
Too Good to be True?

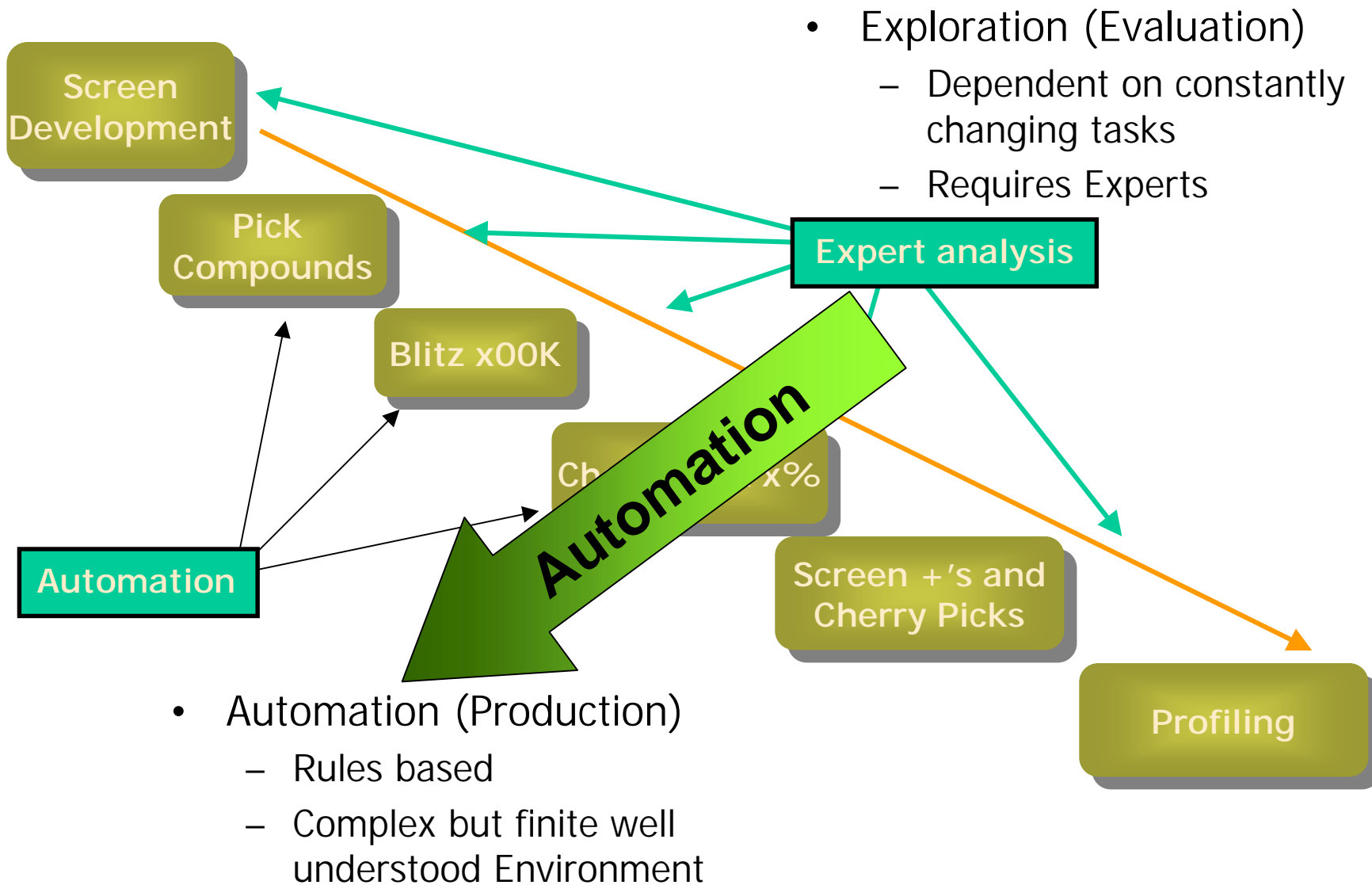
Lessons Learned During the
Evaluation of PipeLine Pilot at
BMS

Jay Gill

Wallingford Data Technologies

Discovery Technologies

Automation of Information in HTS



Why is it Important to Capture Process?

- Standardization - Distribute the experts' knowledge - maximize your time to innovate in your domain, Reduce uncertainty
- Measurement - Once a process is defined it can be measured
- Optimization - Once a process is defined and can be measured, it can be optimized

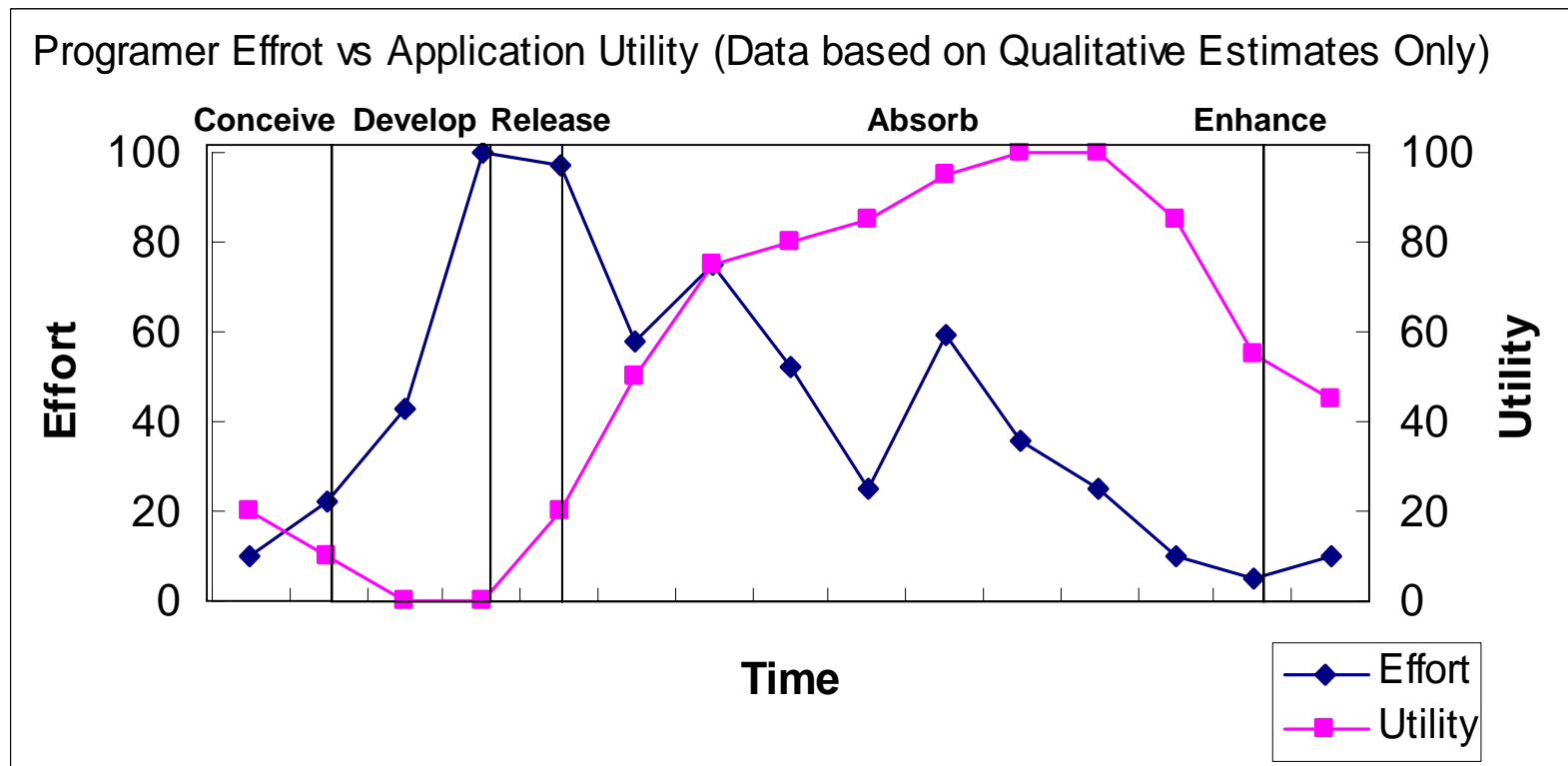
Software development offers one of the best tools to help define and standardize process

The Problem

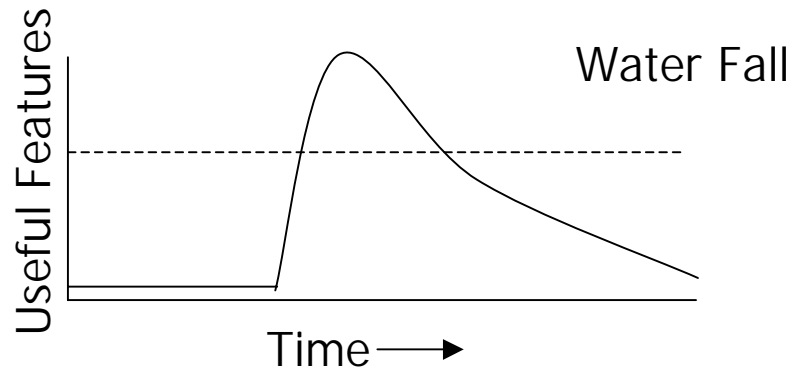
- Not enough developers.
- Developers aren't the right people.
 - Software professionals know software but not the domain.
- The problem keeps changing
“if we knew what we were doing it wouldn't be called research” - A. Einstein
- What To Do?

Potential Solutions

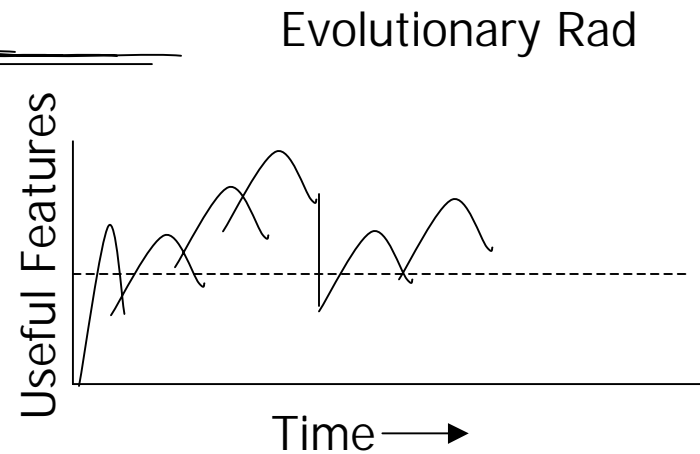
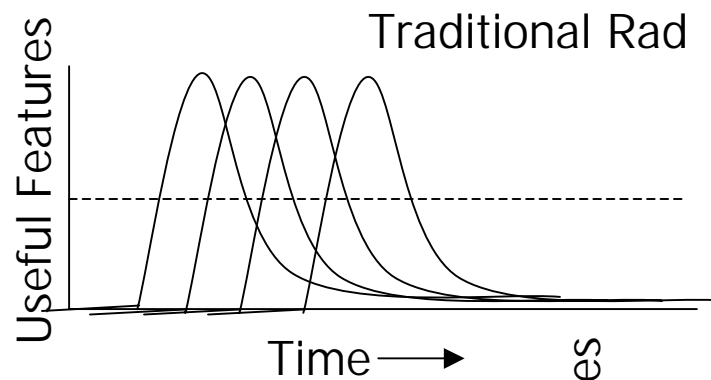
- Develop several “killer” apps
 - become too general
 - take too long
 - are too complex/costly



Develop Lots of Targeted Solutions (Rapid Application Development)



Goal- Increase the area under the Utility curve

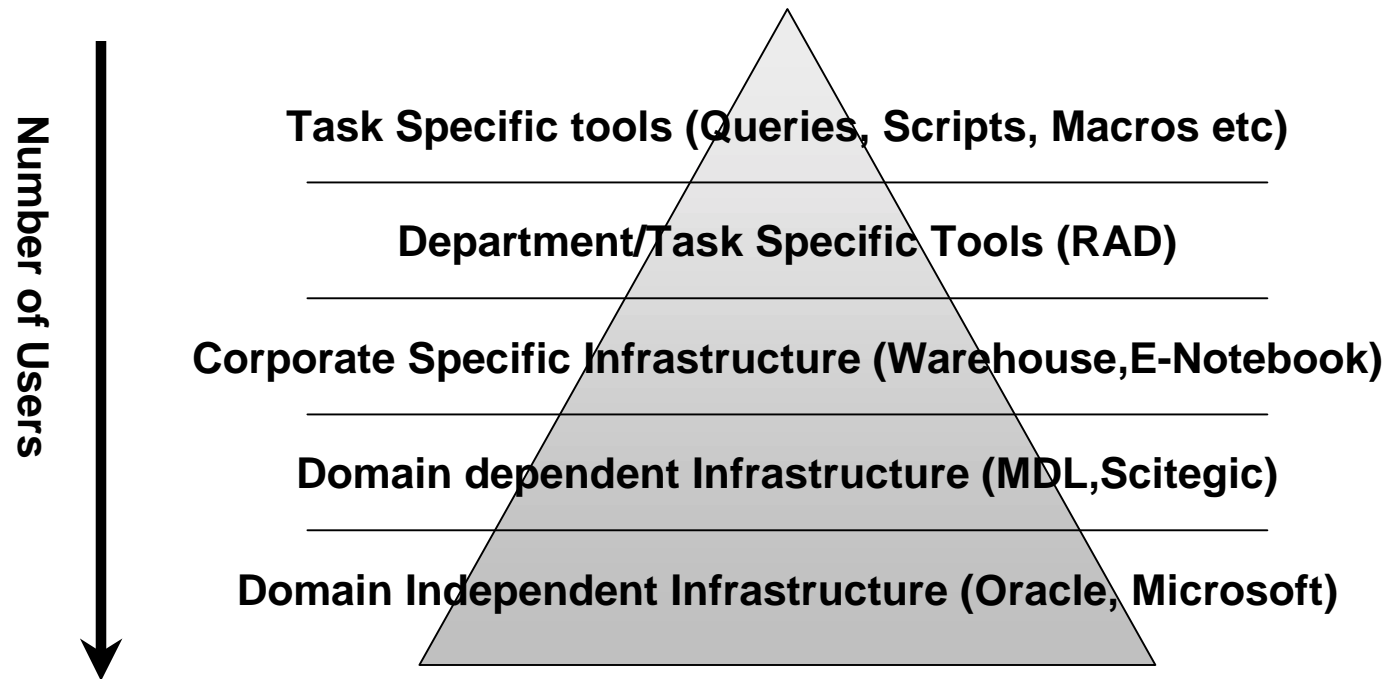


Approaches to RAD

- Work Harder - ???
- Work Smarter - ???
- Get others to do the work for you!

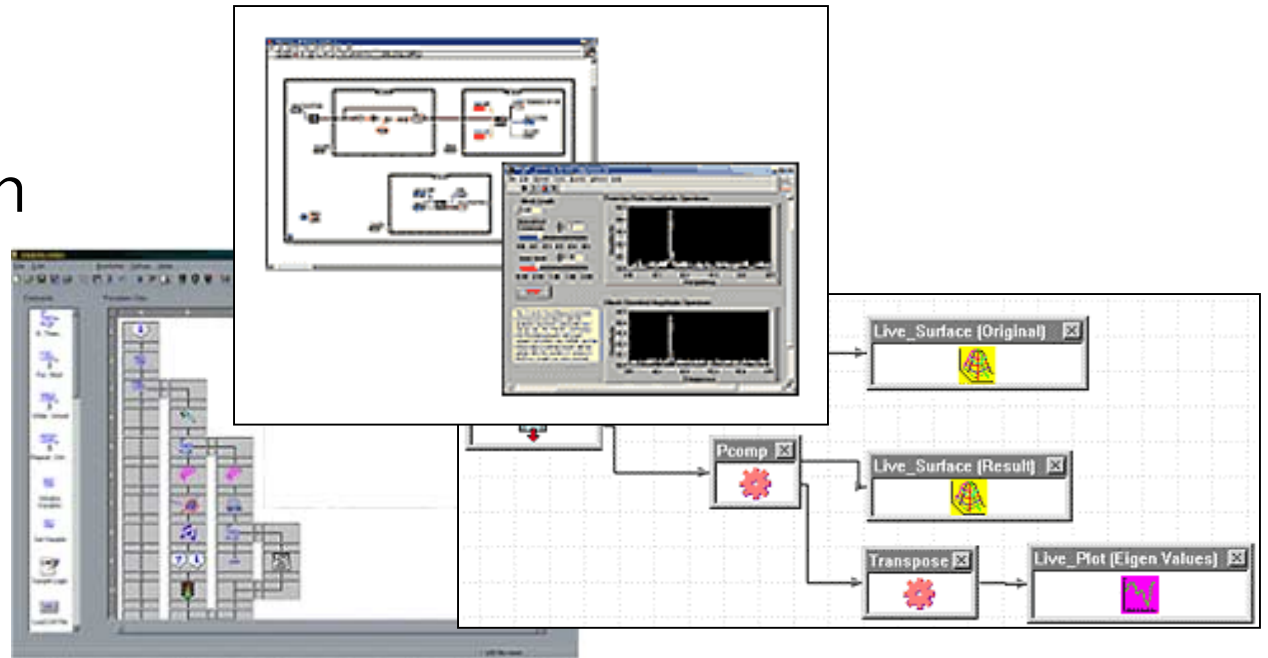
End user programming

- Scientists develop solutions
- RAD developers refine solutions, build the tools
- Programmers “mine” Rad apps for general purpose apps and deploy broadly



Graphical Programming

- Users can connect the dots, but shouldn't care about
 - Flow control
 - Exception Handling
 - Technical details (oracle access, SQL)
 - Needs rapid iteration (test & refine)
- Automation:
 - Lab View,
 - White Carbon
 - Overload
- Analysis
 - IDL



Specific Requirements

- Scalable Approach
 - C++ to Script to End User
 - Integrate with existing and future tools (open - extensible)
- Easy to use
- Deliver Cheminformatic toolset for both the CADD focus and the End User
- Fast
- Scalable Architecture for wide deployment
 - clear path from interface to application
 - Deployment Strategy

Why Evaluate Scitegic?

- In house build - not practical
- Domain Focus
- Molecular Knowledge
- Out of the box tool kit
- Recommendations from other users
- Open approach - easy to extend and customize

Issues: Too many communities

- HTS SAR Analysis
- Library Design
- Substance performance Monitoring
 - Library
 - Compound Acquisition
- Standard Compound/Screening metrics for overall view and optimization
- Developers
- Scientists
- CADD
- Existing large in house projects (Warehouse)

What we did

- Formed Teams
 - HTS Scientists
 - CADD/ChemoInformatics
 - Infrastructure Integration
 - Developers
- All received 2 days INTENSE training
- All were asked to review features which would be useful to their community and assigned specific tasks

What needed to be evaluated?

- What is the quality of the chemistry tools?
- How will the development environment play?
 - With developers?
 - With key stakeholders?
 - Will End Users Adopt?
- Ability to handle large data sets
- Integration with current environment
 - Structures, Data

Developers: Technology tests

- Integration
 - With Oracle - trivial to connect to ODBC sources - more complex to provide generic tools
 - With ISIS & Structures
 - Easiest (fastest) via flat files.
 - Isis worked but required help
 - Haven't moved to ISIS Direct
 - Excel
 - Simple using VBA/VBScript (wouldn't be critical IF we were a Spotfire shop)

Developers: Technology Tests Cont.

- Extensibility
 - Via SOAP and URL's
 - Simple Soap Server (out of the box)
 - Wrote URL “scraper” with help from Scitegic
 - With HTML for Advanced UI development
 - Scitegic helped demonstrate/build a skeleton HTML page display
Deployment

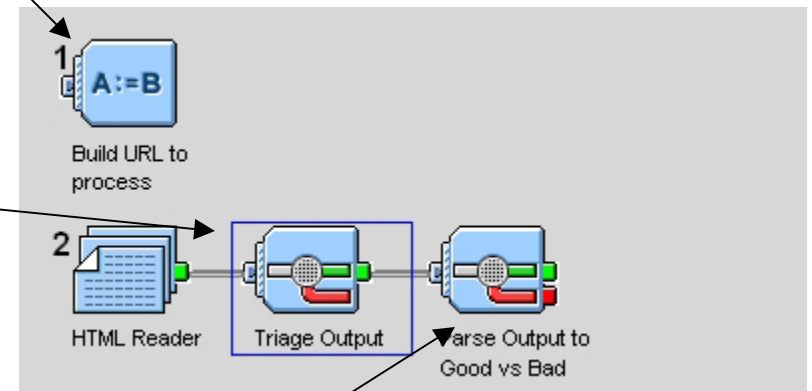
URL "Scraper"

- URL takes a list of BMS#'s in and returns either "Good/Bad"

```
/* Construct URL */  
StartURL := 'http://aurl/annotate.cgi?id=';  
EndURL := '&db=bms&an=' . 'Terminate';  
@AnnotateURL := StartURL . BMSNumber;  
@AnnotateURL := @AnnotateURL . EndURL;
```

```
stringlen(text) > 1 AND  
findsubstring(text, 'Untitled') < 1 AND  
findsubstring(text, 'Text') < 1;
```

```
/* parameter -"ID tab Good/Bad" -Good = 1, bad = 0 */  
r := rsplit(text, '\s');  
Good_Bad := r[2];  
BMSNumber := r[1];  
remove('r', 'text');  
/* Pass records = 1 , fail=0 */  
Good_bad;
```



Developers: Infrastructure

- Deployment
 - Via Developer seat - works fine (\$\$\$)
 - Via Web Interface - Excellent!!!

sci-teg Pipeline Pilot Web Access

sci-teg Pipeline Pilot Web Access

Login Protocols Execution Results

The Pipeline Pilot folders accessible from the web are listed below. Double-click a folder name to go into that folder. Double-click the [...] item to go up one directory level. (If the list of folders is empty, you are in the root folder.)

Click on a protocol to see its description, then click the Execution tab to input parameters and run the protocol or the Trash icon to discard the results of a completed protocol or the Trash icon to discard the results of a completed protocol. Just save your results files elsewhere to preserve them.

Folders: Current folder: Protocols/Web Services/NewLeads

- [...]
- Development
- Lead Sources

Protocols:

- DemoWebService
- HTS Primary Data Analysis
- HTS QC and Analysis
- ListInList
- ListInList
- TestMarksTools

Start Time	Status	Elapsed Time
10:54:22 PM	Finished	0:08

and usage to describe the functioning of their derived protocol.

Infrastructure/Support

- Installation very simple despite Novell
- Management of a few users simple (new features coming)
- Scaling ?? Seems ok in other companies

Cheminformatics and CADD

- Evaluate Chemistry Tools
 - Very impressed with speed
 - Successfully utilized learning tools.
 - Steep learning curve
 - Are currently focusing more closely on the tools.

Timing FAST!

- 1.69M structures read smiles from file, convert to Structure, compute Mol.Wgt, plot histogram (3min 20sec)
- Retrieve HTS data - bin (450K ~3min)
- Complete analysis of HTS data (14Min)

HTS Scientists

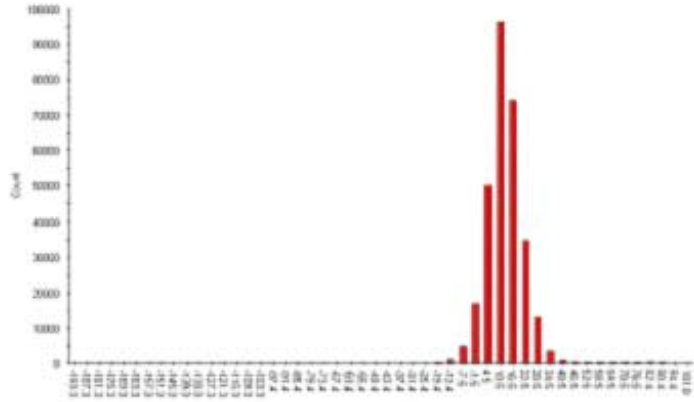
- Test to see if an end user could develop an application in PP.
 - Choose 1 task (QC of HTS Data)
 - Training
 - 2 Days training
 - 3 Days learning
 - 1-2 days IT support mostly to build tools (Oracle integration)
 - Outstanding Scitegic support

HTS Scientists

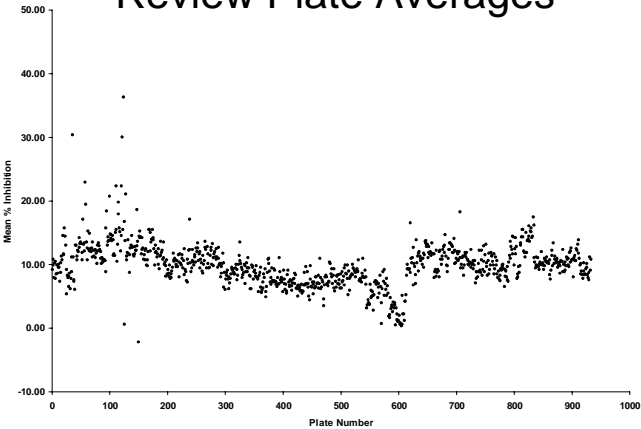
- Choose to build a pipeline to analyze post screen data
 - Test for systematic error
 - Set Hit threshold (interactive)
 - Apply Exclusion lists
 - Generate “Hit List” for the chemists

Tests for Systematic Error

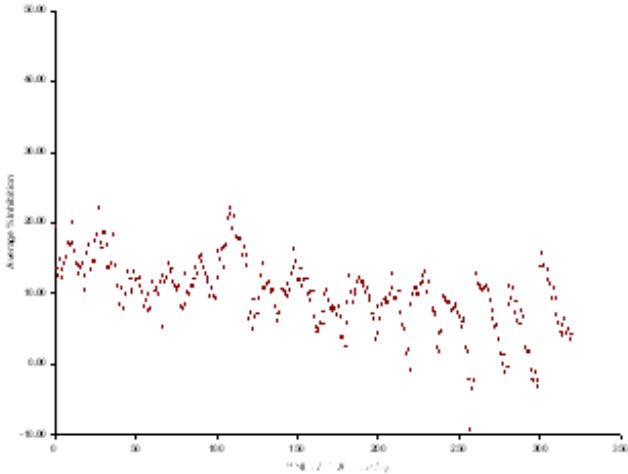
Distribution of Activity



Review Plate Averages



Review Well Averages

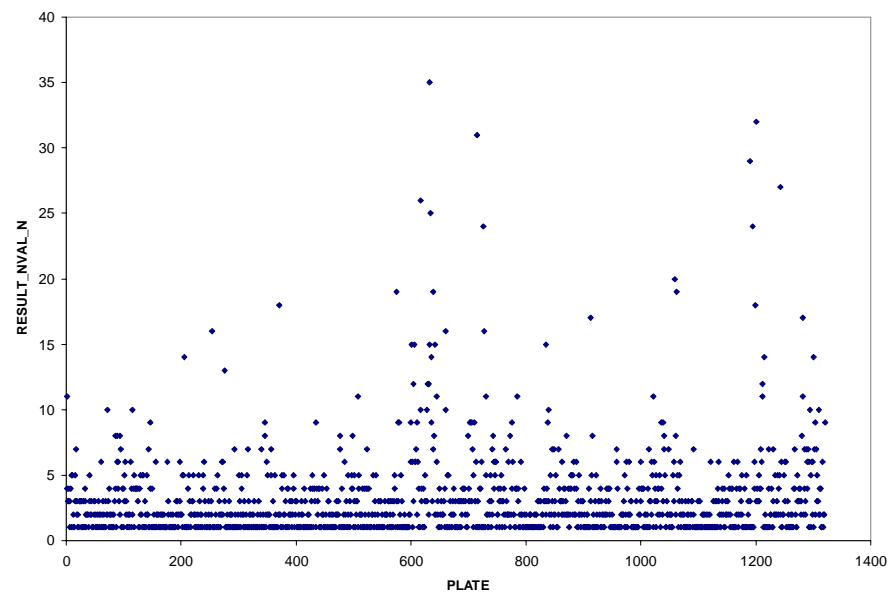


Determine Hits (non-cheminformatically)

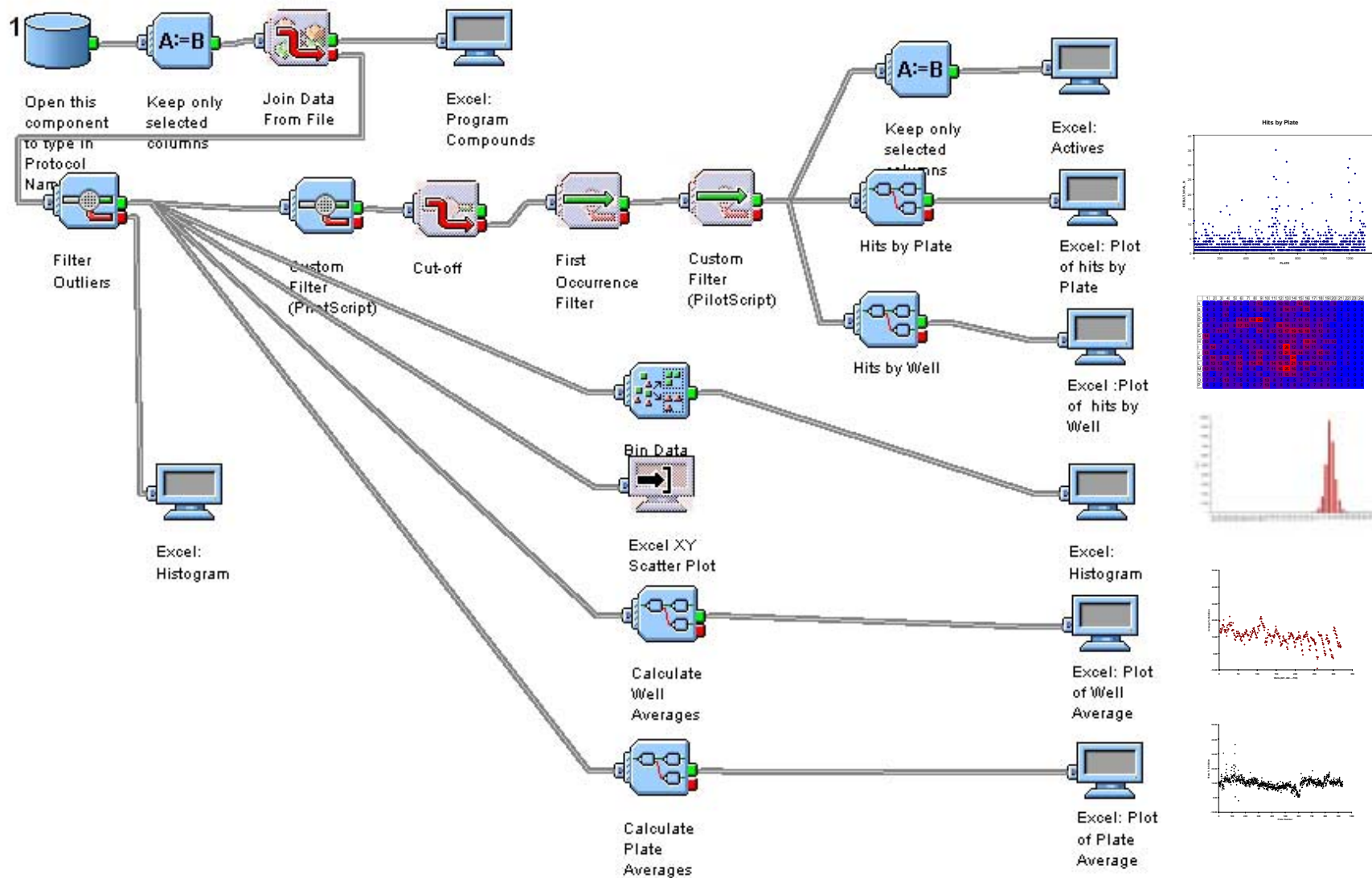
Hits By Well

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
A	2	4	5	11	6	6	3	7	13	3	4	10	10	7	14	10	4	6	3	8	3	0	0	0
B	4	5	3	8	3	6	1	6	9	7	5	8	14	11	8	15	3	6	4	5	2	0	0	0
C	4	3	3	10	2	4	3	3	5	0	5	8	8	7	4	5	4	5	4	6	1	3	0	0
D	3	7	4	5	4	14	11	18	26	10	5	4	10	7	11	11	6	9	7	6	1	2	0	0
E	7	6	6	11	9	17	15	11	10	6	7	18	14	11	10	12	7	11	5	1	3	0	0	0
F	5	7	11	11	6	9	7	3	14	8	12	13	17	19	16	18	10	12	8	5	2	3	0	0
G	14	4	4	3	7	6	6	4	5	8	3	10	9	13	7	9	10	9	4	3	2	2	0	0
H	10	4	4	9	3	4	9	9	5	6	8	10	13	14	7	19	14	7	11	10	1	3	0	0
I	8	14	7	6	7	5	5	9	8	6	6	12	26	12	8	14	11	9	4	4	3	2	0	0
J	13	5	5	4	5	7	4	10	15	9	8	13	21	16	14	10	9	15	10	6	1	5	0	0
K	8	14	9	15	8	14	4	8	14	5	7	12	10	24	4	6	10	10	5	5	2	4	0	0
L	8	13	13	6	10	15	8	14	18	9	14	16	15	21	10	16	12	13	11	10	1	5	0	0
M	12	10	12	8	7	14	4	8	3	7	11	16	25	16	6	10	3	8	9	4	2	1	0	0
N	7	2	7	8	6	10	5	5	9	3	7	11	15	14	9	5	10	7	4	5	2	1	0	0
O	7	7	5	13	7	5	6	3	5	13	8	7	4	5	7	7	4	5	7	3	1	3	0	0
P	6	2	2	8	7	9	2	6	7	13	4	8	5	8	6	4	7	8	8	4	3	3	0	0

Hits by Plate



HTS Data Analysis Pipeline



End User Test - Results

- End user did the work!
- Took a 4 day process to 14 minutes (for a 750K screen
 - Allows “what if” analysis
 - Ease of use encourages acceptance
- End user became major champion
- We’re now “tidying” up and will share via the web interface.

Conclusions

- Well we bought it
 - Deployed to a limited group rather than the whole enchilada
 - Experience since purchase even more positive as we learn more
- Internal Champions Critical both IT and Scientist
 - Scientists must commit time/energy
- Requires a clearly defined Business need which should be emphasized during evaluation.
 - Provides focus/clarity
- Applications must be feasible with clear impact.
 - Need a clear “bottom line ROI”

Where we will deploy PP

- Analysis of Deck Architecture
 - Regular Report on deck structure
- Complete HTS analysis workflow
- Inbound Compound Processing
- General Chemoinformatics Processing

Acknowledgements

- BMS End Users: Ramesh Padmanabha, Mark Hermsmeier, Malcolm Davis, Brad Pearce
- BMS Management: Alastair Binnie, Mike Poss, Ed Petrillo
- Infrastructure: Dan Maloney, Bill Karpowitz, Nelly Masias
- Scitegic: Chris Farmer, Charlie Weatherall, Andrea Barry