Simplifying and Standardizing Data Analysis at Merck with a Common ActivityBase Platform

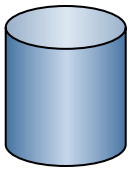
Alexander Szewczak, Ph.D.
Director, Pharmacology and Assay Operations Lead, Merck Boston
IDBS Boston Best Practices Screening Forum
22 May 2014

Background

- Major initiatives at Merck to streamline and globalize operations, rationalize facilities footprint and use more external resources
- Strategic decision made to not develop software internally
- Merck in-house software for analysis of in vitro pharmacology assay data was approaching "end of life."
  - Used primarily for high and medium- throughput production dose-response assays in micro titer plate format, including some uHTS follow up work.
  - Some single concentration screens as well.
  - IC50/EC50 4 parameter fits, agonism/antagonism, kinetics, etc.
  - Template-based (both assay and calculation) with assay meta data included.
  - Data pushed to single corporate repository.
  - Not well suited for ad hoc and one-off experiments.
- Originally developed at Merck Frosst Montreal.
- Deployed across all therapeutic areas and many functional departments at nine major sites in UK, Italy, Spain, Japan, Canada, U.S.

In Vitro Pharmacology Data Workflow

Reader / Raw Data

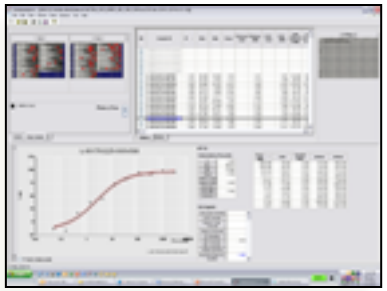


Compound Management Database / Plate map Information

Or

Compound List (standard format experiment)

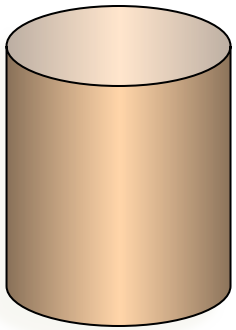
Assay Meta Data



Experiment Analysis

Processed Data

Assay Meta Data



Data Repository

Summarized Data

Full Analysis of All Existing Data

In Vitro & In Vivo Activity plus Chemical Properties

Processed Data -not yet uploaded

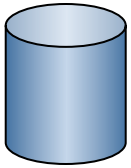
and/or

Previously Summarized Data

Real-time Results

In Vitro Pharmacology Data Workflow

Reader / Raw Data

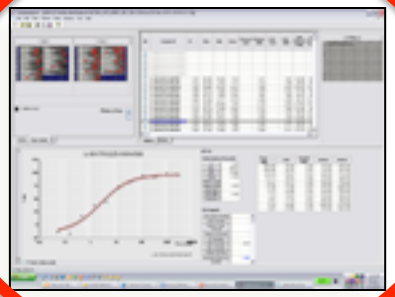


Compound Management Database / Plate map Information

Or

Compound List (standard format experiment)

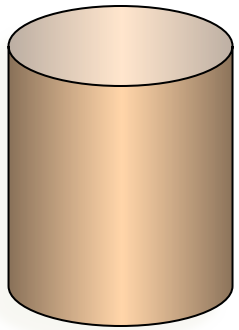
Assay Meta Data



Experiment Analysis

Processed Data

Assay Meta Data



Data Repository

Summarized Data

Full Analysis of All Existing Data

In Vitro & In Vivo Activity plus Chemical Properties

Processed Data -not yet uploaded

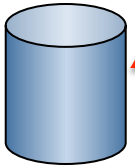
and/or

Previously Summarized Data

Real-time Results

In Vitro Pharmacology Data Workflow

Reader / Raw Data

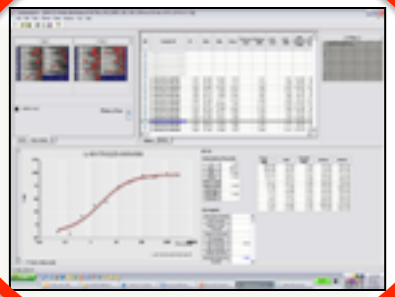


Compound Management Database / Plate map Information

Or

Compound List (standard format experiment)

Assay Meta Data



Experiment Analysis

Key node in the data workflow

Processed Data -not yet uploaded

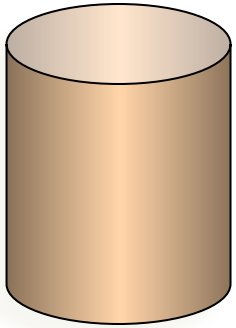
Real-time Results

and/or

Previously Summarized Data

Processed Data

Assay Meta Data



Data Repository

Summarized Data

Full Analysis of All Existing Data

In Vitro & In Vivo Activity plus Chemical Properties

In Vitro Pharmacology Data Workflow

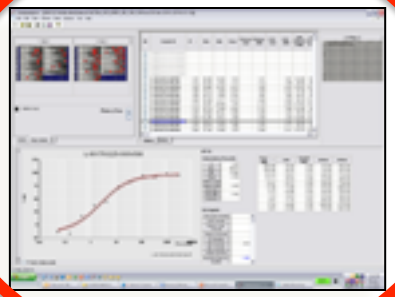
Reader / Raw Data

Compound Management Database / Plate map information

Or

Compound List (standard format experiment)

Assay Meta Data



Experiment Analysis

Key node in the data workflow

Processed Data -not yet uploaded

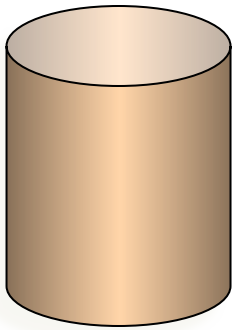
and/or

Previously Summarized Data

Real-time Results

Processed Data

Assay Meta Data



Data Repository

Summarized Data

Full Analysis of All Existing Data

In Vitro & In Vivo Activity plus Chemical Properties

Issues with the legacy system

- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years, about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Issues with the legacy system

- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years, about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Issues with the legacy system

- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years, about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Issues with the legacy system

- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years, about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Issues with the legacy system

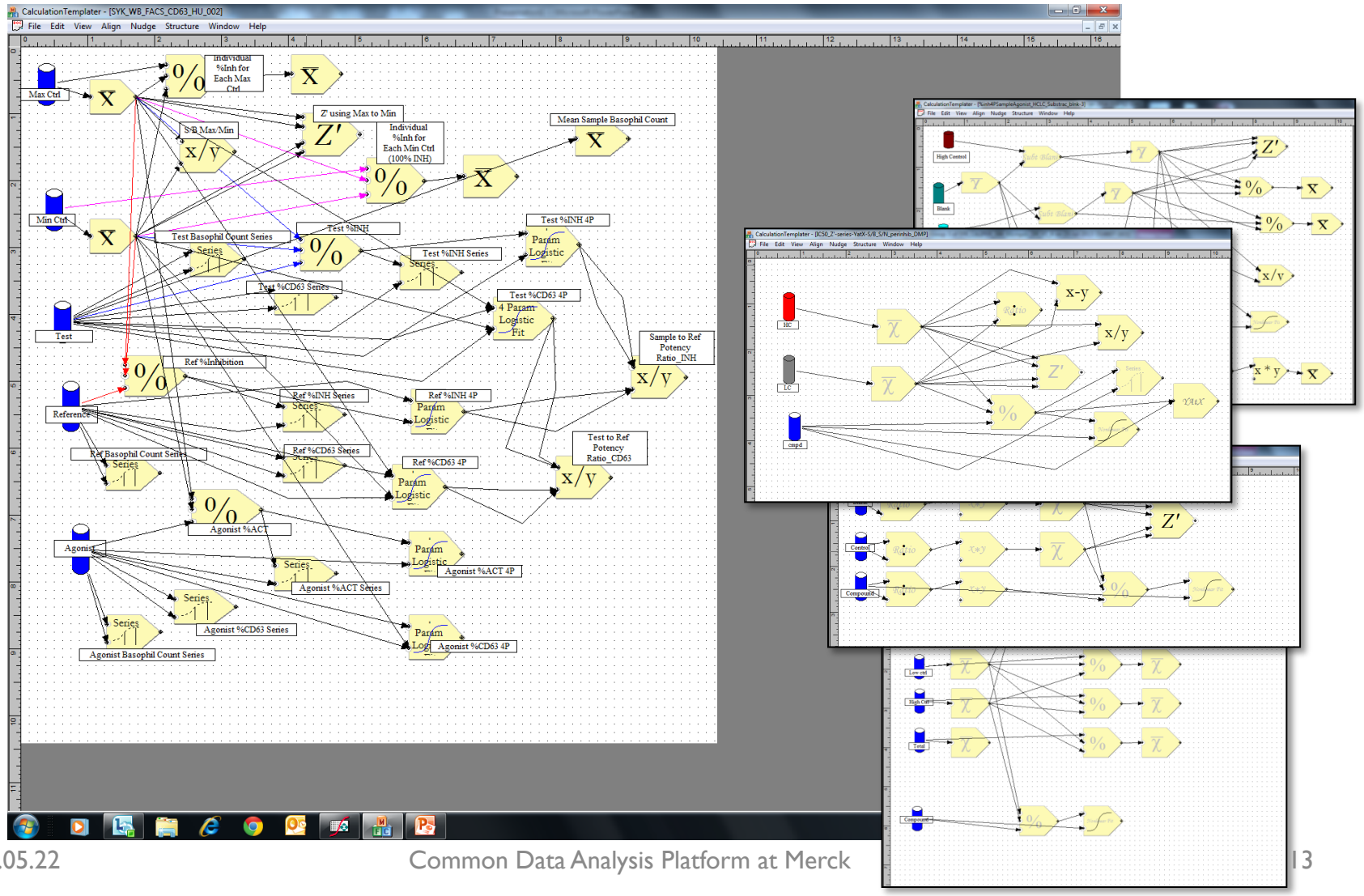
- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years; about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Issues with the legacy system

- Extremely flexible, enabling a wide range of individually-customized data analysis.
- Fostered a culture in which every research team could analyze their data in a custom way.
- Template design typically based on type of experiment or technology (e.g. "HTRF", "AlphaScreen", "luminescence").
- Template creation and maintenance became a huge burden.
  - Each assay required its own template, which was created by a distributed team of IT support staff located across the globe.
  - Support staff would typically start with an existing custom template (usually one of their own) and modify it, rarely starting from scratch (early differentiation!).
  - Over 1800 different calculation templates and 3000 plate formats accumulated over the past 10 years, about 500 users.
  - Graphical design tool for calculation definition – too many degrees of freedom.
  - Many types of normalization.
  - Many different plate formats; little standardization around local hardware.
- Many different ways of carrying out essentially the same measurements and data analysis. Hard to share or transfer assays.

Legacy calculation templates

Incremental innovation to the point of diminishing returns



The benefit of information technology

"**Information technology forces you to organize your processes more logically**. The computer can handle only things to which the answer is yes or no. It cannot handle maybe. It's not the computerization that's important, then; it's the discipline you have to bring to your processes. You have to do your thinking before you computerize it or else the computer simply goes on strike."

- Peter Drucker

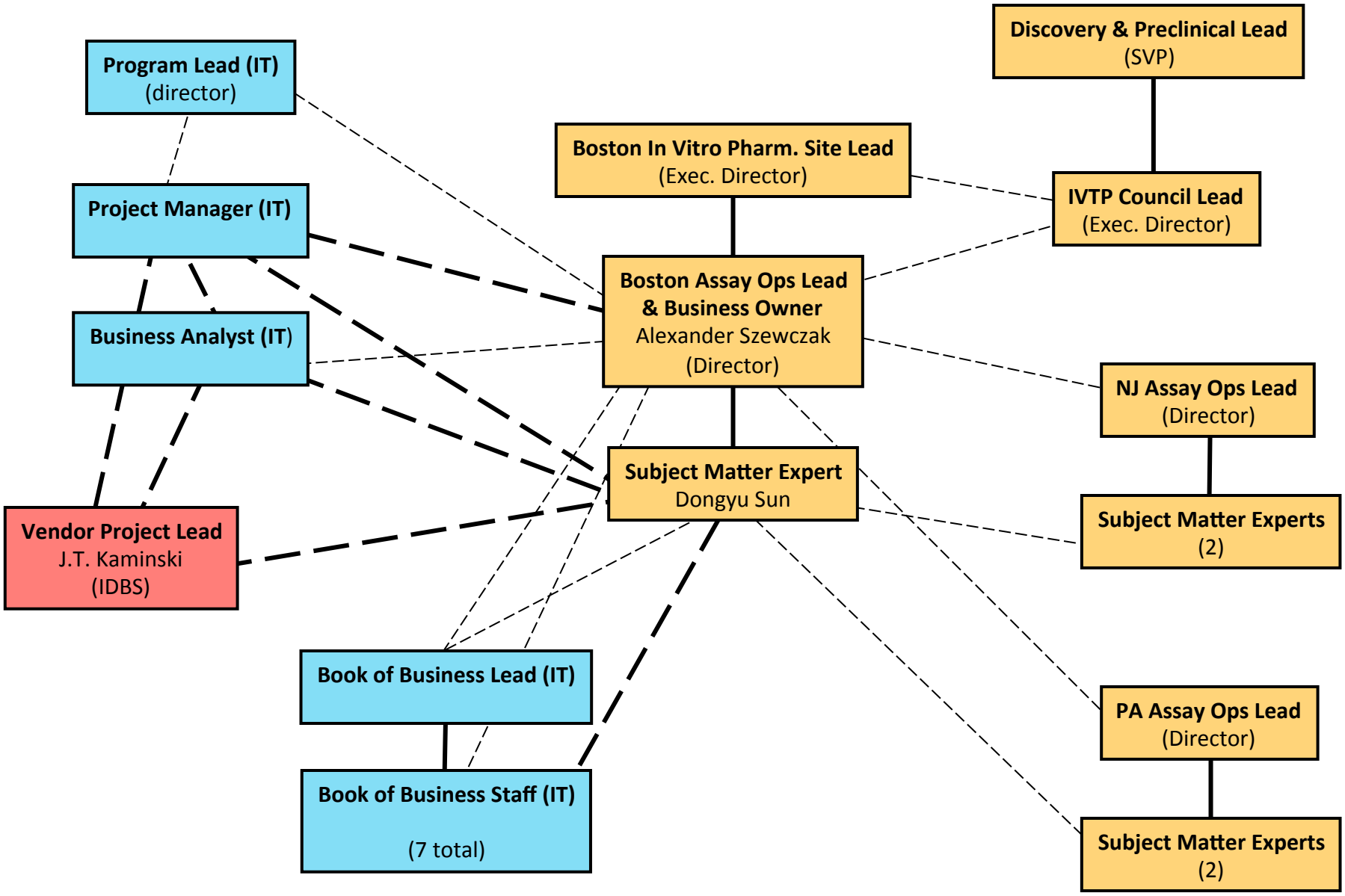
Project Goals

- Adopt ActivityBase XE as replacement platform across Merck.
- Standardize data analysis around uniform best practices.
- Reduce complexity and lower maintenance costs.
  - Fewer assay templates, platemaps, calculations
  - Easier to transfer assays internally and externally
  - Easier to train users and IT staff

Standardization vs. customization

- Need to let teams respond flexibly to project scientific demands.
- Yet often the same basic calculations are being done.
- Standardization facilitates information sharing, training, load balancing, and reduces unnecessary complexity.
- Delayed differentiation more efficient way to customize.

PROJECT STAKEHOLDER NETWORK MAP, 30 Oct 2012



Solid lines = formal reporting, dashed lines = informal relationships. Heavy lines indicate higher level of interactions.
Red box = IDBS (software vendor), Blue boxes = Merck IT, Orange boxes = Merck Discovery/In Vitro Pharmacology. Note lack of interaction between IT Project and IT BoB groups.

Project action plan

1. Build awareness (many researchers didn't know about the project)
   - Email project introduction and status updates
   - Present new software overview at departmental meetings
   - Create project wiki page
2. Get stakeholder buy in (there was resistance to the initial IT plan)
   - Include scientists in the design of analytical templates and software pilots
   - Consult with IT management, IT support staff (somewhat ignored previously), scientists at other sites
3. Communicate, communicate, communicate! (People respond badly when they don't know what's going on or feel left out)
4. Do a pilot rollout at Boston site (corporate-wide is no way to start)
   - Build 4 initial analysis templates & benchmark calculations against existing protocols
   - Test data import, analysis and uploading to corporate database
   - Finish Boston training and template conversions
5. Launch Standards Committee and build company-wide institutional knowledge
   - Funnel standards information directly to the wiki pages
6. Use lessons learned to guide full-scale rollout to remaining Merck sites and external partners (6+ sites, ~300 users)

Global Standards

- Raw data file and import definition
- Standard vocabulary
- Plate formats
- Raw data normalization
- Abase master templates (std. calculations)
- Business rules
- Experiment Condition Grouping ID (ECG_ID)

Raw data file and import definition

Abase native method for import definition is a quick solution for data parsing. But each minor variation requires a new import definition, while potentially adding little or no value.

- Limit plate reader raw data export to standard formats
  - Same assay technology, same output format
  - Allow exceptions when there is a business justification.
- Customized Python import definitions on the front end
  - Recognizes multiple different file formats, allows them to be parsed by same script
  - Called by ActivityBase XE Runner
- These efforts substantially reduced unnecessary complexity of data file outputs and import definitions.

Standard vocabularies

- Variables and result types
  - Avoid IC50, ic50, IC_50, ic_50 ambiguity, etc.
- Assay types and technology names
- Plate formats
- Standard assay naming conventions

  For example:

  (Assay Target)_ (Assay Technology)_(Assay Category)

  TARGET#1_HTRF_ENZ

Raw data normalization

Adopt a single transformation formula: "% Desired Effect" which covers %Inhibition, %Activity, %Binding, etc. in one calculation:

% E = 100 x
(Response - No_Effect_Control) / (Positive_Effect_Control - No_Effect_Control)

"No_Effect_Control" means the appropriate control displaying no effect of interest (i.e. full biological activity for an antagonist or inhibitor screen; alternatively, zero or background activity for agonist/activator screens).

"Positive_Effect_Control" refers to the control that shows the full (or as full as possible) expected effect of interest. (i.e. complete inhibition for an antagonist or inhibitor screen, or complete activation for an agonist or activator screen.

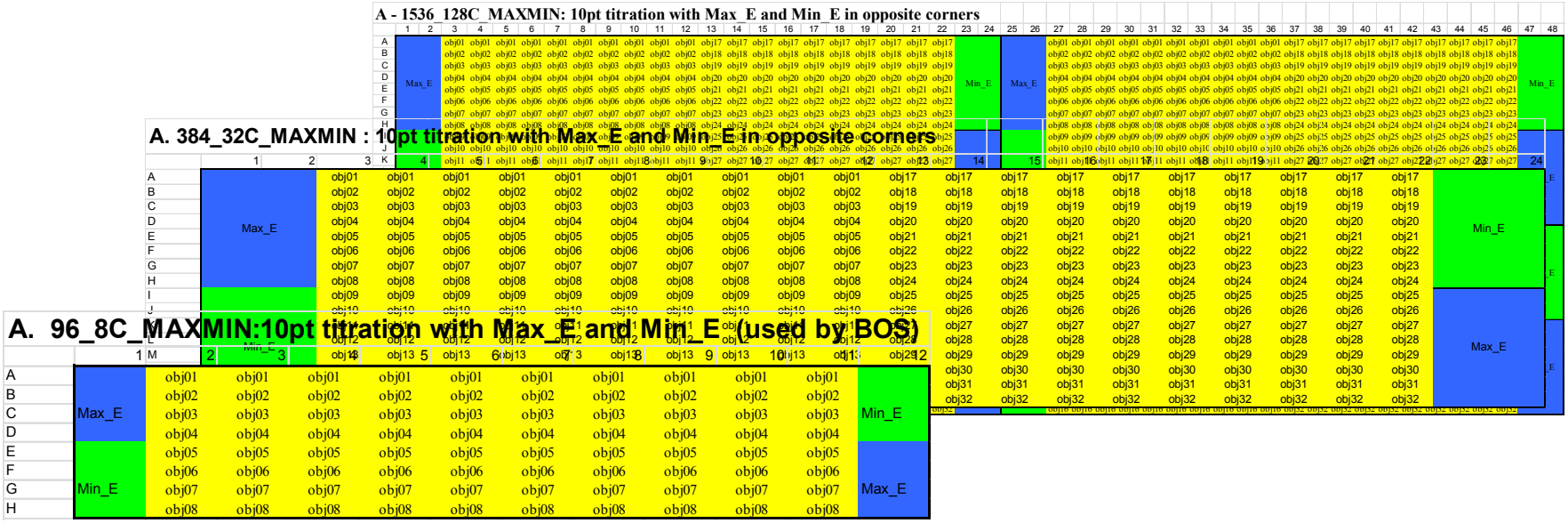
In all cases, "effect" refers to the major effect of interest for which the target is being screened (binding, agonism, activation, antagonism or inhibition, including allosterism).

Adopted from Schering/Organon

Plate Formats

- Plate formats were reduced to about forty total, including 96, 384 and 1536 well plates for single point as well as 6, 7, 8 and 10- point dose response analysis.
- Will use Abase "obsolete" feature for unused plate formats
  - Allow plate formats be phased out when they are no longer in production.
- Symmetric and fixed locations for no effects and positive controls
- Fixed locations for assay standard controls
- Standard nomenclature for plate format
- Synchronize plate formats of different densities
  - Allows compound source plates and assay plates in different formats in an assay, i.e. four 384 well compound source plates to one 1536 well assay plate, vice versa, etc.

Standard Assay Plate Formats



- ❑ 96 well — 1, 8, 10 pts (14 layouts)
- ❑ 384 well — 1, 6, 7, 10 pts (28 layouts)
- ❑ 1536 well — 1, 10 pts (4 layouts)

- ✓ *For both internal and external*
- ✓ *New layouts can be added if the current formats do not meet business needs*
- ✓ *Some layouts will be phased out as existing assays go away*

Fewer Plate Formats Reduce Complexity and Save Resources!

Master Templates

- Key concept: Create a limited set of standard templates, based primarily on number of signals and the need for a standard curve (back fitting).
  - Many different assay technologies use essentially the same calculations.
- Use "off the shelf" master template as the starting point to create appropriate solution (delayed differentiation)

Considerations for template design

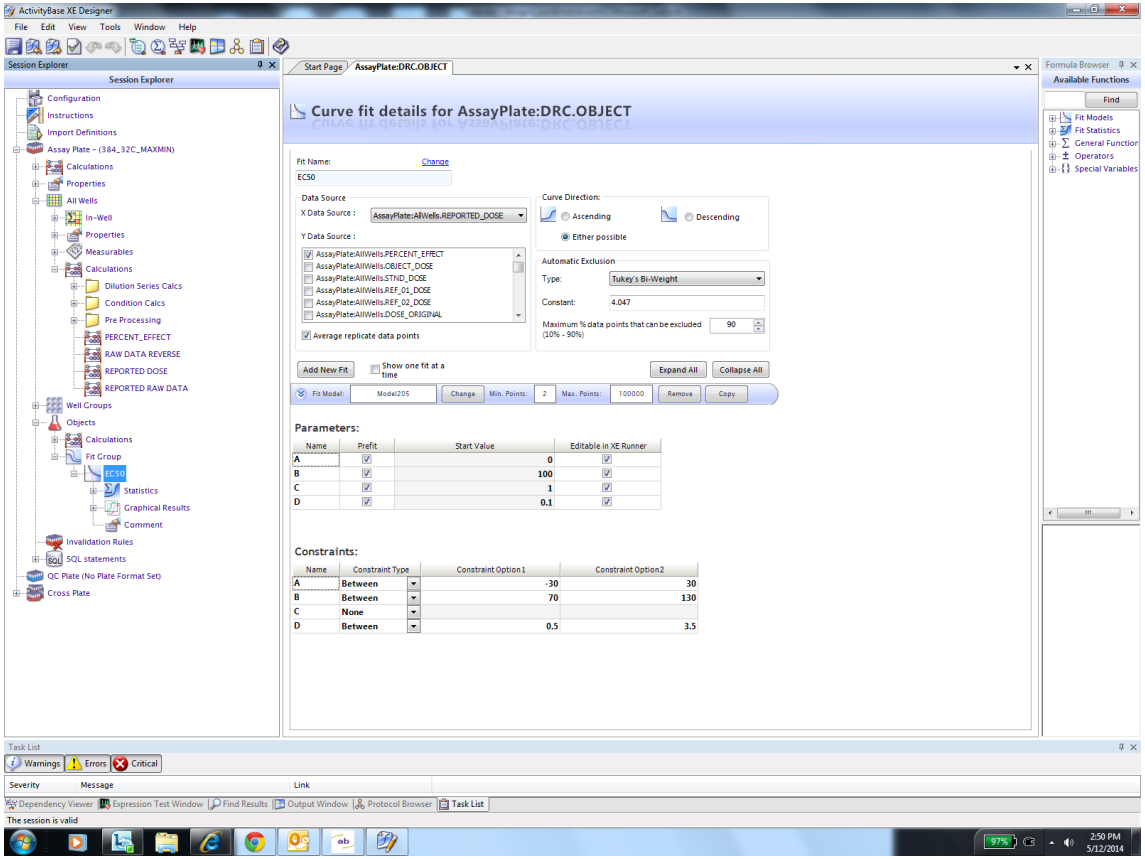
- Number of read(s) in raw data file(s) per plate
  - The primary factor determining the number of master templates
  - With or without data transformation using a standard curve
- Transformation of raw data – pre-processing
  - In-well analysis: separate category (kinetics, FLIPR, etc.)
  - Raw data calculation: i.e. ratio of raw data
- Normalization of transformed data: single standard calculation, %E
- Curve fitting and results
  - EC50, 4P parameters, etc.
- Derived results (Ki, relative efficacy, etc.) can also be calculated.

- Include a superset of features/variables/results. Use them only when needed.

Standard is a substance used in data transformation in an assay
Reference is an object (obj) in a plate and used as an internal assay control

One main calculation definition

IDBS model 205 (4 parameter logistic equation)



$$\%E = 100 \text{ X } \frac{(OBJ - MIN\_E)}{(MAX\_E - MIN\_E)}$$

Master Templates

# of Signals	Assay Category	Assay Technology	Master Template (ADA Calc Templates)
1	• Absorbance • Fluorescence • Luminescence • Chemiluminescence • ELECTROCHEMICAL LUMINISCENCE • HCS • Cell proliferation	• Any OD • FI • ViaLight • SAP ADP-Glo • PathHunter • FLIPR	i. 1S_STD_BK_QC ii. 1S_STD_BK_QC_Backfit
2	• Time-Resolved Fluorescence Resonance Energy Transfer Assay • HCS	• HTRF • ToxBLAzer(BLA) • Lance • Lantha-binding • IMAP-TR_FRET • Lance-Ulight/ SureLight • Lance-cAMP • AlphaScreening	i. 2S_STD_BK_QC ii. 2S_STD_QC_BK_Backfit
3	• HCS • Functional • Other	• FACS • INCELL	i. 3S_STD_BK_QC ii. 3S_STD_BK_QC_Backfit
4	• HCS • Functional • Other	• FACS • INCELL	i. 4S_STD_BK_QC ii. 4S_STD_BK_QC_Backfit
n	• HCS • Kinetics • Functional • Other	• INCELL • FLIPR	i. nS_Series_Analysis_Fitting ii. nS_Series_Analysis_Backfit To be completed

5 other specialty templates have been created

Business rules

Basic rules prevent unrealistic curve fitting

- Limit the range of fitting parameters
  - Maximum: 70 to 130%, minimum: -30 to 30%, slope 0.5 to 3.0
- Override EC50 when outside of tested concentration range
  - EC50 greater than [Max] or less than [Min]
- Apply to all assays
  - Science driven exceptions are allowed

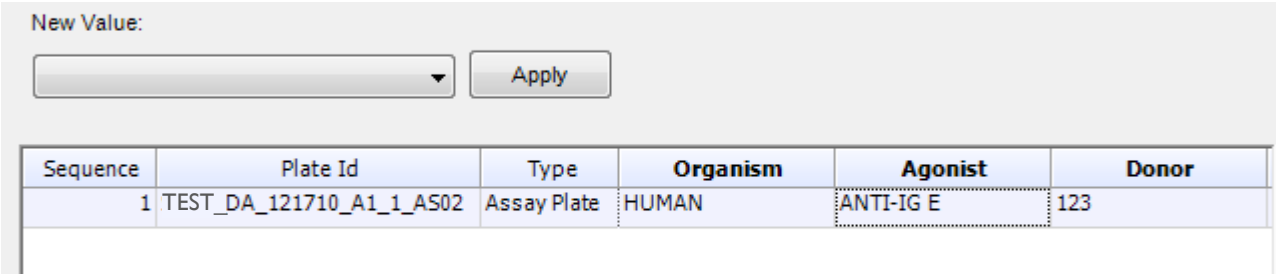
Advanced rules facilitate automatic fit to minimize human QC

- Pre-defined for individual assays, settings tracked in a separate DB
- Deal with more complex fitting issues
  - Caused by biology or specific assay or compound series problems which create fitting issues
- Can be turned on or off in a testset
- Modify or adjust in XERunner if necessary

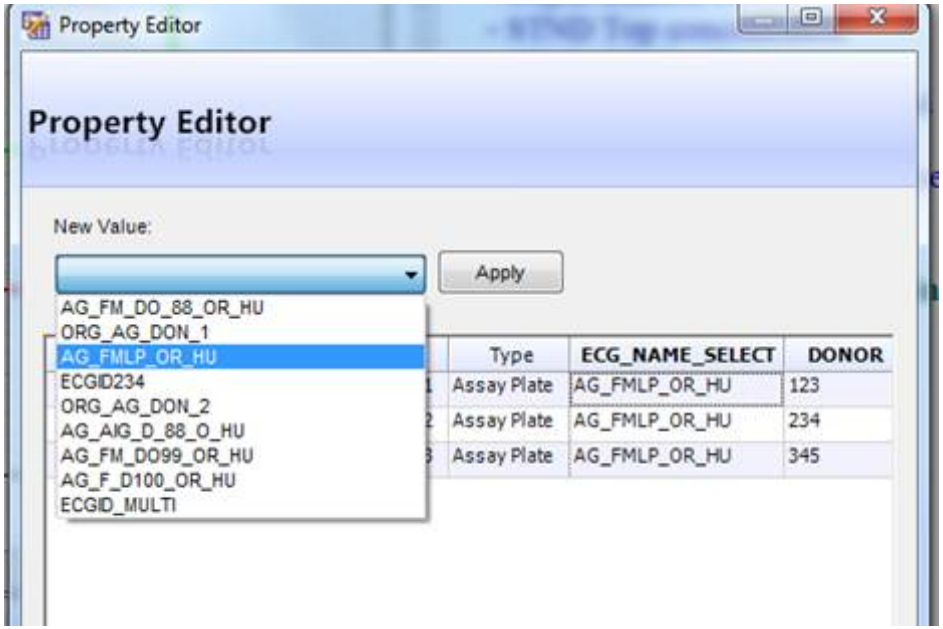
- These rules do not always work; manual inspection still needed

Experiment Condition Grouping_ID or ECG_ID

Original legacy software function for easy selection of summary variant values



Using the ECG_ID, users only need to select one entry: ECG_ID



- ✓ Potential for fewer protocols
- ✓ Fewer templates
- ✓ Simplified testset automation

Values of assay conditions will be displayed in XERunner when you perform the data analysis

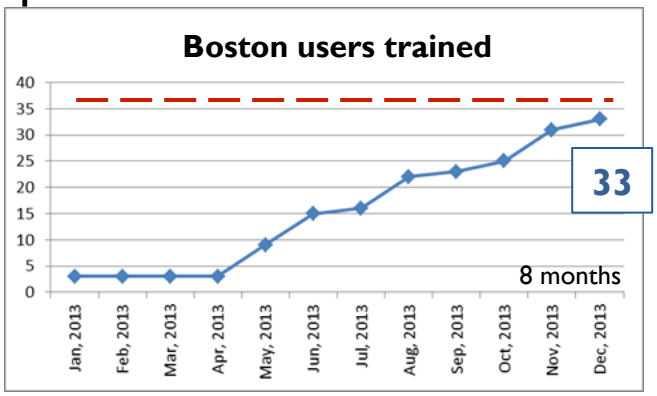
Scott Mosser & team

Pilot at Boston

- Master template creation and testing
- Connections to compound management database (mechanical propagation of well information)
- Data workup
- Data upload to corporate repository
- Training, one-on-one with real data on your assay
- Learning curve for early adopters

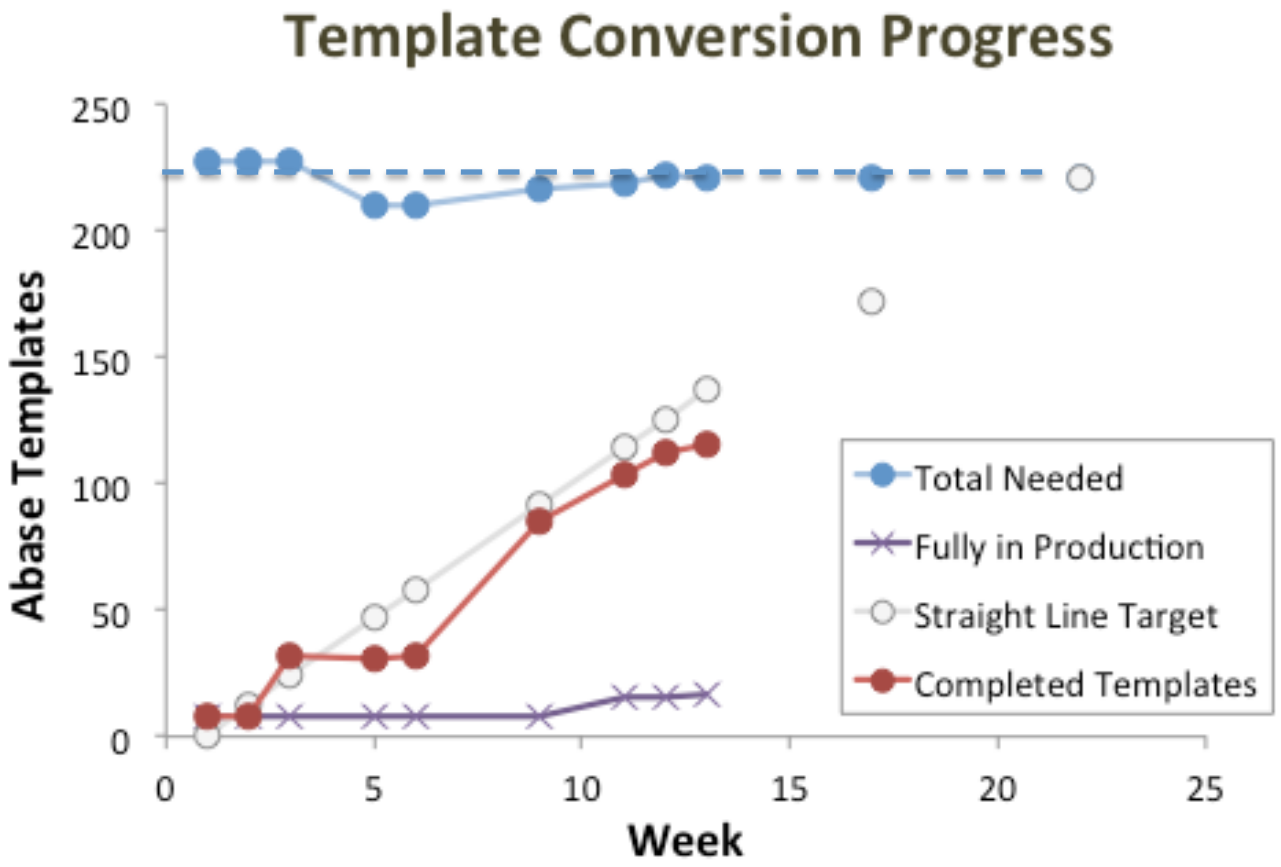
Boston Pilot

- Thirty-three end users trained by one super user (2hr group class session, plus an average of about 2-3hr hands on help for each)
- Additional advanced topics training
- Data analysis standards wiki pages continually updated
- User guide and FAQ document prepared based on user experiences.
  - How to go back and reanalyze my data, how to use advanced curve fitting tools, etc.
- Boston template conversions are essentially complete
  - A total of 22 ActivityBase protocols or templates have been created to date.
  - 13 have generated data for uploading
  - Five were put on hold as the assays were either terminated or transferred.
  - Six being validated currently
  - Three abandoned / never used.



Expanding globally: Conversions of existing assays across the Merck internal and external networks

March 2014 start



Results to date

- Data analysis standardized
- Template creation and maintenance much simpler now
  - Fewer than 20 master templates
  - Under 40 plate formats
  - Templates take 33-50% less time to create (was 2-6 hr / template)
- Making minor changes in calculations no longer requires creating a new template.
  - Saves lots of time by minimizing duplication of effort, preventing delays in data analysis.
- Total time saved for the IT support team is 5-10 hours per week, based on a steady state average of 5 templates a week (current workload is ~ 10 per week).
- Additional 5-6 hours saved per month from flexibility in changing calculations.

For scientists, preliminary results show 30% less time required for data workup for well-behaved assays, but there are still a lot of assays to bring online.

Significant indirect time savings (e.g. assay transfers) are also expected.

Remaining Issues

- Other sites have different compound management and assay LIMS systems and larger groups to train
- Gaining enough experience for scientist to work completely independently takes time
- Culture of hands-on curve fit tweaking is still pervasive (slows analysis considerably)
- Automated curve fitting results are still variable, and expertise is not wide spread

Lessons Learned

- Talk to people early and often! Update status regularly.
- Build awareness, make the case for change.
- Ensure that all key stakeholders are informed and engaged. Repeat & reinforce your message.
- Listen to issues/concerns! (don't make assumptions when holding conversations)
- **Start small & iterate!** Do a pilot if the scale of your change is large.
- It's easy to make superficial technical progress without really changing people's behavior or mindset.
- When you think there's a conflict, tackle it head on – use one-on-one phone calls. They're much better than e-mails and an order of magnitude better than teleconferences.
- Collecting information and standards in one easily accessible location really saves time & keeps people from reinventing the wheel.

Acknowledgements

Dongyu Sun
Charles Fleming
Heena Chauhan
Scott Mosser
David Pechter
Stephen Day
Gordon Inamine
Bharti Gajera
Tom Harkins
Eric Anthony Williams
Mala Kansara

Mike Benjamino (IDBS)
J.T. Kaminski (IDBS)
Mary Jo Wildey
Matthew Sills
Dominic Joseph Maratt
Albert Aguero
Alexander Szewczak
Jonathan Connick
Bruce Beutel

Many others, especially early adopters!