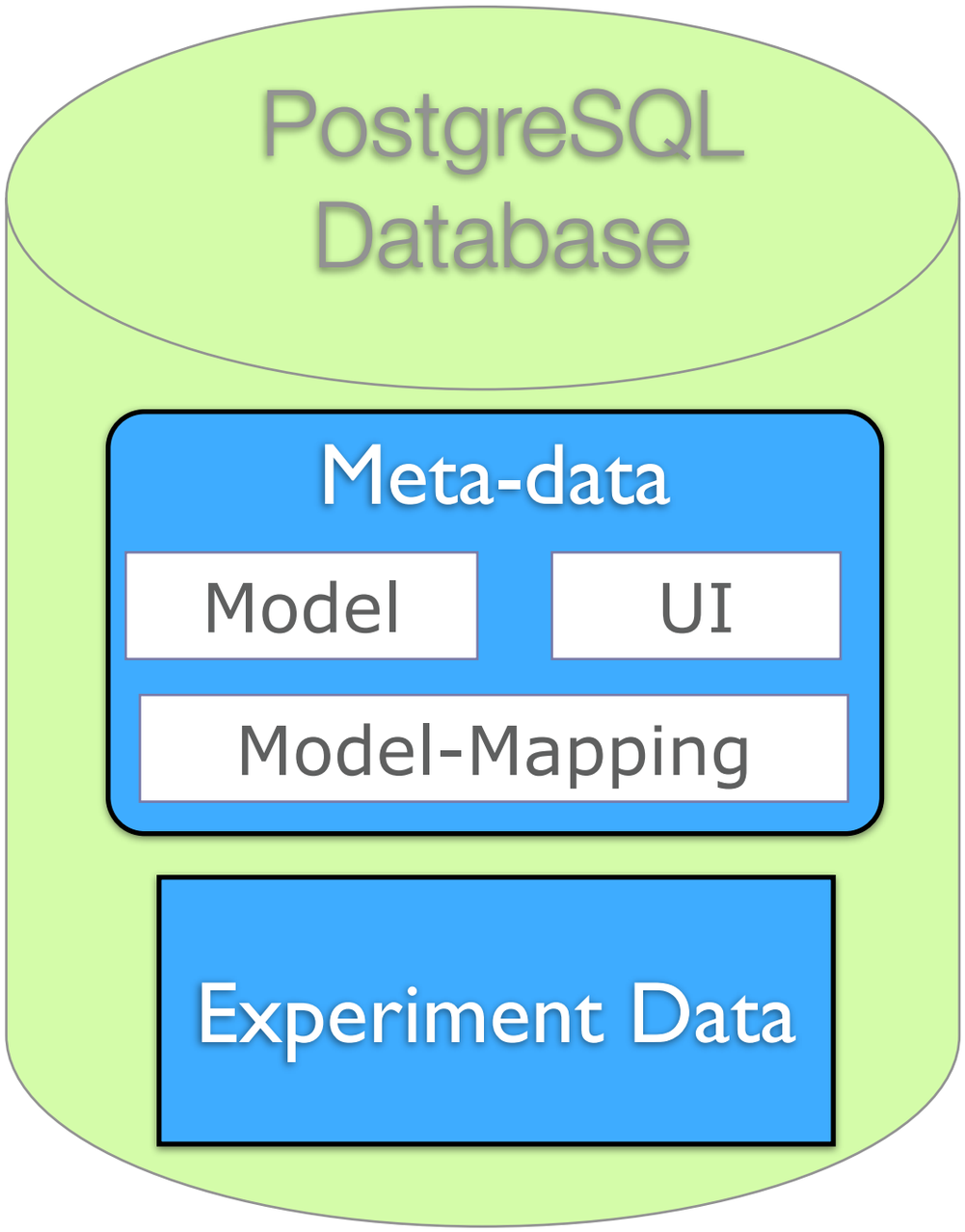
```
CREATE VIEW Subject AS
    Select ID, dob       from
Autoimmune_Disease_Subject
UNION
    Select ID, Modality
    from NOP_Subject;
```

NOP_Subject

ID
dob*

Logo courtesy of http://protege.stanford.edu/

Database



- Meta-data
  - Data model
  - UI customization
  - Model mapping
- Experiment data
  - Schema specific to lab application

Meta-data Tables

RDB_CLASS

Edit Data – Maya's server (localhost:5432) – stevens_v1.2 – :RDB_CLASS

No limit

	cid [PK] integer	frameID integer	name character var	userDefinedN text	clsType character var	parent character var	primaryKey character var	inline boolean	isConcrete boolean	documentatic text	browserPatte character var	tableName character var	vie ch
1	1	11515	Autoimmun		:RDB_CLASS	:REIFIED_SLC	"	FALSE	TRUE	"	Autoimmun	Autoimmun	
2	2	11557	PLE_Subject.		:RDB_CLASS	:REIFIED_SLC	"	FALSE	TRUE	"	PLE_Subject.	PLE_Subject.	
3	3	11225	ANNOTATIO		:XML_CLASS	XML	ID	FALSE	TRUE	"	ANNOTATIO	ANNOTATIO	
4	4	11566	RAY_Subject		:RDB_CLASS	:REIFIED_SLC	"	FALSE	TRUE	"	RAY_Subject	RAY_Subject	
5	5	11448	SPATIAL_AB		:STANDARD	:SPATIAL	:THING	FALSE	TRUE	"	SPATIAL_AB		
6	6	11274	Plasma_sam		:RDB_CLASS	Sample_aliq	ID	FALSE	TRUE	"	Plasma_sam	Plasma_sam	v.F
7	7	11185	Subject		:RDB_CLASS	:THING	:THING	FALSE	FALSE	"	Subject VAL		v.S
8	8	11106	THY_Subject		:RDB_CLASS	Autoimmun	ID	FALSE	TRUE	"	THY_Subject	THY_Subject	v.T
9	9	11262	NOP_Subjec		:RDB_CLASS	Subject	ID	FALSE	TRUE	"	NOP_Subjec	NOP_Subjec	v.N
10	10	11533	EBY_sample.		:RDB_CLASS	:REIFIED_SLC	"	FALSE	TRUE	"	EBY_sample.	EBY_sample.	
11	11	11536	GRANS_sam		:RDB_CLASS	:REIFIED_SLC	"	FALSE	TRUE	"	GRANS_sam	GRANS_sam	
12	12	11158	Biopsy		:RDB_CLASS	:THING	ID	FALSE	TRUE	"	Biopsy VAL(i	Biopsy	v.B

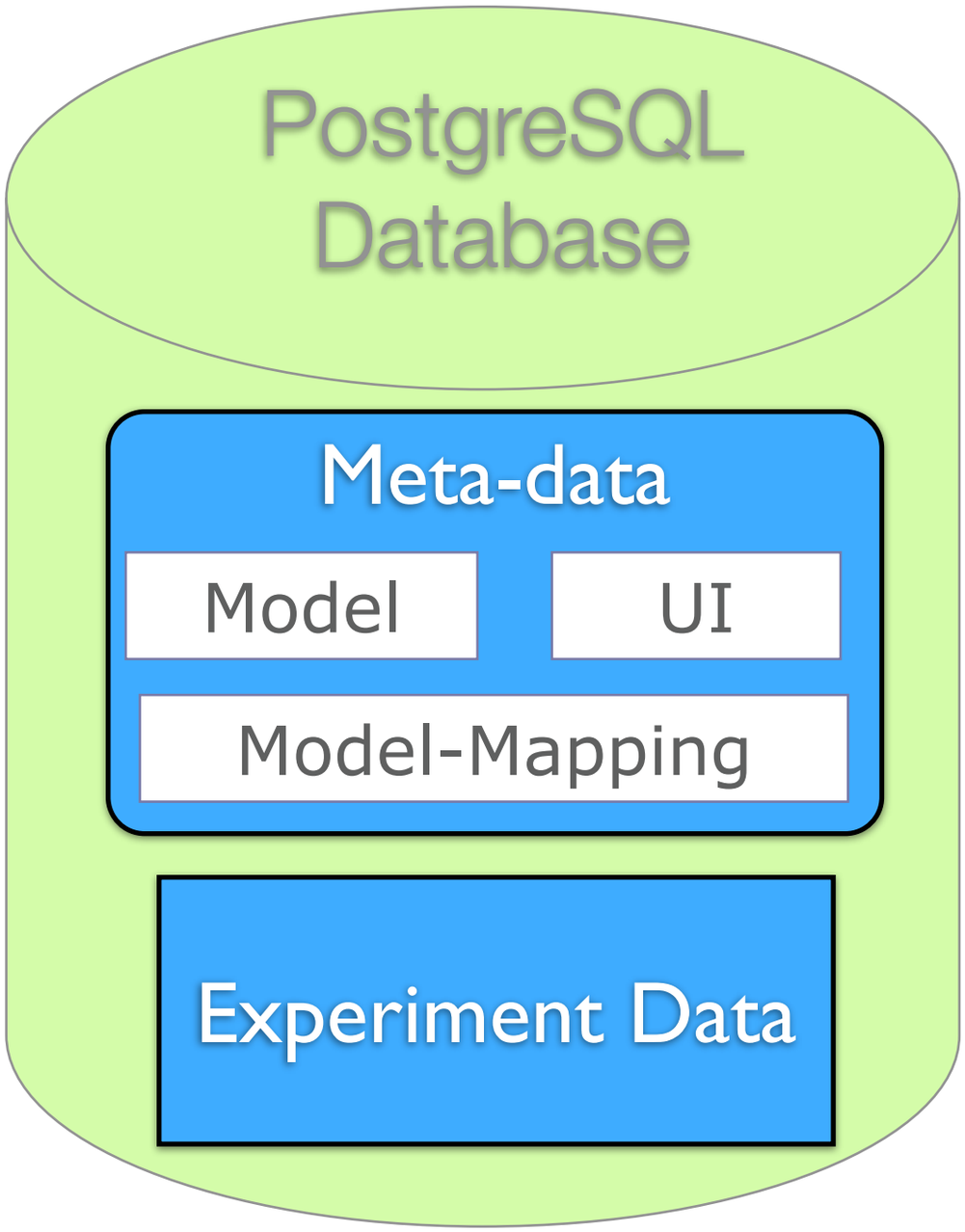
RDB_ATTRIBUTE

Edit Data – Maya's server (localhost:5432) – stevens_v1.2 – :RDB_ATTRIBUTE

No limit

aid [PK] integer	frameID integer	domainCls character var	name character var	userDefinedN text	slotType character var	protegeValue character var	defaultValues text	allowedCls character var	slotInverse character var	numericMin numeric	numericMax numeric	cardinali integer
38	11114	Family_Stud	zip	"	:RDB_ATTRI	String	"	"	"			0
64	11197	Family_Stud	state	"	:RDB_ATTRI	String	"	"	"			0
91	11109	Family_Stud	FamilyStudy	FS ID	:RDB_ATTRI	String	"	"	"			0
135	11097	Family_Stud	address	"	:RDB_ATTRI	String	"	"	"			0
269	11206	Family_Stud	home_phon	hm phone #	:RDB_ATTRI	String	"	"	"			0
270	11207	Family_Stud	Family_Mem	"	:RDB_ATTRI	Instance	"	Subject	belong_to_f			0
342	11291	Family_Stud	wk_phone	wk phone#	:RDB_ATTRI	String	"	"	"			0

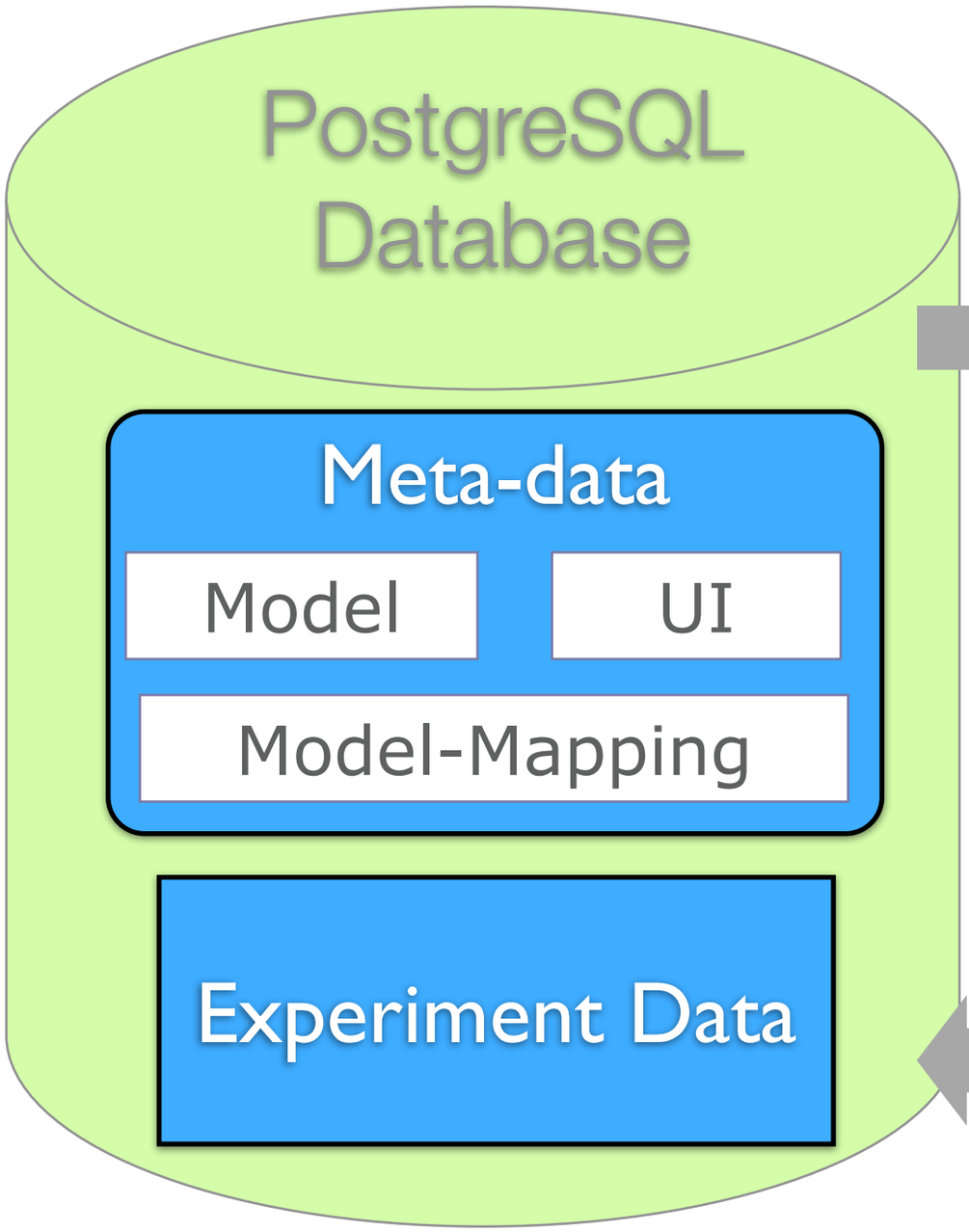
Part 3: Set-up Web Server Application



Server Application

- Config database connectivity
- Load meta-data
- Interpret model (no hard-coded data objects)
- Dynamically generate web pages for browsing experiment data and data entry

Part 3: Set-up Web Server Application



Server Application

- Config database connectivity
- Load meta-data
- Interpret model (no hard-coded data objects)
- Dynamically generate web pages for browsing experiment data and data entry

Browse Data on the Web

EMS **Database: sur_v3**

EMS Home | Logged in as Hao Li (ADMINISTRATOR) | Logout | Help

Trial Protocol (Parent Class = Set)

Create New Instance
View Class Definition
Search

Instance

Block Naming 1

Protocol Name	Block Naming 1
Protocol Description	Name targets in sequence
patient response timestamp	(Class= File)
stimulus onset timestamp	(Class= File)
CD Archive	abc
electrodes	(Class= Electrode) Electrode 2 Electrode 3
trials	(Class= … id | index | presentati… descriptor 1 | target 2 | distractor 3 | target

EMS **Database: sur_v3**

EMS Home | Logged in as Hao Li (ADMINISTRATOR) | Logout | Help

Electrode (Parent Class = :THING)

Create New Instance
View Class Definition
Search

Instance

Electrode 3

Referenced by electrodes of Trial Protocol	(Class= Trial Protocol) Block Naming 1
electrode number	3
electrode site number	2
electrode site name	
Distance From Temporal Tip	4 mm
Electrode Depth	3.5
neurons	(Class= Neuron)

id	neuron number	Neuron Firing File	neuron width	neuron amplitude
1			3	20
2				3.5
3				

CLASS BROWSER

For Project: sur_v2.1

Class Hie

- :THING
- :SYSTEM-CLASS
- Patient
- Surgery
- Neuron
- Ordered_Set
  - Trial_Protocol
- Ordered_Set_Element
  - Trial
- File
- Single_Unit_Recording
- Electrode
- fMRI

CLASS EDITOR

For Class: Trial_Protocol (instance of :ORDB-CLASS)

Name: Trial_Protocol

User Assigned Name: Trial Protocol

Role: Concrete

Template Slots

Name	Cardinality	Type
electrodes	multiple	Instance of Electrode
elements	multiple	Instance of Trial
Protocol_Description	single	String
Protocol_Name	singl	
stimulus_onset_timest...	singl	
unit_firing_timestamp	singl	

electrodes (instance of :ORDB-SLOT)

Name: electrodes

User Assigned Name: Electrodes

Database Type:

Documentation: In each trial protocol, there are multiple electrodes recording at the same time.

Slot Sequence I: Protocol_Description

Database Type Parameter:

Cardinality

- required — at least
- multiple — at most

View Widget: OBJECT-LINK

Unique

Value Type: Instance

View Widget Parameter:

Maximum:

Minimum:

Allowed Classes: Electrode

Form Widget: COMBO

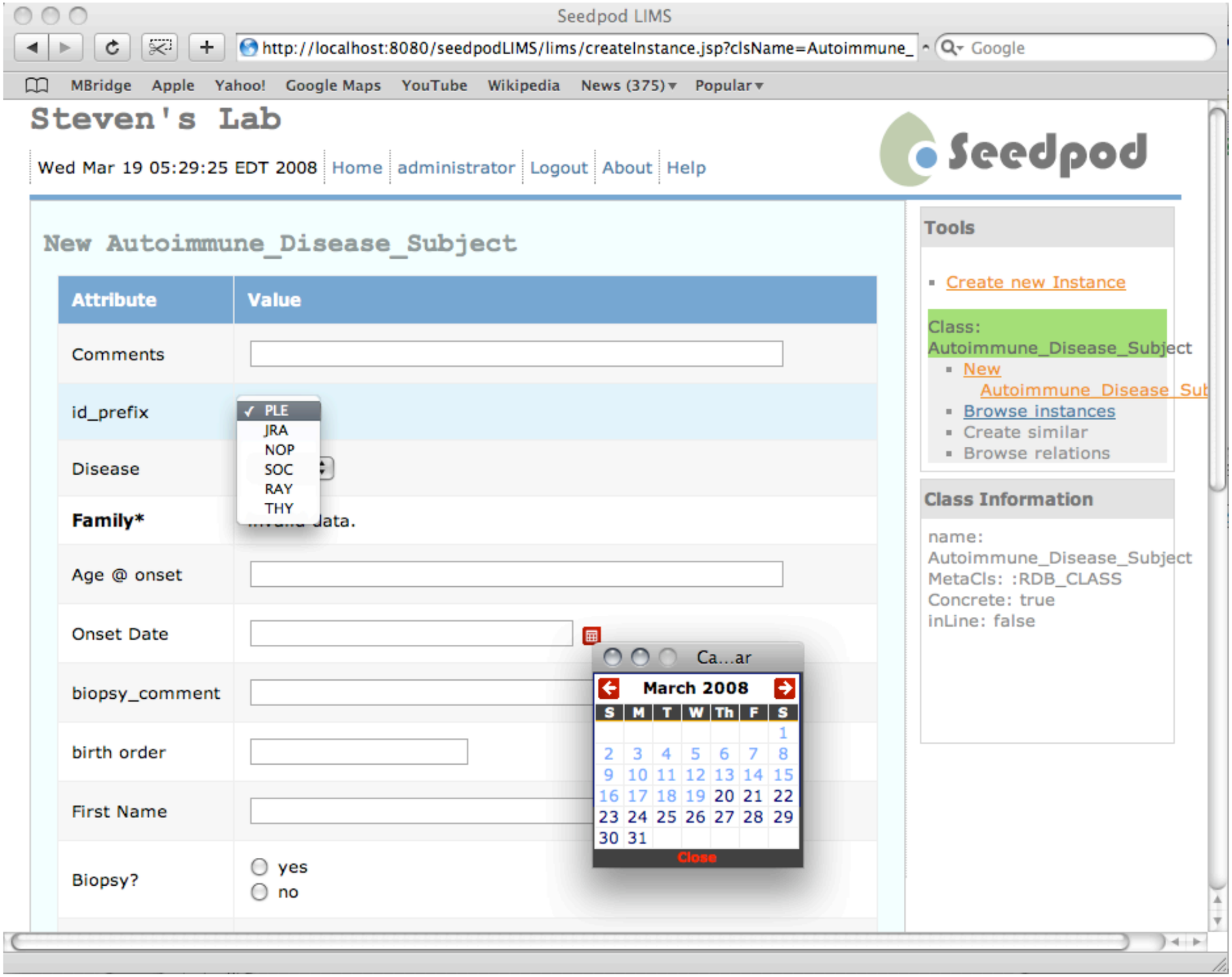
Inverse Slot:

Default Values:

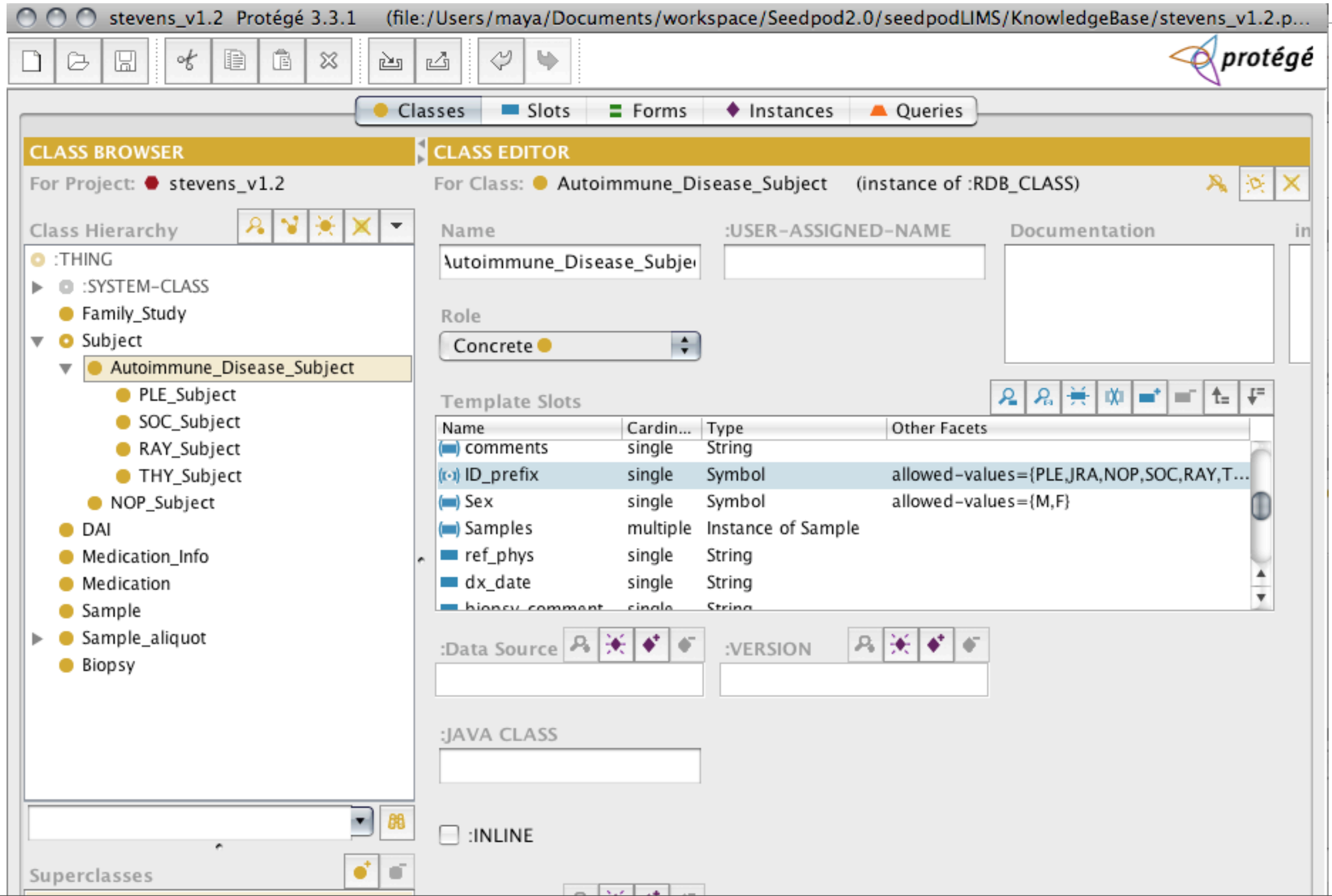
Template Value:

- TEXT
- RADIO
- SELECT
- DATE
- TEXTAREA
- COMBO
- CHECKBOX
- NUMERIC

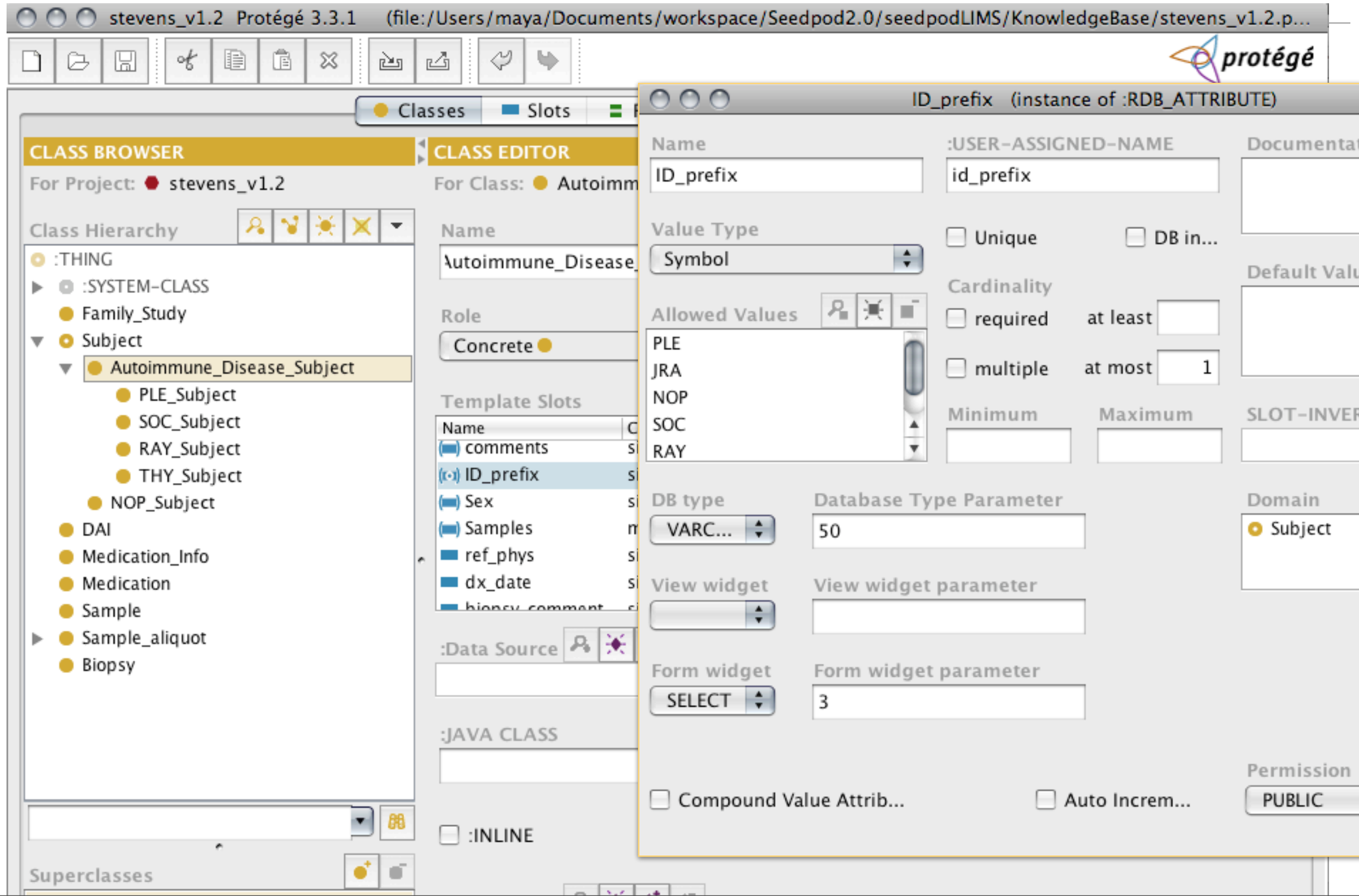
Data Entry


Seedpod LIMS
http://localhost:8080/seedpodLIMS/lims/createInstance.jsp?clsName=Autoimmune_
Google
MBridge
Apple
Yahoo!
Google Maps
YouTube
Wikipedia
News (375)
Popular
Steven's Lab
Wed Mar 19 05:29:25 EDT 2008
Home
administrator
Logout
About
Help
Seedpod
New Autoimmune_Disease_Subject
Attribute
Value
Comments
id_prefix
PLE
JRA
NOP
SOC
RAY
THY
Disease
Family*
Age @ onset
Onset Date
biopsy_comment
birth order
First Name
Biopsy?
yes
no
March 2008
S M T W Th F S
Close
Tools
Create new Instance
Class: Autoimmune_Disease_Subject
New Autoimmune_Disease_Sub
Browse instances
Create similar
Browse relations
Class Information
name: Autoimmune_Disease_Subject
MetaCls: :RDB_CLASS
Concrete: true
inLine: false

Customize GUI Components



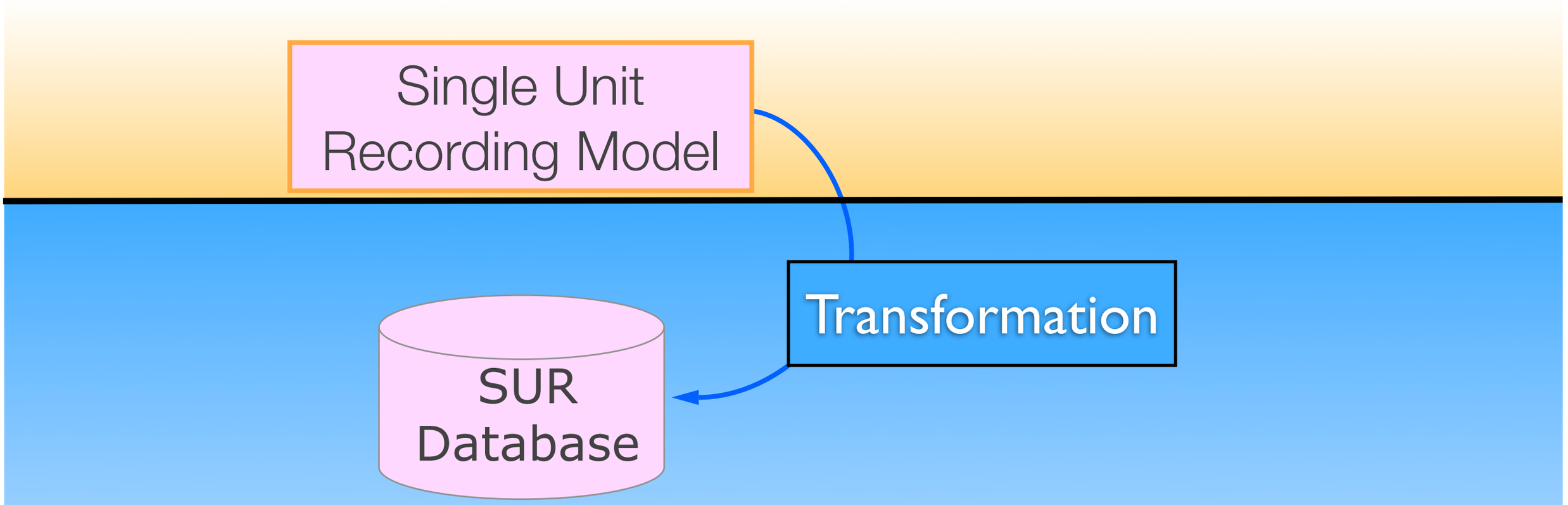
Customize GUI Components



Model Independent Application

Single Unit
Recording Model

Model Independent Application



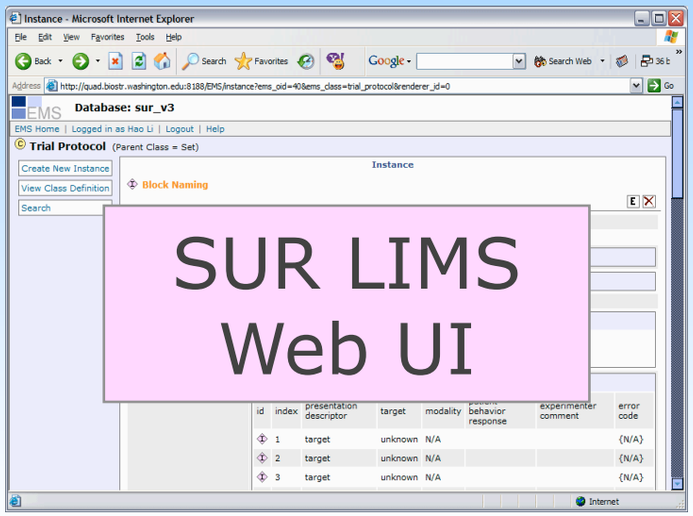
Model Independent Application

Single Unit Recording Model

Transformation

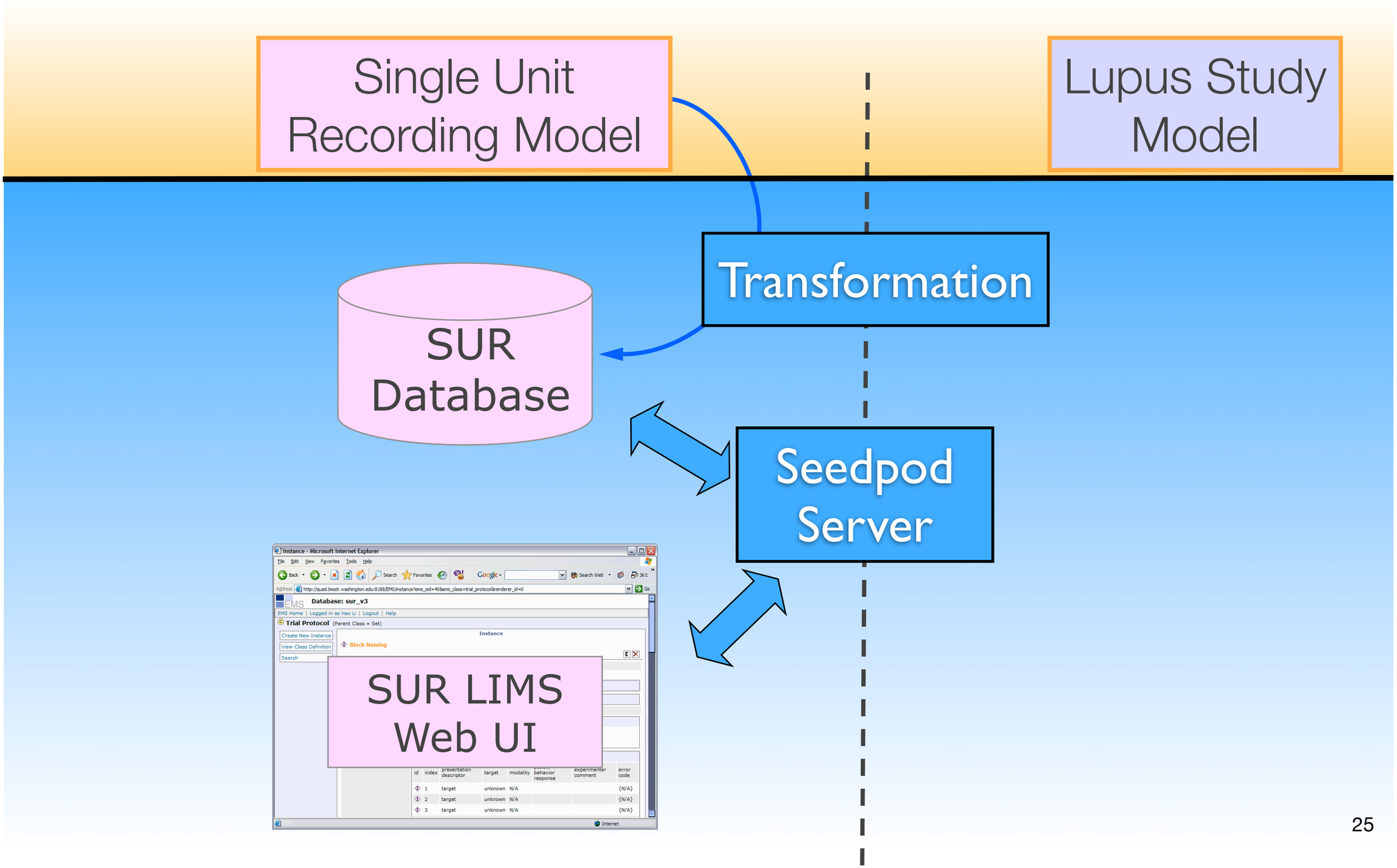
SUR Database

Seedpod Server

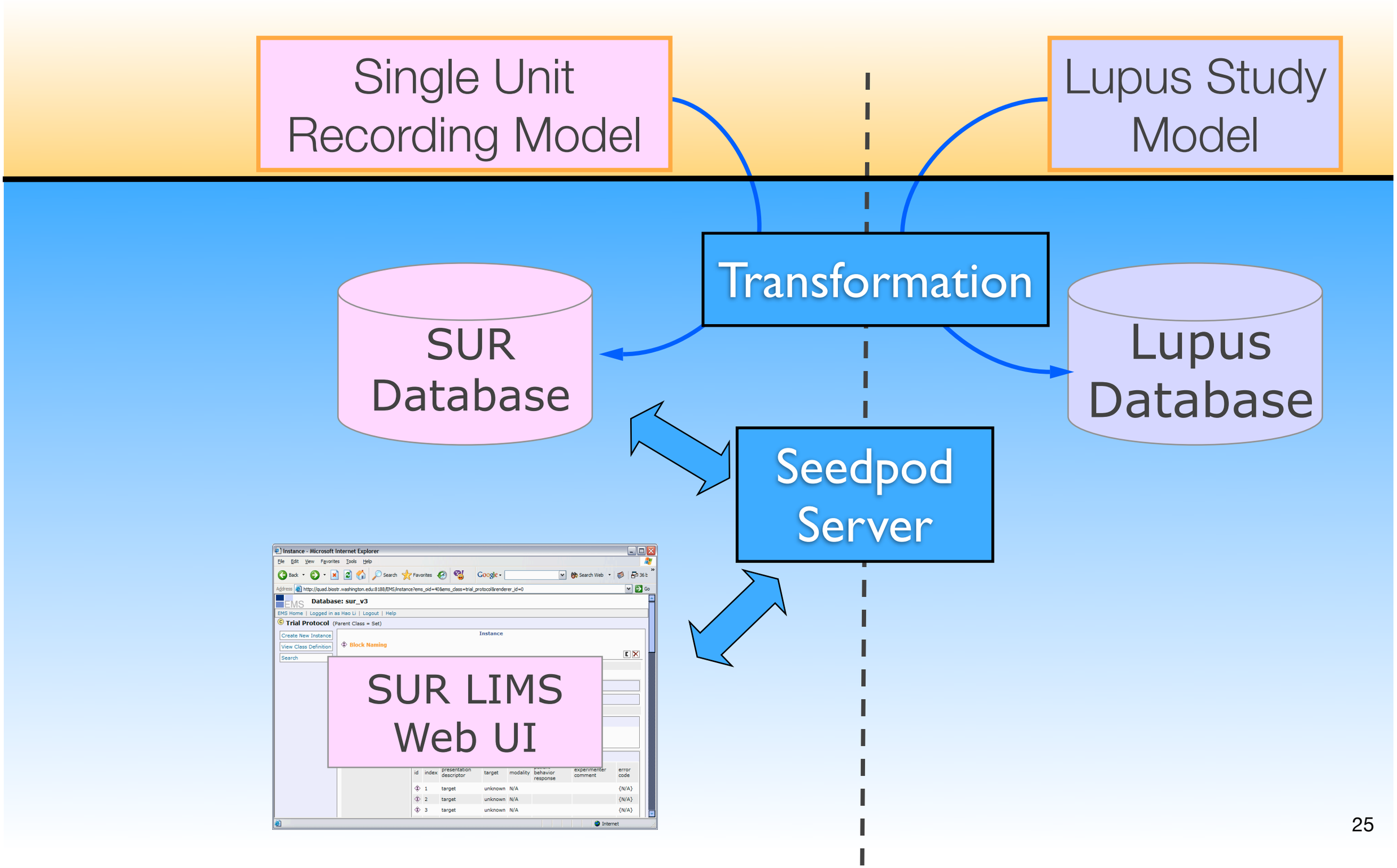


SUR LIMS Web UI

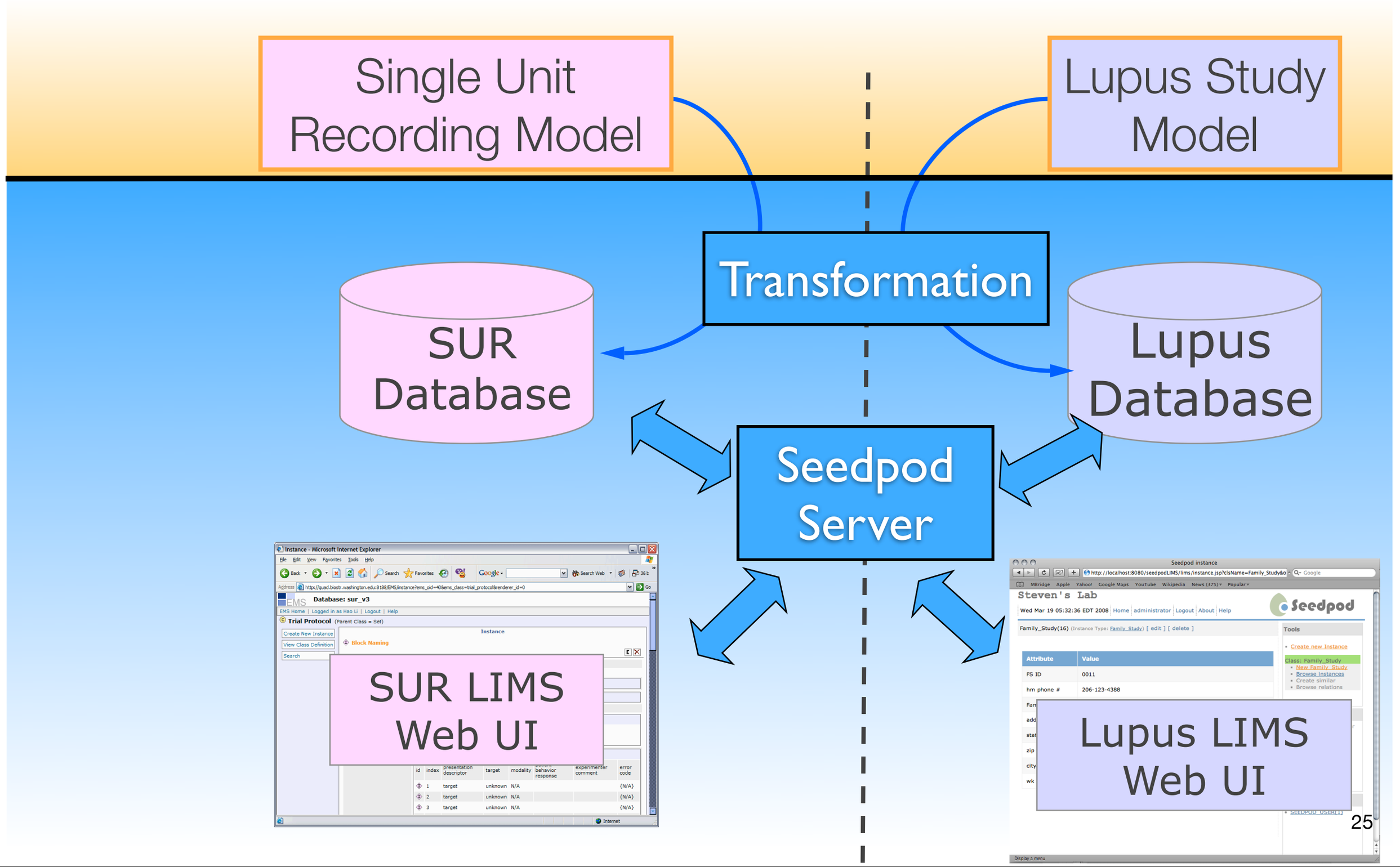
Model Independent Application



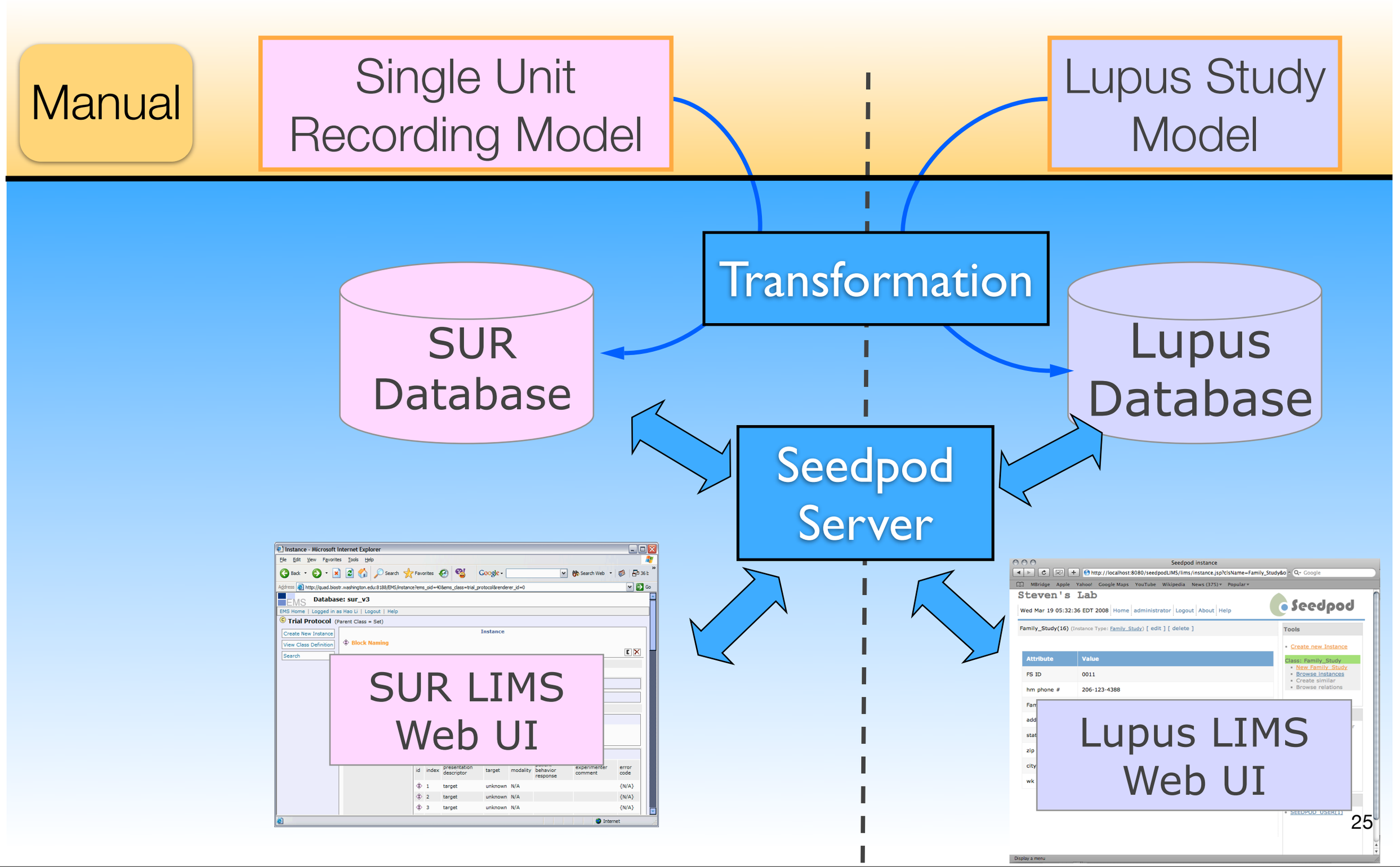
Model Independent Application



Model Independent Application



Model Independent Application



Model Independent Application

Manual

Automatic

Single Unit Recording Model

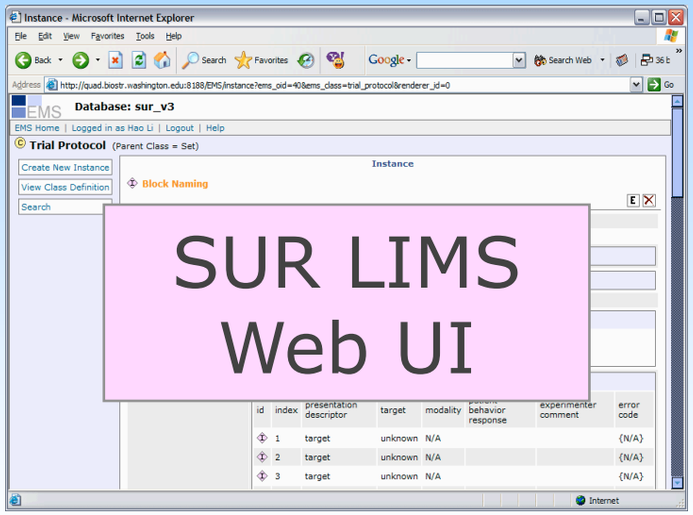
Lupus Study Model

Transformation

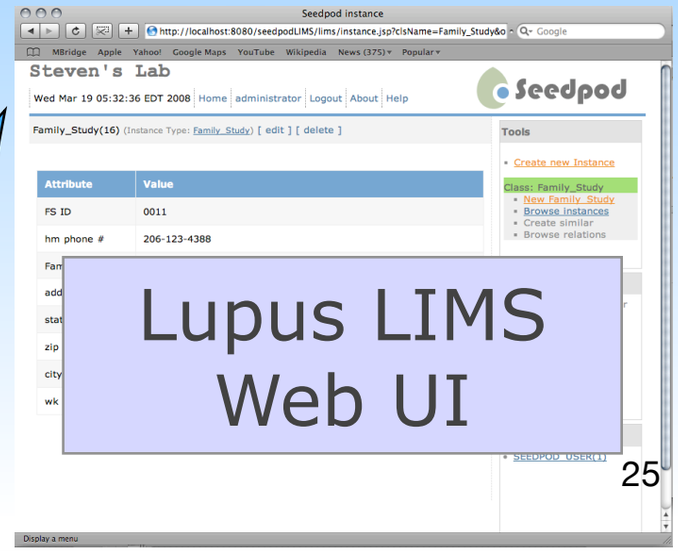
SUR Database

Lupus Database

Seedpod Server



SUR LIMS Web UI



Lupus LIMS Web UI

System Evaluation

Seedpod Aim	Significance	Test
More robust way to manage data	Allowing scientists to store and manage data in a relational database improves efficiency and query capabilities.	Compare a user query in Excel vs Relational database. Compare the amount of time and effort required. For example, ask for an aggregate question, ask someone to do that in excel. And then have a DBA to write the same query and get the result.
Reduce expertise required to LIMS development	Less expertise required by using Seedpod lowers the complexity and increases efficiency of system development	In the past, LIMS needs to be developed with a domain subject expert, someone who understands model design, relational database implementation, system engineering, and software development. Seedpod only requires a DSE and someone that understands how to model the data in Protégé.
Bridge the gap between domain subject matters and engineers	Reducing the upfront knowledge learning on the account of the researchers	Researchers and data model engineers can converse purely in terms of the domain subject and stay away from technical jargons of database development. The model makes it easier to say something about the data.
Lower cost	Increased reusability	The system only needs to be set up once and can be used to serve a community of labs.
Reduce the amount of time to make changes	Development can now keep up with change of experiment protocol	In the past, when experiment protocol changes, it may take another round of development time to create an experiment management system. Seedpod uses the amount of time to understand data model and the time to model.

Transformation Evaluation

Seedpod Aim	Significance	Test
Completeness	Is all information from protege model captured	List all entities in Protege and see if they have 100% coverage in the RDB Schema
Minimality	No redundant information	Test for a normalized database. Data is not captured redundantly.
Understandability	Subjective evaluation of whether the schema is easily understandable. (This is usually a trade off with completeness and minimality)	Have another DBA look at the schema and see if that's understandable.

Conclusion

- Seedpod
  - Provides a data management solution for biomedical researchers
  - Cost-effectiveness development for biomedical informaticists
- Future work:
  - A better modeling environment
  - Change management, protocol management
  - Data analysis, query interface, security and privacy

Thank you!

- Questions?

Supported by NIH Human Brain Project Grant RO1 MH/DC02310, and National Library of Medicine Training Grant program number: T15-LM07442.