



Helium: A Data Driven User Tool for SAR Analysis

May 17th 2011
Karen Worsfold

Today's presentation – Questions to be answered

- Why did GSK undertake development of Helium?
- How was the Helium idea conceived?
- How did GSK develop Helium to ensure user satisfaction and acceptance?
- What is the architecture of Helium?
- What functionality does Helium provide?
- What synergies does GSK see with combining Excel, JChem for Excel and Helium?
- What is Helium's current status and future plans?

Where we came from

IT Resources Supporting SAR in Discovery

90's

00's

Today

Spotfire[®]

CPT

InforSense

ChemRetriever 3.0

DRUG DISCOVERY
INNOVATION & EXCELLENCE

Chemically Aware Spreadsheet

Discovery Explorer

AutoStructure

accelrys[®]

GSK CPDP
Compound Profiling Data Portal

ISISNET

MDL ISIS/Base

Where we are headed

IT Resources Supporting SAR in Discovery

90's

00's

Today

Future?



Excel 2007



Jchem for Excel



Instant JChem

The Birth of 'Helium'

- One stop shopping for all of your SAR data needs
- Giving the user control over what data they wanted rather than needing to know where the data was stored
- First prototype was a blank screen.....users had too much flexibility and didn't know where to start!
- Next incarnation was integrated with Tibco Spotfire
 - Helium to gather the data and Spotfire to view the data in multiple ways (tables, scatterplots, bar charts etc)
 - Spotfire had the ability for plug-in panels to be created eg a Structure Viewer
 - Spotfire supported large data sets
 - At the time, cost was not a significant factor

The 'Cold Turkey'

- Using Agile Development Techniques we ran a Proof of Concept to see if it could also include a forms view allowing us to replace ISIS Base
 - Gave to Discovery scientists and asked them to stop using their legacy tools and just use Helium with Spotfire.
 - As they identified a task they needed to do, we developed the Helium feature to enable it
 - Significant Improvement was made in usability
 - Data retrieval functionality and usability of Helium were received 'Very Positively'
 - Less positive feedback with the attempt to integrate a "forms" view
 - Still received resistance from the average bench chemist due to Spotfire

What next for 'Helium'?

- Decision to provide a simple grid view tool, a more complex visualisation tool and a forms view tool
 - **Helium in Excel**
 - **Helium in Tibco Spotfire**
 - **Instant Jchem**
- Decision on what tool to use was based on what they wanted to do with the data and not what data they needed

'Helium', the Sequel

- In 2009, GSK purchases the Chemaxon Suite of tools
 - JChem for Excel
 - Instant JChem
 - JChem cartridge
- And so Helium was to be integrated with in Excel and use JChem for Excel API to manage structures

Benefits

- JChem for Excel enhances a familiar and comfortable tool
- Instant JChem provides flexible forms tool
- JChem Cartridge underpins our chemistry web services

The screenshot shows the Microsoft Excel interface with the JChem for Excel ribbon active. The spreadsheet contains the following data:

External Id	Structure	Database
MFC00003777		acd
MFC00003784		acd
MFC00039585		acd
MFC00041457		acd
MFC00044428		acd

The Helium Task Panel on the right includes the following options:

- Calculate Derived Properties
- Calculate Simple Properties
- Fragment R-groups
- Get Compound Number for Structure
- Get Similar Compounds
- Remove Isotopes
- Remove Salts / Solvates
- Remove Stereochemistry
- Search Sub-structure in GSK databases
- Standardise Valences
- Sub-structure Search within table

The Development Approach – Keys to Success

Internally

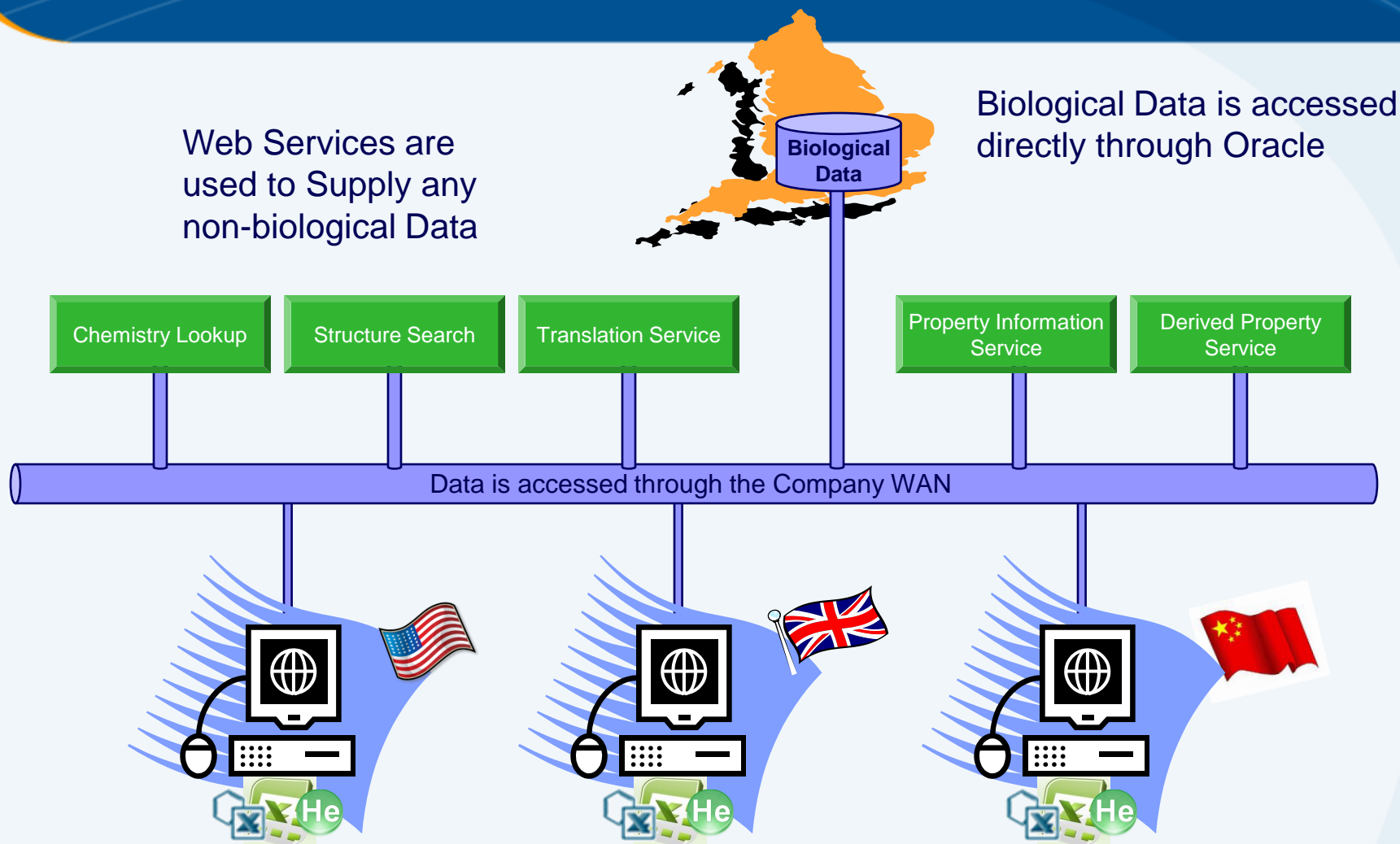
- User interaction and Ownership
- Identified Lead End User (Senior Researcher from Discovery)
- Interactive End User Group covering all disciplines and sites within R&D
- Agile development approach with regular deliverables
- Weekly End User Group meetings
- Extended End Users engaged at key points
- Accessed “Live” data, making the tool immediately useful
- ‘Viral’ release of the software to R&D

Externally

- Superb support and response from ChemAxon
- Collaborative relationship in solving problems
- Excellent communication between the two companies



Helium Architecture



- ❖ Clients can be in any R&D location, and can be from many different disciplines (Biology, Chemistry, Computational Chemistry, Compound Management)
- ❖ Dataset sizes can range from tens of compounds or structures to tens of thousands

Keeping Helium up to date

- In Excel

- Microsoft ClickOnce for managing the customization
 - Client code regularly checks for updates and installs
 - Insures that update installation is simple
 - Keeps installations of the software from “falling behind”

- In Spotfire

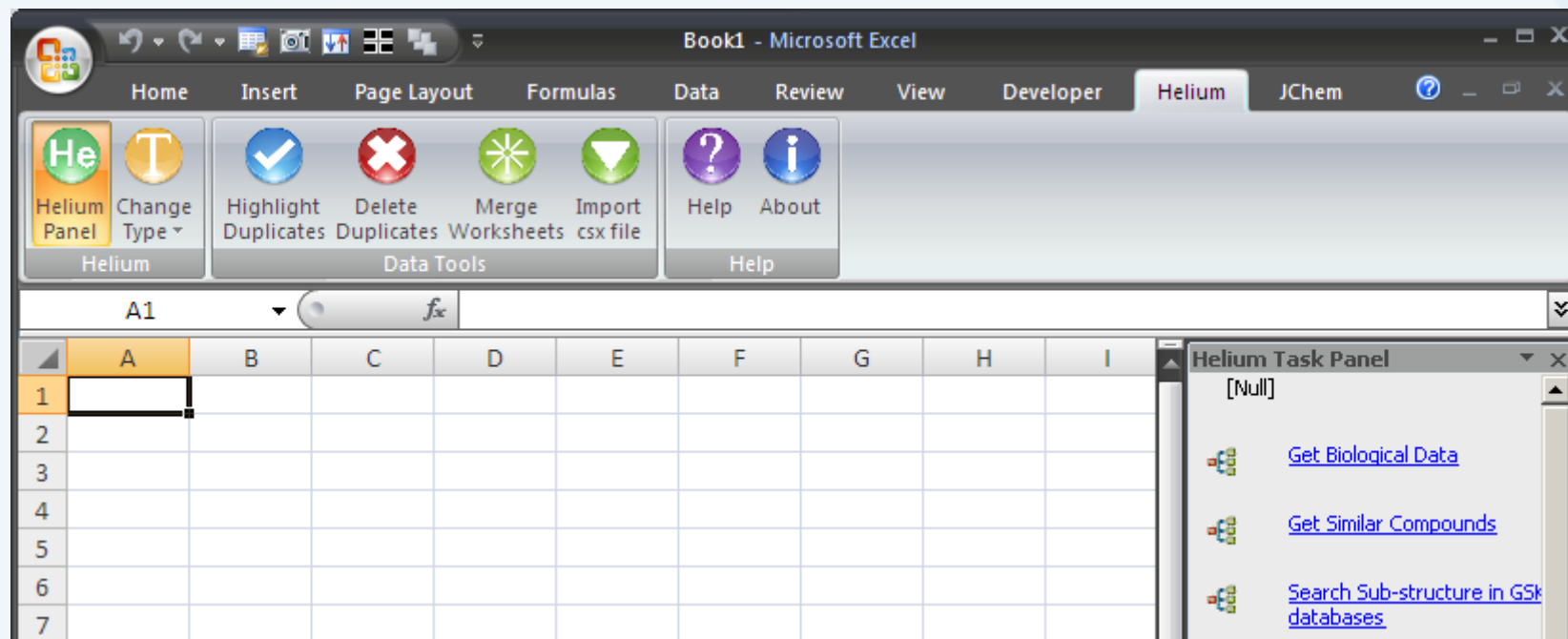
- Each log-on to the server checks for updates and prompts user to install

The components of Helium

❖ Microsoft Excel 2007

❖ JChem for Excel

❖ Helium



- General functionality for formatting, sorting, presentation and creating equations
- Chemical Structure presentation, file import and export, etc.
- Access to GSK web services, GSK biological data and consistency with legacy data

Helium User Interface

The screenshot shows the Helium User Interface integrated into Microsoft Excel. The ribbon includes the following groups:

- Helium Panel:** Helium, Change Type
- Data Tools:** Highlight Duplicates, Delete Duplicates, Merge Worksheets, Import csx file
- Help:** Help, About

The worksheet contains the following data:

External Id	Structure	Database
MFCD00003777	<chem>C1=CC=C2C=CC=CC2=C1</chem>	acd
MFCD00003784	<chem>O=C1C=CC(=C1)C2=CC=CC=C2</chem>	acd
MFCD00039585	<chem>CC1=CC=C2C=CC=CC2=C1</chem>	acd
MFCD00041457	<chem>C1=CC=C2C(=C1)OC2</chem>	acd

The Helium Task Panel on the right displays the following tasks:

- Calculate Derived Properties
- Calculate Simple Properties
- Fragment R-groups
- Get Compound Number for Structure
- Get Similar Compounds
- Remove Isotopes
- Remove Salts / Solvates
- Remove Stereochemistry
- Search Sub-structure in GSK databases
- Standardise Valences
- Sub-structure Search within table

Helium Ribbon – for non data specific tasks

JChem Ribbon to access JC4XL Functionality

Datatype Sensitive Task Panel. This display appears when a "structure" data type is selected,

Worksheet with datatyped Helium Data

Helium Functionality

The screenshot shows the Microsoft Excel interface with the Helium Panel ribbon selected. The ribbon includes buttons for Helium Panel, Change Type, Highlight Duplicates, Delete Duplicates, Merge Worksheets, Import csx file, Help, and About. A dropdown menu is open for the 'Change Type' button, listing the following options: Compound Number, External Id, Null, Parent Compound Number, Prep Lnb Ref, Project Id, SMILES, Structure (which is selected with a green checkmark), and User Id. The background shows a spreadsheet with a chemical structure in cell B2 and the text 'Database' in cell C2.

- Datotyping assigned by regular expressions
- Generic data types(e.g. Project Id) can be manually set

Helium Functionality List

Data Type	Task
Project ID	<i>Get Compound Numbers</i>
Project ID	<i>Validate Project ID</i>
Compound Number	<i>Get Biological Data</i>
Compound Number	<i>Get Critical Results</i>
Compound Number	<i>Get LNB Refs</i>
Compound Number	<i>Get Parent</i>
Compound Number	<i>Get SMILES</i>
Compound Number	<i>Get Structure</i>
Compound Number	<i>Get Structure for Parent Compound</i>
Compound Number	<i>Get Versions</i>
Compound Number	<i>Identify duplicate structures</i>
Compound Number	<i>Validate Compounds</i>
SMILES	<i>Calculate Derived Properties</i>
SMILES	<i>Calculate Simple Properties</i>
SMILES	<i>Canonicalize</i>
SMILES	<i>Get Exact Structures</i>
SMILES	<i>Get Similar Compounds</i>
SMILES	<i>Remove Isotopes</i>
SMILES	<i>Remove Salts / Solvates</i>
SMILES	<i>Remove Stereochemistry</i>
SMILES	<i>Search Sub-structure in GSK databases</i>
SMILES	<i>Standardize Valences</i>
SMILES	<i>Sub-structure search within table</i>
External ID	<i>Get SMILES</i>
External ID	<i>Get Structure</i>

Data Type	Task
<i>Lnb Ref</i>	<i>Get Compounds Numbers</i>
Lnb Ref	<i>Get Registration Information</i>
Lnb Ref	<i>Get SMILES</i>
Lnb Ref	<i>Get Structure</i>
Lnb Ref	<i>Get Versions</i>
Lnb Ref	<i>Validate LnbRefs</i>
User Id	<i>Get Registered Compounds LNB Refs</i>
Structure	<i>Calculate Derived Properties</i>
Structure	<i>Calculate Simple Properties</i>
Structure	<i>Get Exact Structures</i>
Structure	<i>Get Similar Compounds</i>
Structure	<i>Remove Isotopes</i>
Structure	<i>Remove Salts / Solvates</i>
Structure	<i>Remove Stereochemistry</i>
Structure	<i>Sub-structure Search in GSK databases</i>
Structure	<i>Standardize Valences</i>
Structure	<i>Sub-structure Search within table</i>
Parent Compound Number	<i>Get Parent</i>
Parent Compound Number	<i>Get SMILES</i>
Parent Compound Number	<i>Get Structures</i>
Parent Compound Number	<i>Get Versions</i>
Parent Compound Number	<i>Validate Compounds</i>
Null	<i>Get Biological Data</i>
Null	<i>Get Similar Compounds</i>
Null	<i>Search Sub-structure in GSK databases</i>

Helium Performance (worst case conditions)

Activity	Time
Retrieve 33,885 Compound Numbers from 110 Project Ids	28"
Retrieve SMILES from 33,885 Compound Numbers	5' 12"
Retrieve Structures from 33,885 Compound Numbers	26' 28"
Canonicalize 33,885 Structures	8' 05"
Remove Salts and Solvates from 33,885 Structures	9' 35"
Remove Stereochemistry from 33,885 Structures	10' 05"
Get Lot Ids from 33,885 Compound Numbers	11' 40"
Validate 33,885 Compound Numbers	11' 38"
Check for Duplicate Structures for 33,885 Compound Numbers	9' 12"
Retrieve Parent Compound Number for 33,885 Compound Numbers	8' 30"
Retrieve MF, MW, BEW for 33,885 Compound Numbers	37' 47"

*This performance level represents a significant improvement over previous tools!
Typical SAR datasets are much smaller*

Helium, JChem and Excel – the Possibilities

- With the various tools available within Excel 2007, Helium and JChem for Excel, the researcher has an extensive, flexible tool for interrogating data:

The screenshot shows the Microsoft Excel 2007 interface. The ribbon includes tabs for Home, Insert, Page Layout, Formulas, Data, Review, View, Developer, Helium, and JChem. The 'Conditional Formatting' menu is open, showing options like Highlight Cells Rules, Top/Bottom Rules, Data Bars, Color Scale, and Icon Sets. The 'Color Scale' option is highlighted. The spreadsheet data is as follows:

	A	B	C	D	E	J	K	L	M
4	MFCD00039585	<chem>C1=CC=C2C=CC=CC12</chem>	-6.8	0.003	1				
5	MFCD00041457	<chem>C1=CC=C2C(=O)OC12</chem>	-3.8	0.003	1				
6	MFCD00044428	<chem>CC(=O)C1=CC=C2C=CC=C12</chem>	0.35	0.591	1				
7	MFCD00046319	<chem>C1=CC=C2C=CC=CC12</chem>	-2.8	0.003	2	0.252	20	3.258	140
8	MFCD00051614	<chem>C1=CC=C2C(=O)C=C1C2</chem>	-1	0.18	2	0.992	19	2.373	212
9	MFCD00057817	<chem>C1=CC=C2C=CC=CC12</chem>	6.3	0.591	2	1.283	27	3.157	288
10	MFCD00060877	<chem>C1=CC=C2C=CC=CC12</chem>	-5.2	0.003	1	0.192	13	4.408	141
11	MFCD00061044	<chem>C1=CC=C2C=CC=CC12</chem>	-6	0.003	1	0.196	12	3.889	132

Excel conditional formatting options give great highlighting options for identifying patterns in your data.

Helium, JChem and Excel – the Possibilities

- With the various tools available within Excel 2007, Helium and JChem for Excel, the researcher has an extensive, flexible tool for interrogating data:

External Id	Structure	Core found	R1	R2	R3	abe	alp	arir
MFC00072162		Yes	CN(C[*])CC[*]	[*]C1=CC=CC=	CN(C[*])CC[*]	30.8	0.919	2
MFC00083115		Yes	[*][H]	[*]CCC#N		-3.05	0.059	1
MFC00085585		Yes	[*]C1=CC=CC=	[*]C1=CC=CC=	[*]C1=CC=CC=	5.05	0.348	4

Easy column filtering

For Example after R-Group decomposition, you can convert structure to SMILES and then use the SMILES as filters in Excel to segregate your data

Helium, JChem and Excel - the Benefits

- Bringing new capabilities to be included in SAR
- Easily allows comparison of assay data vs. off target activity vs. liability
- A single, non-threatening interface to Discovery data
- Most researchers are familiar with Excel and its functionality
- Excel files are universally portable and support data interchange
- One application to support and maintain vs. the legacy list of multiple tools
- We benefit from the advancement of Excel in the future (Excel 2010)

Helium in Spotfire

Demo Structures and Helium.dxp - TIBCO Spotfire

File Edit View Insert Tools Help

Cover Page Table and Structure Viewer

Helium Data Type Panel

Active data table:
Data Table
CompoundNumber

ID
 Structure
 clogp
 clogp error code
 hba
 hbd

Helium Task Panel

Data Table

[CompoundNumber]

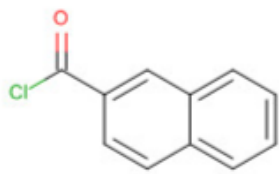
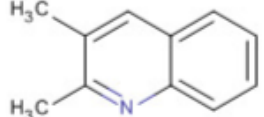
- Get All Versions
- Get Biological Data
- Get Compound Availability
- Get Critres: All Results
- Get Enhanced CritRes
- Get Parent
- Get Parent Smiles
- Get Prep LNB Refs
- Get registered MW, MF, and

ID	Structure	clogp	clogp er	Marking:
gsk1	C1=CC=CC=C1	2.14	0	<input checked="" type="checkbox"/> Marking
gsk2	C1=CC2=CC=CC=C2C=C1	3.32	0	<input type="checkbox"/> Marking
gsk3	CC1=CC2=CC=CC=C2N=C1C	2.98	0	<input type="checkbox"/> Marking
gsk4	ClC(=O)C1=CC2=CC=CC=C2C=C1	2.74	59	<input type="checkbox"/> Marking

Data table: Data Table

Structure Viewer

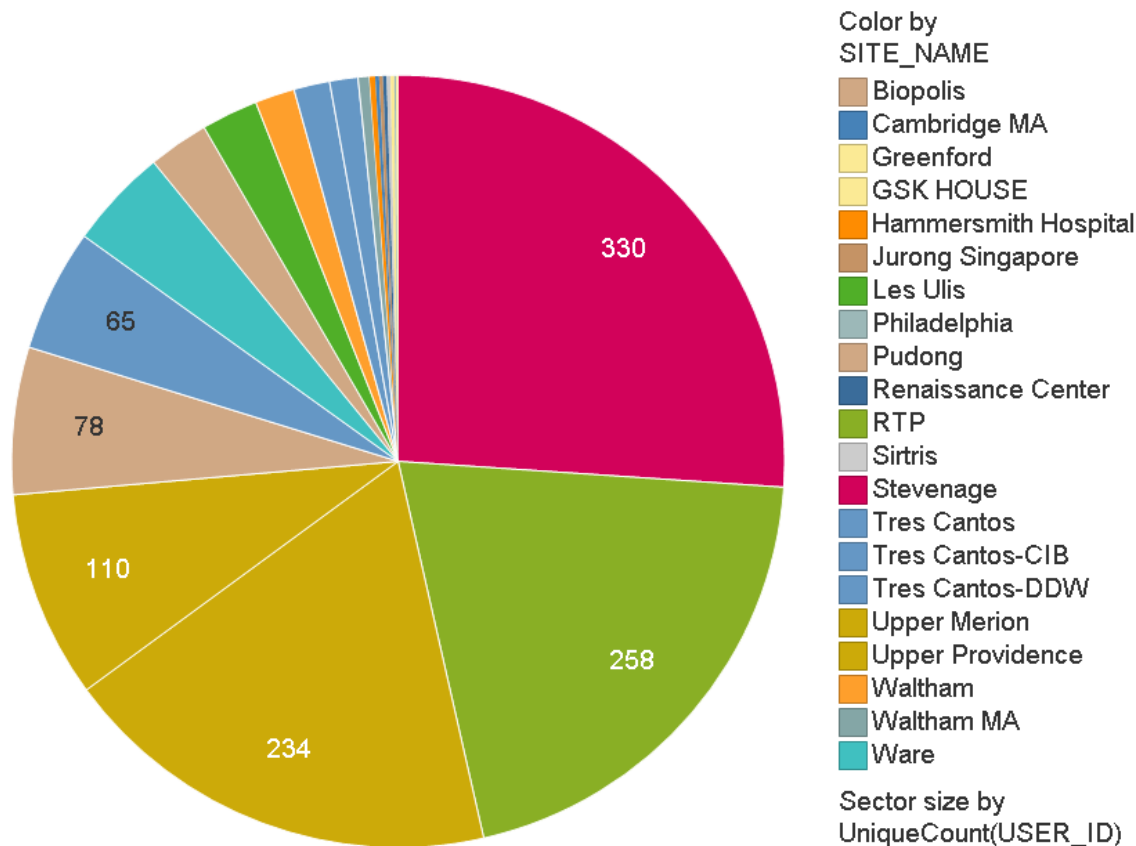
Menu



gsk3 gsk4

Current Deployment Status

- Helium for Excel Version 2.1 is available currently with functionality to replace 6 legacy applications and being used by >1200 researchers.
- Helium for Spotfire was released in Feb 2011 and installed by ~500 users



Bio-IT World Best Practices Awards competition



Winner of Knowledge Management
Best Practices:
GlaxoSmithKline
Helium in Excel: A New Paradigm for
Data Insight
(nominated by Ceiba Solutions)



<http://www.prweb.com/releases/2011/04/prweb5236694.htm>

Future Plans

- Some of the plans/ideas we are working on
 - Implementing Instant JChem to replace ISIS H-Views for forms based data delivery
 - Working with Chemaxon to improve copy / paste functionality of OLE objects (especially ChemDraw) to other Microsoft applications
 - Making Helium available outside of GSK
 - Enhancing Helium to access more datasources such as inventories keyed on barcodes
 - Helium in a Browser

Thank you for your attention!

Questions?