Flexible Solution and Integrated Web Interface for Processing and Uploading of Diverse Bioassay Data

D. Bondarev$*, A. Filikov$, L. Herman$%, J. Patterson $&, C. Wang$, O. Stradella$

$Genzyme a Sanofi Company, Waltham, MA; *currently at Saber Informatics, Burlington, MA; %currently at Lee Herman Consulting LLC, Natick, MA;
&currently at Pattermatix Consulting, LLC , Maynard, MA

Abstract

The Sanofi Bioassay Data Capture system enables efficient management, processing, analysis, and storage of diverse bioassay data.

Then main differentiation requirements vs similar existing commercial applications were: ease of use, fixed but customizable workflow, capable of handling any assay type and any analysis algorithm, strict enforcement of business rules and QA procedures.

We designed and implemented a web application for bioassay data capture by *in vitro* biologists. Along the way we learned what screening biologists really want in a great web application; we present our findings here.

The key feature of this project is its collaborative nature. Since our research IT group is located onsite with the *in vitro* group, the collaboration was informal and we could iterate the development very quickly.

The application was built primarily with Pipeline Pilot. Web-based user interface design delivers convenience and user-friendliness without sacrificing performance.

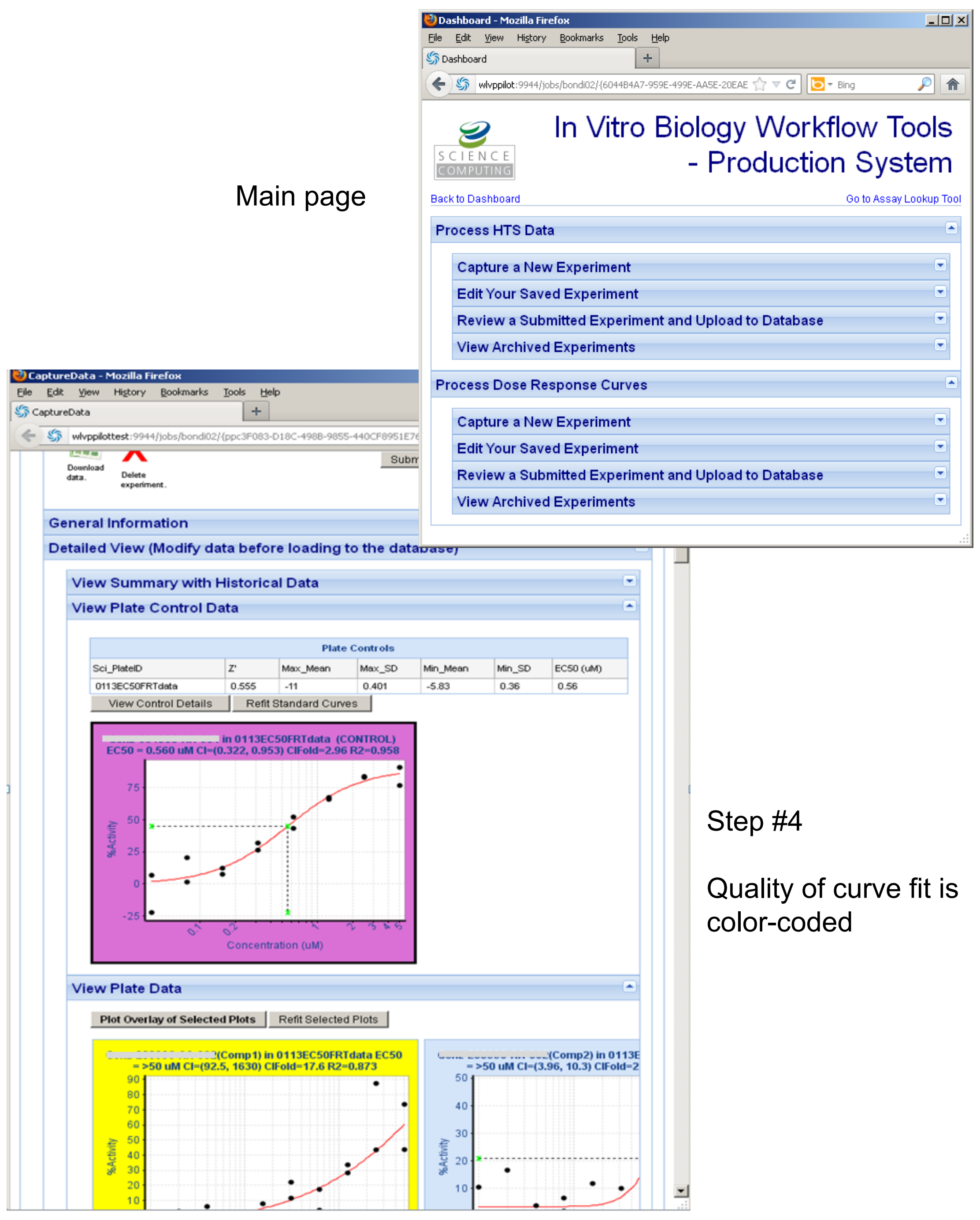
Current key features

- Web-based application built on the Pipeline Pilot platform.
- Automation of capture and storage of results from primary screening and dose response experiments.
- Streamlined and user friendly workflow designed by biologists. Workflows are fixed at the assay type (or even individual assay) level, exposing only a small and controlled subset of options to the user. This increases the homogeneity and quality of the results since they are all processed using the same approved and tested procedure.
- Includes a multi-step review and approval process, and email notifications.
- Data capture, analysis, review, visualization and upload in a single user interface.
- Support for assays of any type including multi-parametric, kinetics, SPR, etc.
- Several curve fitting models with user-controlled parameters provide high flexibility for data processing.
- Export to excel and pdf formats.
- Users can create annotations at the data point and project level; annotations are stored as part of data capture.

Collaboratively designed workflow

The application workflow is a product of a collaborative design with assay scientists – biologists' contribution is crucial – it is them who generate and process data. The main steps are outlined below.

1. Define experiment metadata (name, date, parameters, plate layout, etc.)
2. Capture experiment data from input files, validate the data
3. Display data for editing (curve re-fit, control details, etc.)
4. Edit and submit for review. Email notification sent to operator and reviewer
5. Reviewer: Edit (optional) and load to database. Email notification with report sent to operator and reviewer
6. View archived experiment (optional)



In Vitro Biology Workflow Tools - Production System

Back to Dashboard | Go to Assay Lookup Tool

Process HTS Data
- Capture a New Experiment
- Edit Your Saved Experiment
- Review a Submitted Experiment and Upload to Database
- View Archived Experiments

Process Dose Response Curves
- Capture a New Experiment
- Edit Your Saved Experiment
- Review a Submitted Experiment and Upload to Database
- View Archived Experiments

General Information

Detailed View (Modify data before loading to the database)

View Summary with Historical Data

View Plate Control Data

Plate Controls						
Sci_PlateID	Z'	Max_Mean	Max_SD	Min_Mean	Min_SD	EC50 (uM)
0113EC50FRTdata	0.555	-11	0.401	-5.83	0.36	0.56

Main page

Step #4

Quality of curve fit is color-coded

Reusable analysis templates

While there were different assays in use. many of these shared similar statistical analysis requirements. We used Pipeline Pilot components to build data analytics workflows which made it straightforward to re-use data capture and analysis templates across multiple assays with minor modifications.

In addition to data analysis there were differences as well as similarities in the user interface design across assays. Again, we were able to re-use repeatable workflow fragments as Pipeline Pilot components – this time for web-based user interface functionality, look and feel.

Data QC and visualization

Good visualization of data is very important for efficient analysis. Below are two examples of our visualization tools that are helpful in understanding results and assessing their quality.

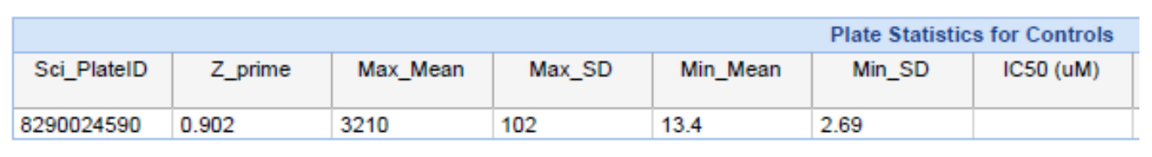
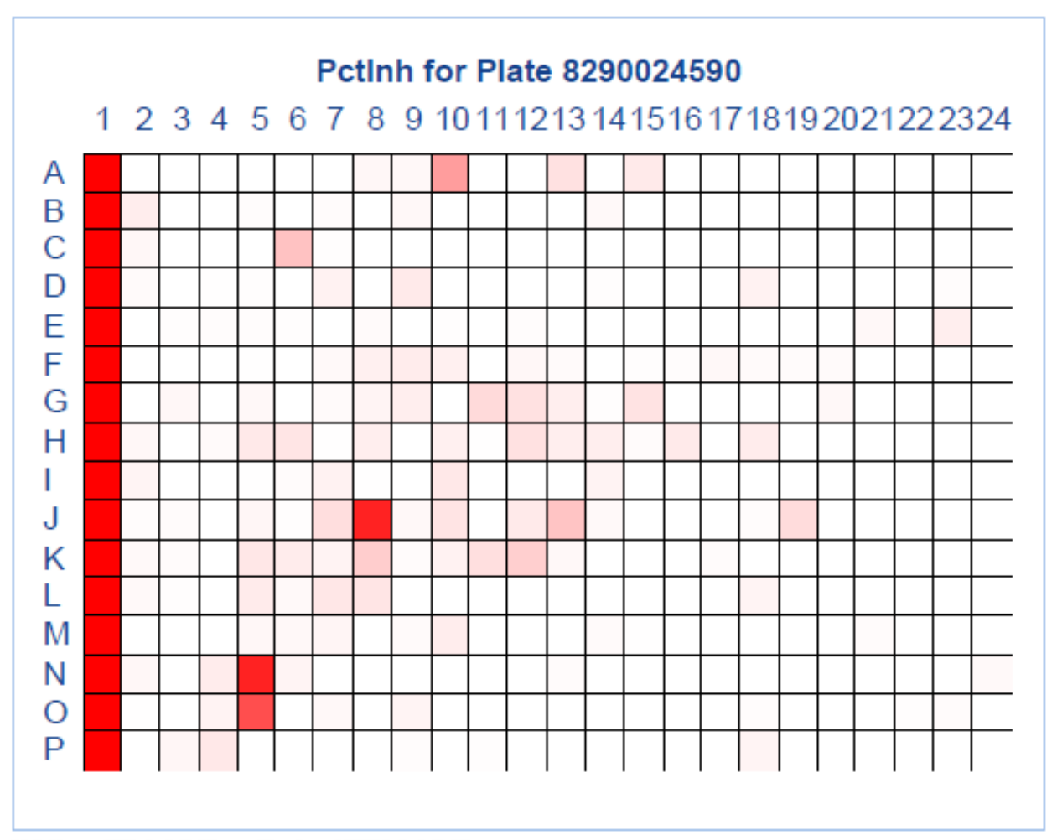


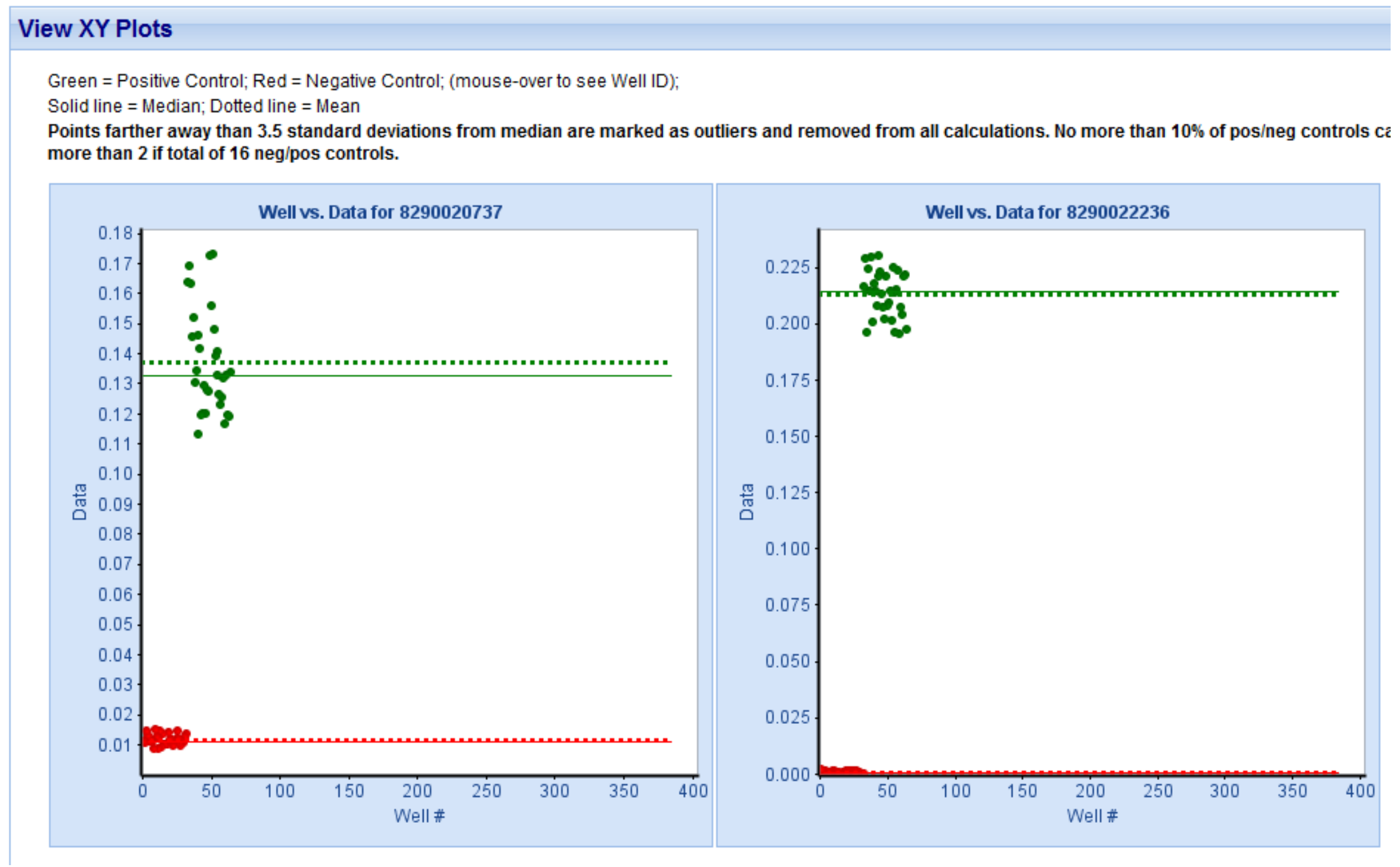
Plate Statistics for Controls						
Sci_PlateID	Z_prime	Max_Mean	Max_SD	Min_Mean	Min_SD	IC50 (uM)
8290024590	0.902	3210	102	13.4	2.69	



White = 0% Inhibition; Red = 100% Inhibition

Data on a plate presented as a heat map:

Quality metrics of an assay run can be visually assessed:



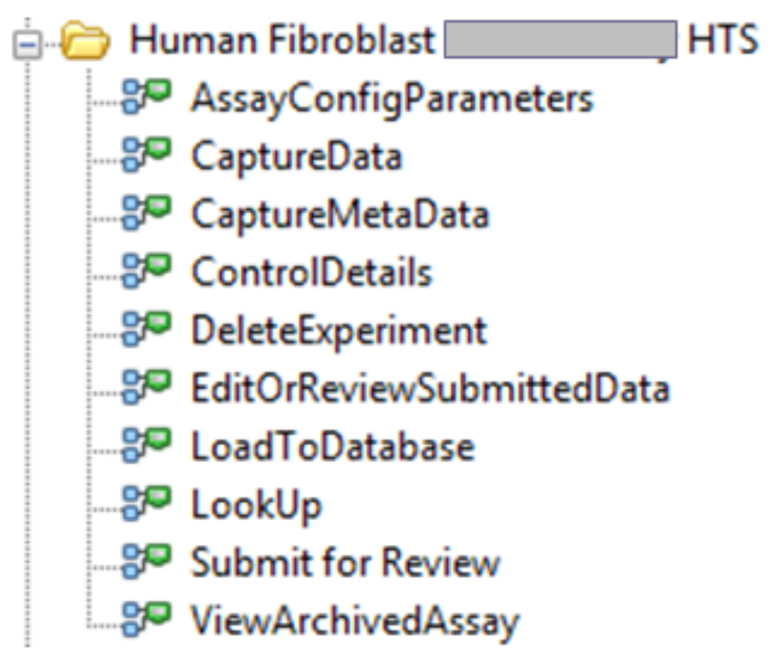
View XY Plots

Green = Positive Control; Red = Negative Control; (mouse-over to see Well ID);
Solid line = Median; Dotted line = Mean
Points farther away than 3.5 standard deviations from median are marked as outliers and removed from all calculations. No more than 10% of pos/neg controls ca more than 2 if total of 16 neg/pos controls.

Modular design

Because of our existing legacy data, specific workflow requirements, and the need for algorithmic flexibility in assay data processing, a significant requirement in the project was to integrate with our existing research IT infrastructure: databases, plate registration systems, and downstream analytics in Seurat.

We decided to use Pipeline Pilot and to utilize its database integration, statistical, and web reporting components. The component-based architecture allowed us to adapt to the changing research IT environment in the course of this project while keeping the application operational.



Human Fibroblast [redacted] HTS
- AssayConfigParameters
- CaptureData
- CaptureMetaData
- ControlDetails
- DeleteExperiment
- EditOrReviewSubmittedData
- LoadToDatabase
- LookUp
- Submit for Review
- ViewArchivedAssay

As the project progressed our users asked for additional functionality. With a modular design we could incrementally extend the application without rewriting the existing functionality. For example, we added email notifications with live links and PDF report attachments in messages. Reviewers could enter the application by clicking on the links – and they would continue the QC and review workflow in exactly the same way as if they started from the main portal page.

The modular design also makes possible to impose unique fixed workflows to all assays within one assay type or even to individual assays when necessary.

Acknowledgments

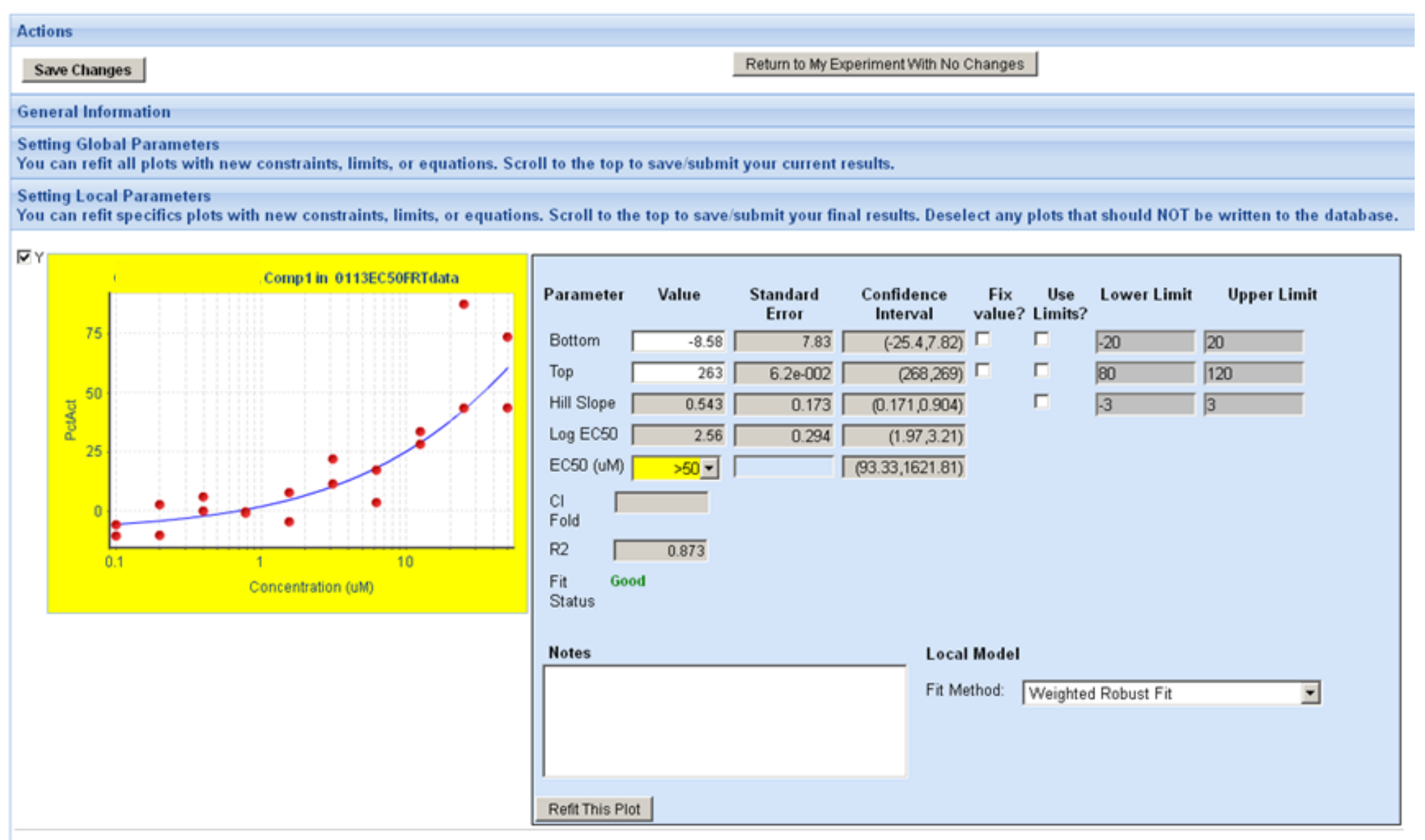
David Stepp	Elina Makino
Tatiana Gladysheva	Theresa Kuntzweiler
Alexander Michel	Alissa Sheils
Beirong Liang	Gary Asmussen
Gregory Hurlbut	Maria Fitzgerald
Maureen Olszewski	Paul Mason
Sanjay Danthi	Timothy Weeden

Curve fitting algorithms

Our *in vitro* data was unsuitable for typical off-the-shelf automated fitting tools: in too many cases there would be no fit at all (and experiments would have to be re-done) or arbitrary constrains were required to force a fit.

We implemented and successfully validated a set of minimally constrained fit algorithms known to work with these types of data. The fit has been implemented as a multi-step workflow with logical branches tailored to different categories of assays. Within each branch there is a cascade of progressively more constrained fit attempts each of which is automatically evaluated against a set of fit threshold criteria. The criteria are uniform across all assay types regardless of the specific fit parameters or method.

Biologists review automated fit results and can manually override (refit) the curves if needed. Typically less than 5% of curves benefit from a manual refit. Such overrides are automatically tracked by the application and users are given an opportunity to add comments to the fit results.



Setting Global Parameters
You can refit all plots with new constraints, limits, or equations. Scroll to the top to save/submit your current results.

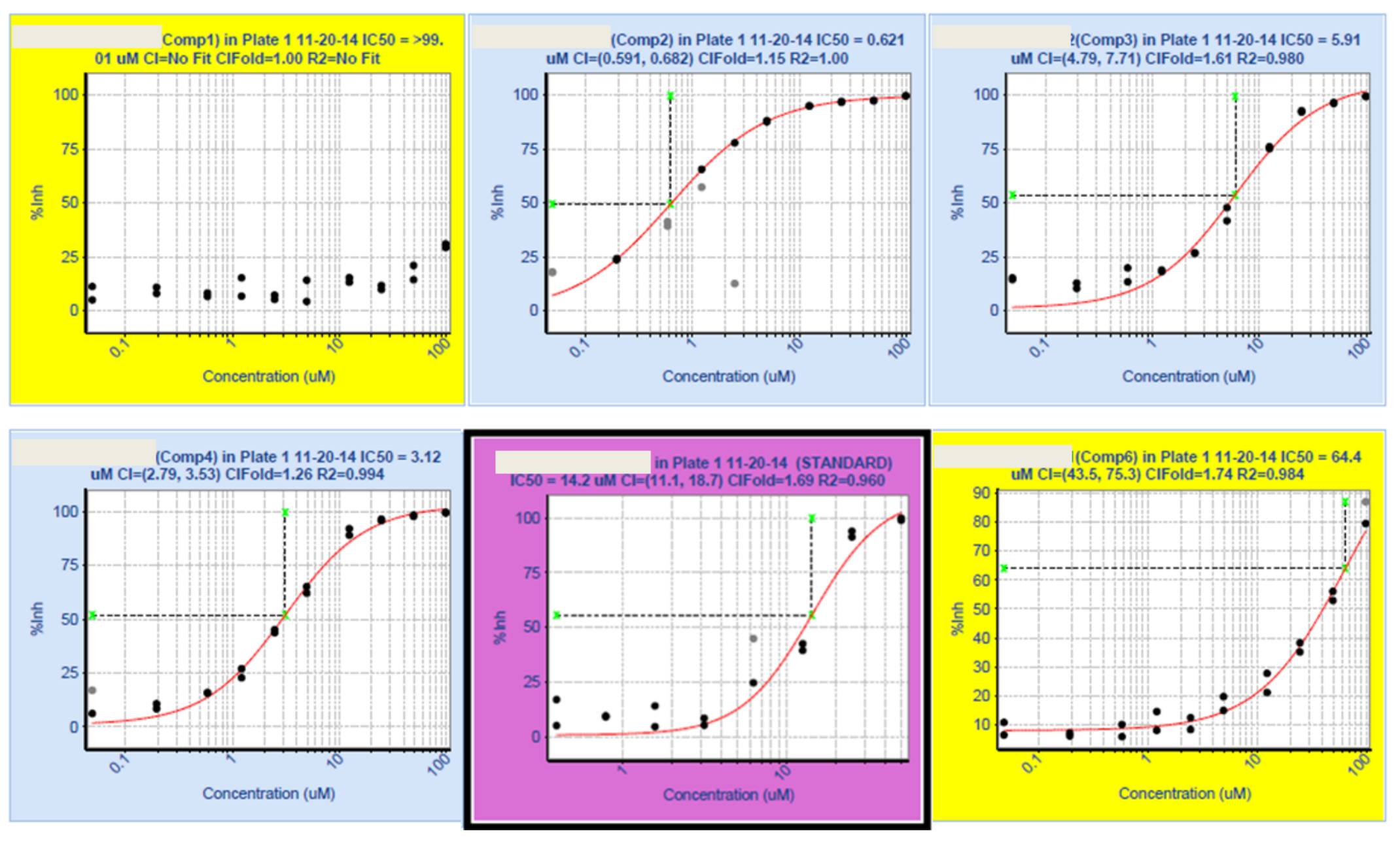
Setting Local Parameters
You can refit specifics plots with new constraints, limits, or equations. Scroll to the top to save/submit your final results. Deselect any plots that should NOT be written to the database.

Parameter	Value	Standard Error	Confidence Interval	Fix value?	Use Limits?	Lower Limit	Upper Limit
Bottom	-8.58	7.83	(-25.4,7.82)	☐	☐	-20	20
Top	263	6.2e-002	(268,269)	☐	☐	80	120
Hill Slope	0.543	0.173	(0.171,0.904)		☐	-3	3
Log EC50	2.56	0.294	(1.97,3.21)				
EC50 (uM)	>50		(93.33,1621.81)				
CI Fold							
R2	0.873						
Fit Status	Good						

Notes

Local Model
Fit Method: Weighted Robust Fit

Fitting algorithms work in high throughput manner. Only low quality data require manual intervention. Quality of curve fit is color-coded



Reporting tools

Requirements for the reports that users get after uploading differ very much between assays – some assays/users are happy with basic reports, but some request reporting non-standard statistics or other calculations. Our open design, together with the Pipeline Pilot platform, allows us to accommodate any requests. This automates and therefore enhances efficiency of scientist's workflow, which definitely contributed to the success of the application.

Conclusions

- **Biologists' workflow design recommendations were key to success**
- **Optimized compartmentalization of the whole application into smaller modules enabled flexibility in adapting to different assays**
- **Well-designed visualization tools are instrumental to understanding results and assessing their quality**
- **Implementing a set of minimally constrained curve fitting algorithms allowed for flexibility in data processing and helped rescue significant amounts of data**
- **Web-based design and use of the Pipeline Pilot platform give convenience and user-friendliness without sacrificing performance**
- **Seemingly small convenience features like confirmation emails are important to users**

SANOFI BOSTON R&D HUB