

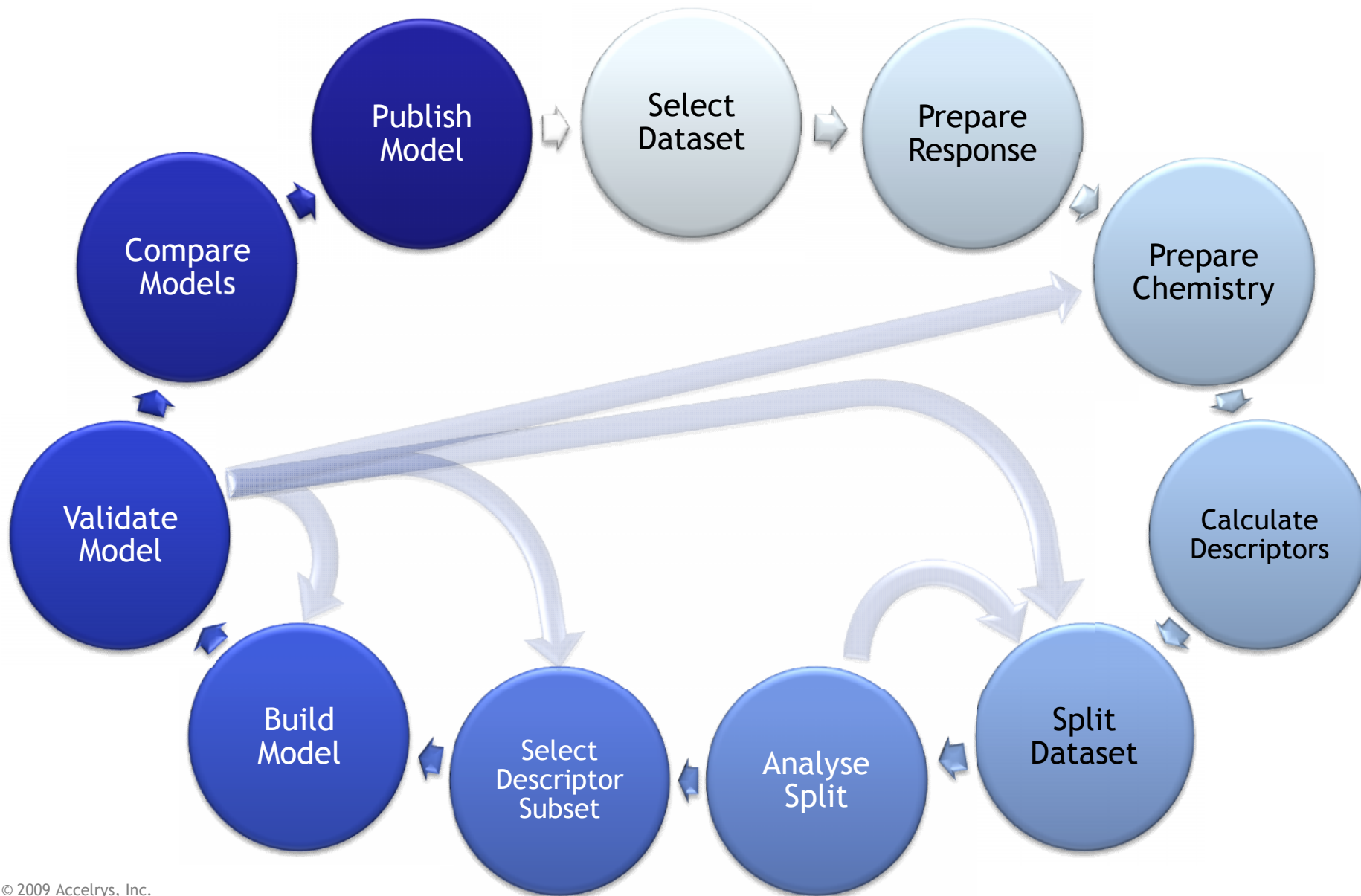
QSAR Workbench

A Project Driven Approach to Efficient Generation of QSAR Models

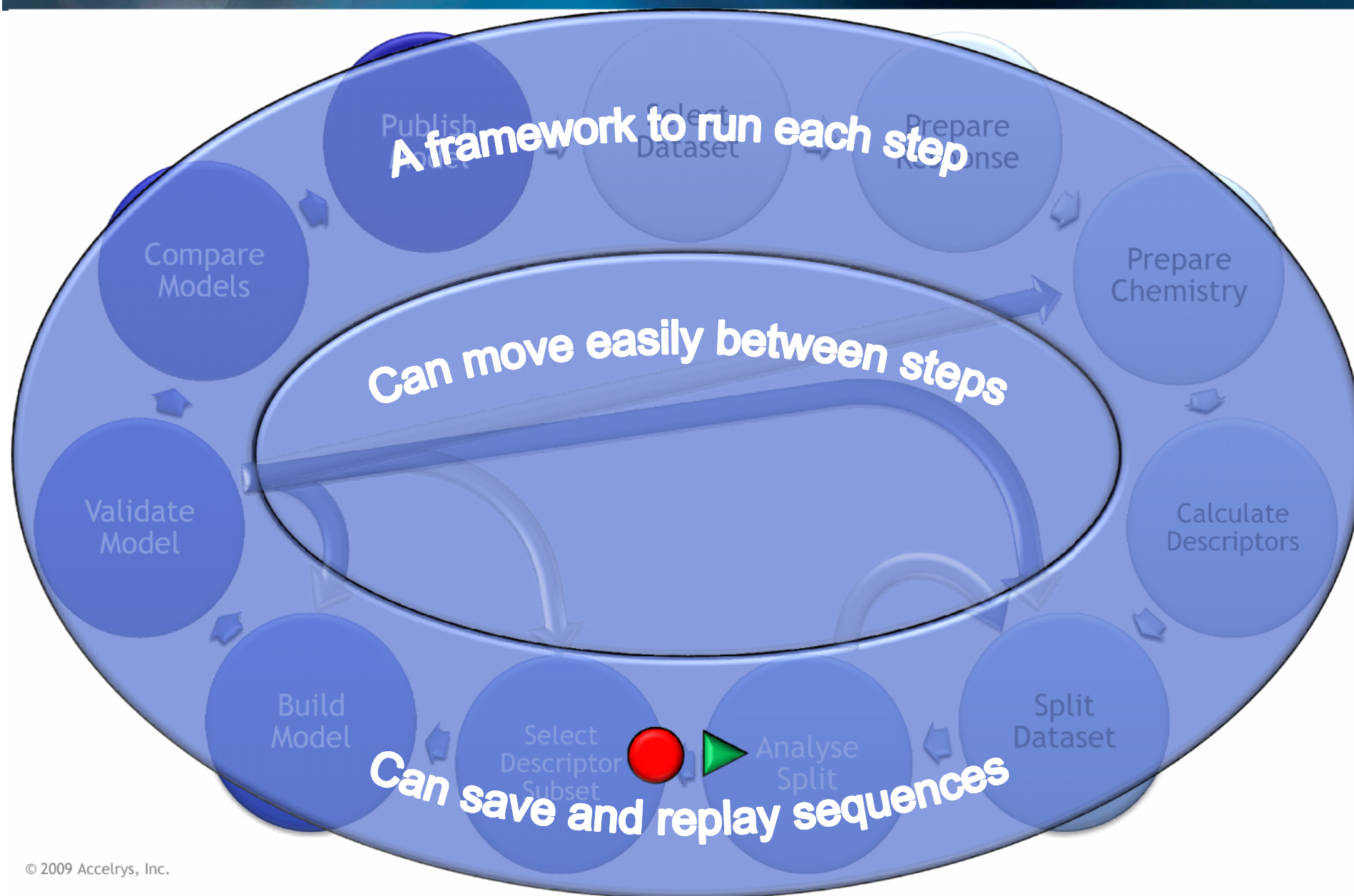
Tien Luu
Noj Malcolm

This presentation and/or any related documents contains statements regarding our plans or expectations for future features, enhancements or functionalities of current or future products (collectively "Enhancements"). Our plans or expectations are subject to change at any time at our discretion. Accordingly, Accelrys is making no representation, undertaking no commitment or legal obligation to create, develop or license any product or Enhancements. The presentation, documents or any related statements are not intended to, nor shall, create any legal obligation upon Accelrys, and shall not be relied upon in purchasing any product. Any such obligation shall only result from a written agreement executed by both parties. In addition, information disclosed in this presentation and related documents, whether oral or written, is confidential or proprietary information of Accelrys. It shall be used only for the purpose of furthering our business relationship, and shall not be disclosed to third parties.

The QSAR Workflow



The QSAR Workflow



Framework Benefits



QSAR Workbench - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/index.html

accelrys[®] QSAR Workbench

Welcome to QSAR Workbench, please use the links below to create a new project or open a previously created project

Create your QSAR project and load dataset

Projects

Delete

Project Name	Last Accessed	Status
MMP3	28 Jun 2010 17:16	✓✓✓✓✓✓

Create New Project

Create New Project

Step 2 of 2: Enter Response Details

Required fields are marked with an asterisk *

Project Name	CS3
Data Source	C:\Program Files\Accelrys\IP...
Data Source Format	SD
Response Property Field *	Activity
Response Property Type *	<input type="radio"/> Continuous <input checked="" type="radio"/> Categorical

Back Hide Preview Submit

Name	Activity
f1_10	1
f1_125	1
f1_139	1
f1_2	1
f1_138	1
f1_47	1

Open Project

Build Model

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3

accelrys[®] QSAR Workbench Project Name: CS3

Prepare Data Split data Descriptors Build Model Validate Model Publish Model

Prepare Data Tasks

- 01 Prepare Chemistry
- 02 Prepare Response
- 03 Review Chemistry Normalisation

The following protocol reads the input chemistry and standardizes the structures according to the selected options.
Please select which transformations to perform:

Submit

Normalisation Rule:	Do Action?
Generate 3D Coordinates:	<input checked="" type="radio"/> Yes <input type="radio"/> No
Add Hydrogens:	<input checked="" type="radio"/> Yes <input type="radio"/> No
Strip Salts:	<input checked="" type="radio"/> Yes <input type="radio"/> No
Standardize Molecule:	<input checked="" type="radio"/> Yes <input type="radio"/> No
Ionize at pH:	<input checked="" type="radio"/> Yes <input type="radio"/> No

Prepare Data Results

Results will appear here as tasks are performed

Curate your chemistry and data
Customise desired curation steps

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3&index=5

QSAR Workbench Project Name: CS3

Prepare Data | Split data | Descriptors | Build Model | Validate Model | Publish Model

Split data Tasks

- 01 Split Data
- 02 Visually Select Split
- 03 Analyse Split

Split data Results

File Name

- MyRandom85 Analy...
- MyRandom85 Details
- MyRandomChemistr...
- MyRandomChemistr...
- Visually Select Split

Create Training Set From Selection

Submit

Data Set Size: 752 Molecules

Legend	
Key	Data
Blue	Training Set
Red	Test Set
Green	Data Set
Orange	Selected

Response Property

Chemistry Based Splits

Click on any Data Set Split

Property Based Splits

Click on any Data Set Split

Chemistry Based Clusters

Click on any Data Set Split

Property Based Clusters

Click on any Data Set Split

PCA Plots

Test_pca_chem_PC2 vs Test_pca_chem_PC1

Test_pca_prop_PC2 vs Test_pca_prop_PC1

Create Training and Test sets with set methods or interactively Analyse Training and Test sets

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3&index=5

QSAR Workbench: CS3

Project Name: CS3

Prepare Data | Split data | **Descriptors** | Build Model | Validate Model | Publish Model

Descriptors Tasks

- 01 Calculate Descriptors
- 02 Create Descriptor Subset
- 03 Analyse Descriptor Subset
- 04 Combine Descriptor Subset

The following protocol calculates molecular descriptors for the project compounds, according to the selected options. Please select which descriptors to calculate:

1D descriptors

- Element Counts

2D descriptors

- AlogP Group
- Estate Keys
- Fingerprints
- Molecular Properties2D
- Molecular Property Counts
- Surface Area and Volume2D
- Topological Descriptors

3D descriptors

- Dipole
- Jurs Descriptors
- Molecular Properties3D
- Surface Area and Volume3D

Descriptors Results

File Name

- Calculate Descriptors Result 1
- Create Descriptor Subset Result

Calculate descriptors
and create
combinations of
descriptors (subsets)
Analyse descriptors

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3&index=5

QSAR Workbench: CS3

Project Name: CS3

Prepare Data | Split data | Descriptors | **Build Model** | Validate Model | Publish Model

Build Model Tasks

- 00 Set Learner Defaults
- 01 Build Single Model
- 02 Build Multiple Models**
- 03 Apply Model to Training Set

Builds multiple statistical models. A combinatorial expansion of the selected Model Types, Descriptor Subsets, and Dataset Splits. Please select the types of model to build and the dataset splits and descriptor subsets to use.

Model Types

- Scitegic Bayes
- Scitegic RP Tree
- Scitegic RP Forest
- R SVM
- R NN

Categorical Model Test :

Descriptor Subsets

- Select All
- EstateKeys
- Fingerprints
- Topological
- EstateKeysMolecular
- Molecular

Dataset Splits

- Select All
- MyRandom85
- MyRandomChemistry85

Parameter Set

- Select All
- Default

Build Model Results

File Name

- Build Multiple Models Summary
- Set Learner Defaults Result

Build single or multiple QSAR models with any combinations of dataset splits, descriptor subsets and model types

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3&index=5

QSAR Workbench: CS3

Project Name: CS3

Prepare Data | Split data | Descriptors | Build Model | **Validate Model** | Publish Model

Compare Training and Test set performance of all models
Apply models to external datasets

Validate Model Tasks

- 01 Apply Model to Test Set
- 02 Apply Model to External Set
- 04 Compare Multiple Models

Scitegic Bayes Models

Model ID	Training Set ROC	Test Set ROC	Split ID		
Model 11	1.000	1.000	MyRandom85	EstateKeys	Default
Model 12	1.000	1.000	MyRandom85	EstateKeysMolecular	Default
Model 13	1.000	1.000	MyRandom85	Fingerprints	Default
Model 14	0.9940	0.9950	MyRandom85	Molecular	Default
Model 15	0.9590	0.9340	MyRandom85	Topological	Default
Model 16	1.000	1.000	MyRandomChemistry85	EstateKeys	Default
Model 17	1.000	1.000	MyRandomChemistry85	EstateKeysMolecular	Default
Model 18	1.000	1.000	MyRandomChemistry85	Fingerprints	Default
Model 19	0.9940	0.9850	MyRandomChemistry85	Molecular	Default
Model 20	0.9610	0.9330	MyRandomChemistry85	Topological	Default

Viewing Details for: Model_20

Model Statistics | Model Parameters | **Charts** | Model Help Text

ROC Curve for Model_20 (Accuracy 0.961: Excellent)

ROC Curve for Model_20 (Accuracy 0.933: Excellent)

Validate Model Results

File Name ▲

- Model Comparison Report for CS3

QSAR Workbench: CS3 - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/project.html?name=CS3&index=5

QSAR Workbench: CS3

Project Name: CS3

Prepare Data | Split data | Descriptors | Build Model | Validate Model | Publish Model

Publish model for wider usage
Capture workflow of model building for replay

Publish Model Tasks

- 01 Publish Single Model
- 02 Publish Method**
- 03 Export Data

Publish Method

Publish a method (a series of tasks executed in a certain order) using the form below. Once a method is published other users can replicate the steps taken to build models in this project. Individual task options can be viewed as a tooltip on the Task Name.

Please enter a unique name for this method

Method Name *

<input type="checkbox"/>	Task Name	Original Run Date
<input checked="" type="checkbox"/>	01 Prepare Chemistry	28 Jun 2010 17:33:44
<input checked="" type="checkbox"/>	01 Split Data	28 Jun 2010 17:33:47
<input checked="" type="checkbox"/>	01 Split Data	28 Jun 2010 17:34:01
<input checked="" type="checkbox"/>	01 Calculate Descriptors	28 Jun 2010 17:34:57
<input checked="" type="checkbox"/>	02 Create Descriptor Subset	28 Jun 2010 17:34:58
<input checked="" type="checkbox"/>	02 Create Descriptor Subset	28 Jun 2010 17:34:59
<input checked="" type="checkbox"/>	02 Create Descriptor Subset	28 Jun 2010 17:35:01
<input checked="" type="checkbox"/>	02 Create Descriptor Subset	28 Jun 2010 17:35:02
<input checked="" type="checkbox"/>	04 Combine Descriptor Subset	28 Jun 2010 17:35:04
<input checked="" type="checkbox"/>	02 Build Multiple Models	28 Jun 2010 17:37:28
<input checked="" type="checkbox"/>	02 Build Multiple Models	28 Jun 2010 17:39:27

Results will appear here as tasks are performed

QSAR Workbench - Microsoft Internet Explorer provided by Accelrys

http://genius:2544/webapps/qsarworkbench/index.html

accelrys QSAR Workbench

Welcome to QSAR Workbench, please use the links below to create a new project or open a previously created project

Projects		
✘ Delete		
Project Name	Last Accessed	Status
FXa		✓✓✓✓✓✓
CS3	28 Jun 2010 17:37	✓✓✓✓✓✓
MMP3		✓✓✓✓✓✓

Create New Project

Build Model

Project Name: FXa

Choose a published method to run from the list below

Published Method * MMP3-Bayes-SVM

Submit

Open Project

Build Model

Replay model building process on new dataset

Acknowledgements

- Noj Malcolm
- Marc Meunier
- Patricia Gestoso-Souto
- Guillaume Paillard
- Adrian Stevens
- Richard Cox