

## Approaches to Information Integration



**Ben Gardner**

**Therapeutic Area Scientific Information Services**

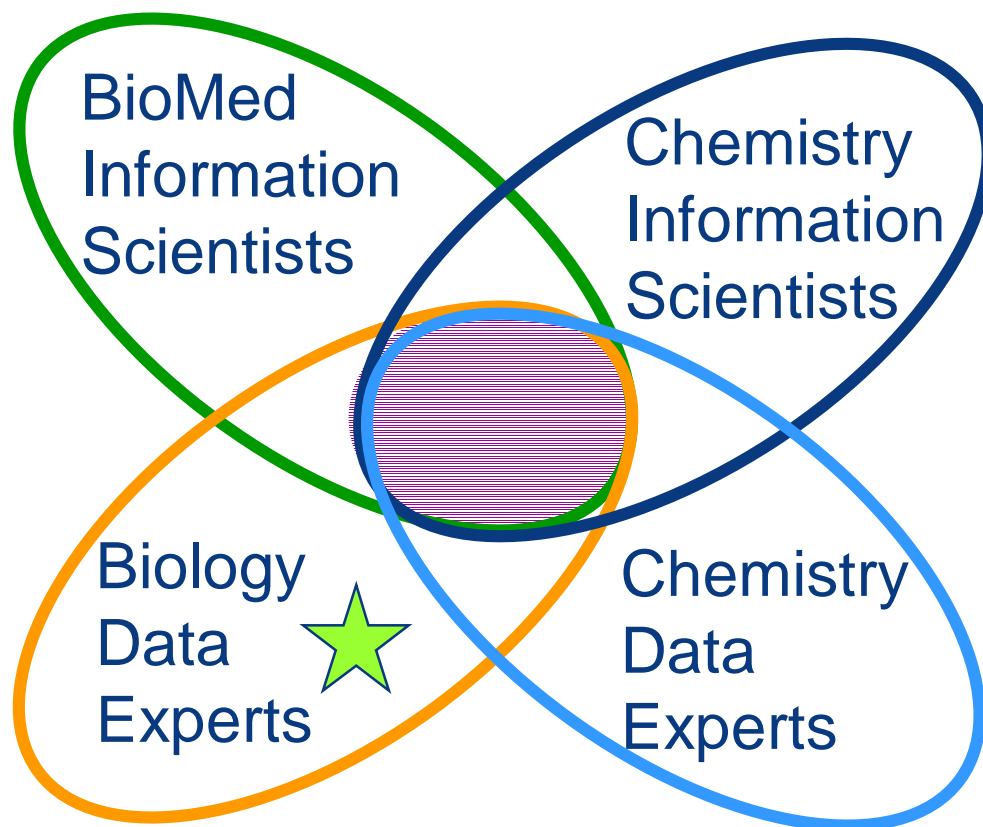
**ICIC Infortics (October 2007)**



**Worldwide Technology**  
***R&D Informatics***

*Pfizer Internal Use*

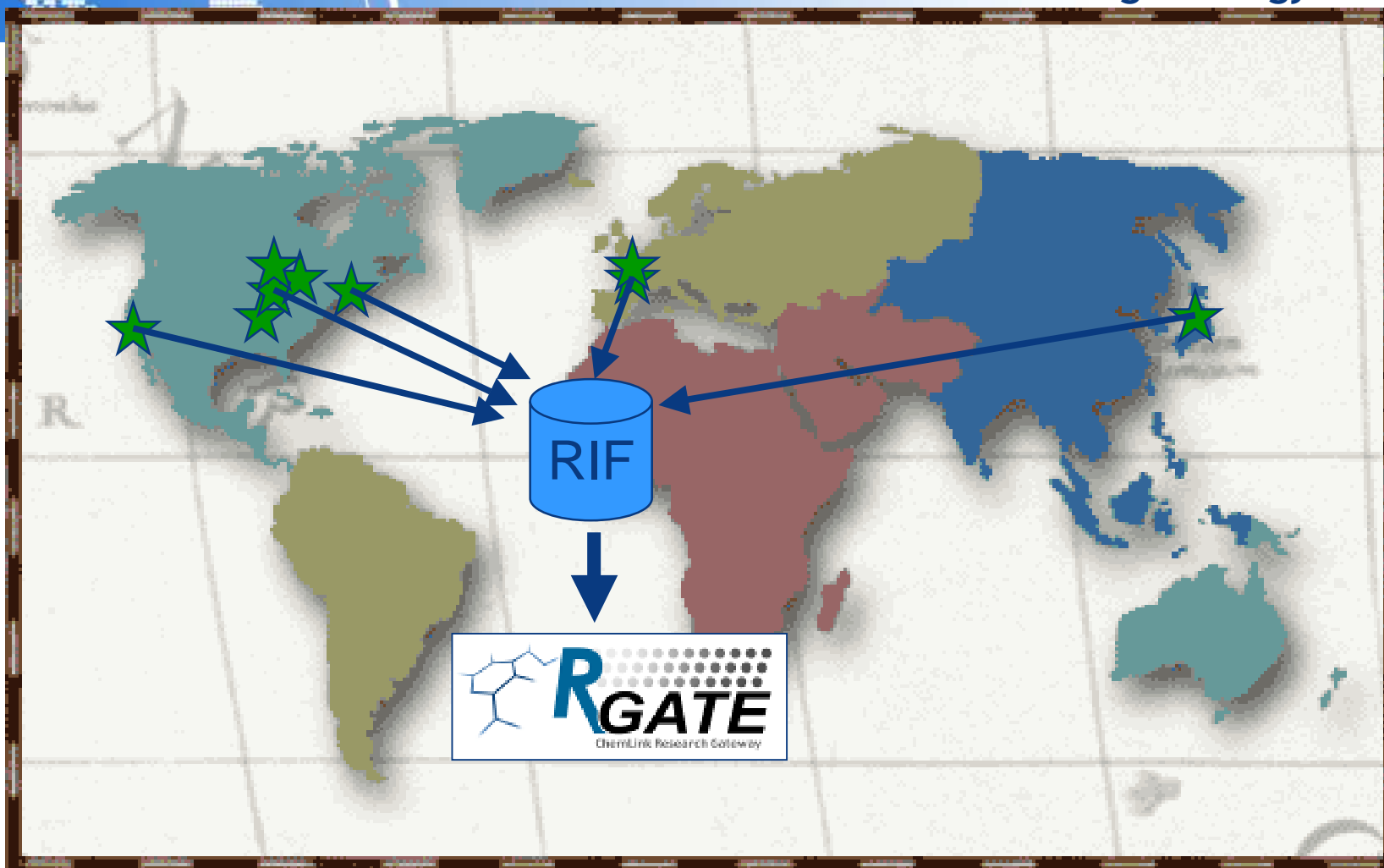
- Therapeutic Area Scientific Information Services
  - ◆ Bringing **Information Scientist** and **Information Technology** expertise together



- Formal organisational recognition of the importance of data
- Access data across the whole of Pfizer (R, D & C)
- Create synergies between IS and IT
- Encourage innovation in data space
- Leverage internal & external data

- Integration of data
  - ◆ Across sites and lines
  - ◆ Public domain, vendor supplied and internal
- Data access
- Web2.0 and information integration

## Utilise a Data Warehousing Strategy



- Research Information Factory (RIF)
- Single global repository for all research data
- Data queried via Data viewers i.e. RGate



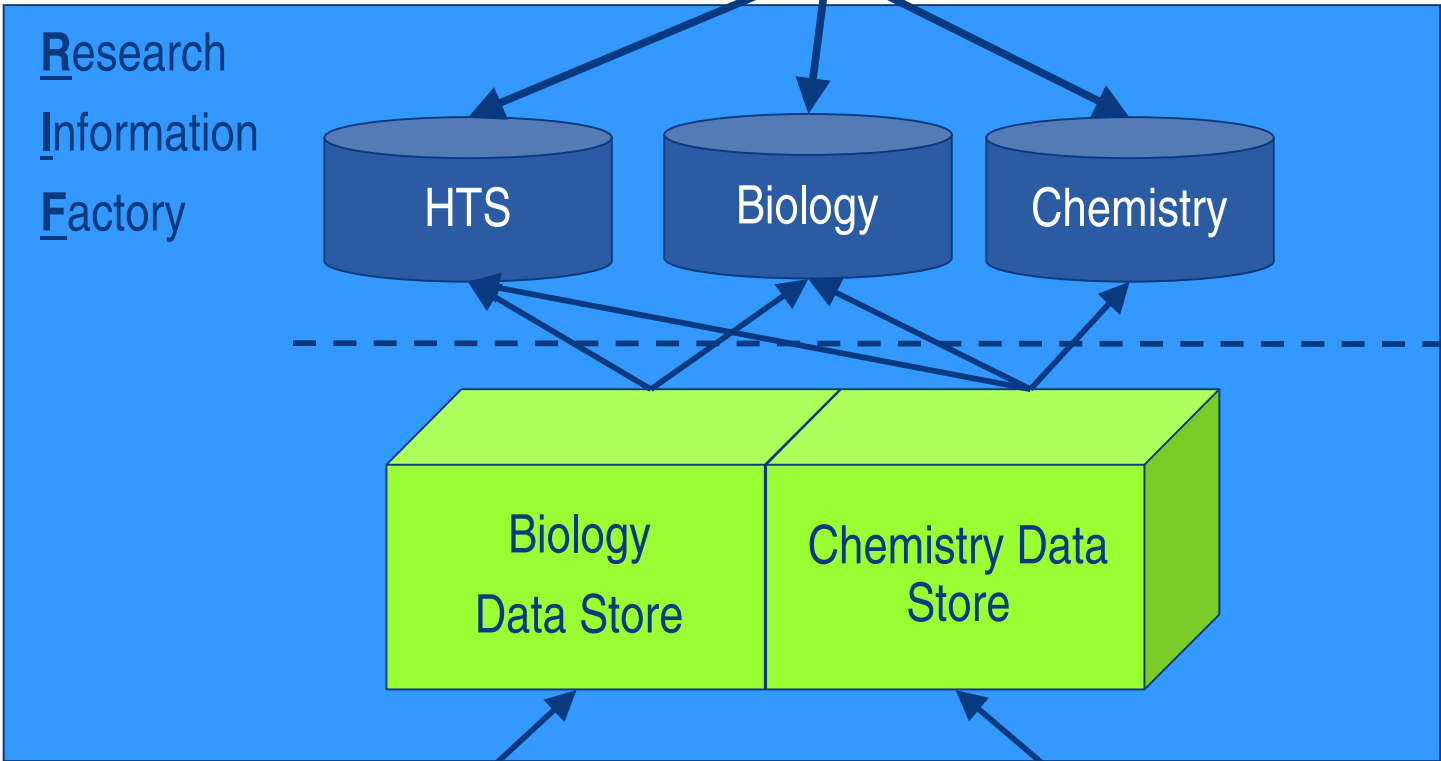


# A simplified view of the RIF



**Data viewer**

Analogous to a shop window



**Data optimised for searching**

Analogous to a showroom

**Data storage**

Analogous to a retail warehouse

GePR, SiGHTS, Bioloader,  
Galileo, Watson, etc.

GeCR, GCM, PCE

**Data capture applications**



**Worldwide Technology**  
*R&D Informatics*

# RGate: Single Global Platform for Data Access & Analysis

**I Want To...**

- Structure Search**
  - Build Structure Search Query
- Queries**
  - Retrieve a Saved Query
  - Build a New Query
  - Build a New Boolean Logic Query
- Template (SAR Table/Form)**
  - Retrieve a Saved Template
  - Build a New Template
- List Management**
  - Retrieve a Saved List
  - Build a New List
  - Combine Lists
  - Import a List
  - Manage Assays
- Virtual Compound Registration**
  - Register Virtual Compounds
  - VCR Worklist
- MoViT Session Files**
  - Retrieve a Saved MoViT Session File
- Preferences**
  - Change My Preferences

**My QuickLinks**

- Rat PK Form (Sandwich)
- Functional SAR tracking (Ben)
- IVT Multi-Protocol (14-8-06)

[View and Manage your QuickLinks](#)

**Quick Viewer**

ID:

Compound ID:

MW:  Formula:

[Quick Structure on Clipboard](#)

[Build Structure Search](#)

**Useful Links**

**Registration**

- GeCR
- GePR

**Mattrack by Site**

- Groton
- Ann Arbor
- St. Louis
- Nagoya
- La Jolla
- Kalamazoo
- Sandwich

**Other Links**

- Chemistry eNotebook
- Drugstore
- Digital Library
- Infosource
- Leviathan
- Live Link
- MoViT
- Pfizer Intranet Home
- PGVL Hub
- PharmaMatrix
- Spotfire Structure Viewer

**Training & Help**

- Computer Based Training Modules(CBTs)
- What's New in RGate v2.1

- A way of life for ~3000 Users
- User base spread across the whole of Pfizer
- Research → Development → Patents → Manufacturing

## RGate: Single Global Platform for Data Access & Analysis

The screenshot displays the RGATE Research Gateway interface. The main window is titled "Query Summary (New Query #1) (4 assays, 28 values to be displayed)". It features a "Build or Modify Your Query" section with a table of search criteria. The table includes columns for "What To Search", "Criteria", and "Show". The criteria listed are: Compound ID, Parent ID, Dose, % Binding, Dose Regimen, Target Outcome, Target Book, Organism Species, and Organism Strain. The "% Binding" criterion is checked. Below the table, there are options to "Use OR between all criteria tabs" and "Exclude virtual compounds from results". The interface also includes a "Query Criteria Choices" list on the left, a "Send To SAR Table" button, and a "Query Not Runnable" message.

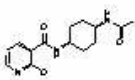
What To Search	Criteria	Show
Compound ID	<a href="#">Add Criteria</a>	<input checked="" type="checkbox"/>
Parent ID	<a href="#">Add Criteria</a>	<input type="checkbox"/>
Dose	<a href="#">Add Criteria nM</a>	<input checked="" type="checkbox"/>
% Binding	<a href="#">Add Criteria</a>	<input checked="" type="checkbox"/>
Dose Regimen	<a href="#">Add Criteria</a>	<input type="checkbox"/>
Target Outcome	<a href="#">Add Criteria</a>	<input type="checkbox"/>
Target Book	<a href="#">Add Criteria</a>	<input type="checkbox"/>
Organism Species	<a href="#">Add Criteria</a>	<input type="checkbox"/>
Organism Strain	<a href="#">Add Criteria</a>	<input type="checkbox"/>

- Build queries using a user friendly interface
- Access >1.8 billion data points
- Search by screen, compound, molecular target, project, etc





# RGate: Single Global Platform for Data Access & Analysis

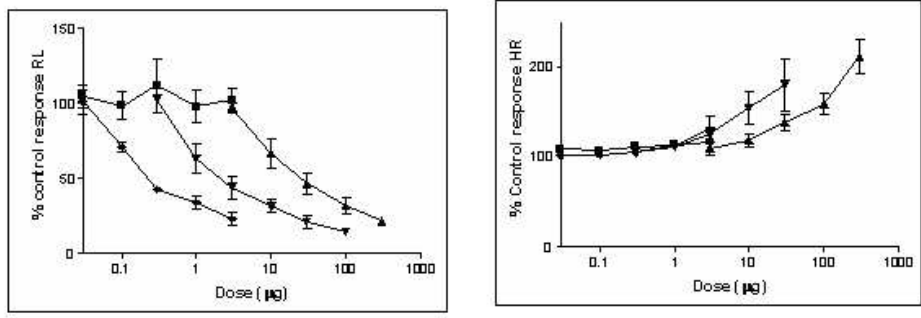
result	Sel	Compound ID	Structure	S6678 TEST (U):Effect	S6678 TEST (U):MED (mg/kg)	S6678 TEST (U):Endpoint Comme	S6678 TEST (U)	S6678 T
		UK-500001		Faster than salmeterol Same as salmeterol Same as salmeterol	>9.00 >10.0 >10.0	Longer efficacy than salmeterol Same efficacy as salmeterol Same efficacy as salmeterol	50.0 >100 >100	HR and Bp No AE at h No AE at h

**Click to drill down to experimental details**

Compound ID	Batch ID	Reference Link (Effect Observed)	Reference Link (MED)	Reference Link (Minimum Toxic Dose)	Experiment Date	Notebook Reference	Compound Level Comment	Experimenter	Source Site
UK-500001	007977-083-00		<a href="#">test2</a>						

**Click hyperlink and open associated document**

My test graphs 1. doc (Read-Only) - Microsoft Word



Test docs to data set 1

Page 1 Sec 1 1/1 At 2.5cm Ln 1 Col 1 REC TRK EXT OVR. English (U.K.)



**Worldwide Technology**  
**R&D Informatics**

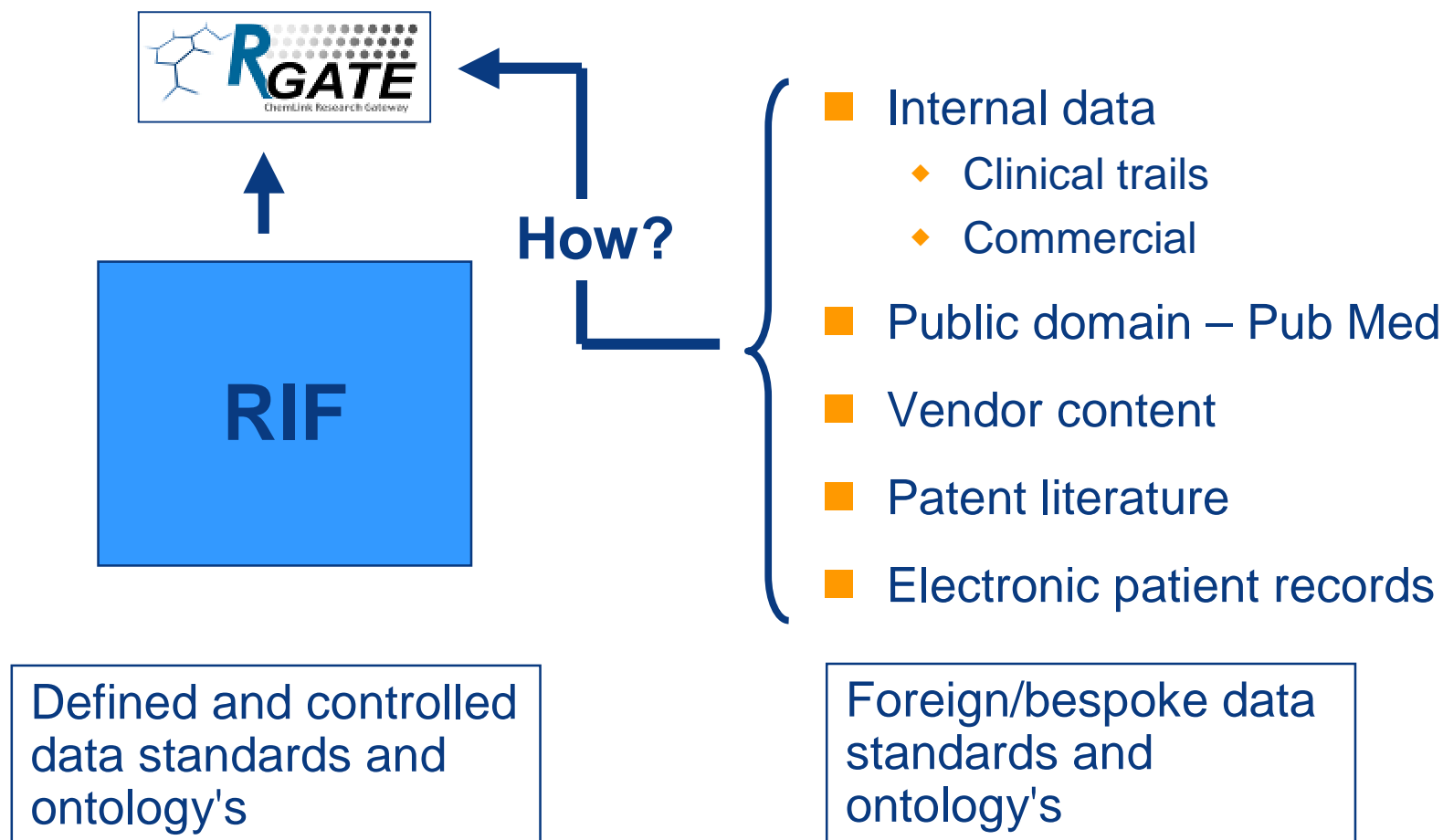
### Pro's

- Easy solution for global data access and sharing
- Incorporates business rules/Global consistency
- Scalability – All in one place
- More than project database – trend analysis

### Con's

- Difficult to incorporate legacy data
- Performance issues
- Scalability – interdependencies /complexity

## BUT what about integrating external data



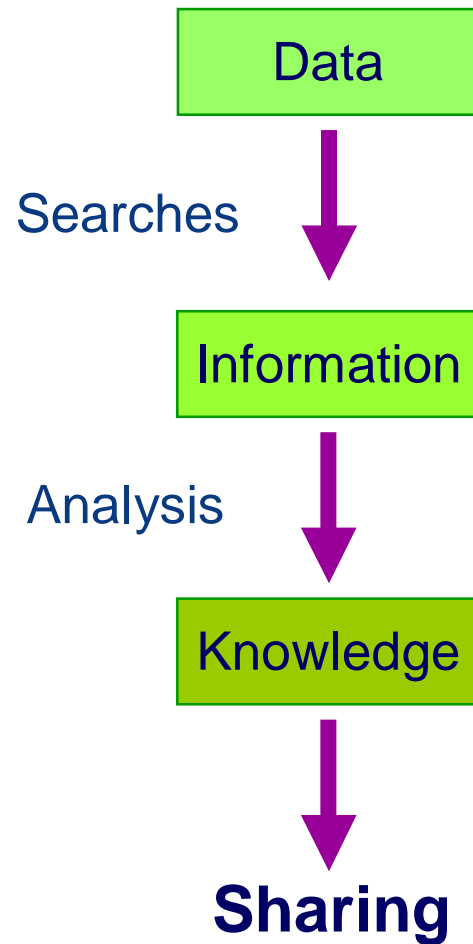


## How can this be made easier?

- Need to propose and adopt industry wide standards
  - ◆ Common keys/ontology's
    - Single disease ontology
    - Single definition of molecular targets
    - Smiles strings & consistent definition of stereochemistry, tautomerism, etc
  - ◆ Content providers intergrate standards in to their data bases
- More vendors could enable simple data integration by
  - ◆ Provide content via web services
  - ◆ Allow direct querying of database
  - ◆ Provide access to the data base schema

**Competitive edge should not come from who owns the data but who can utilise it best**





- Integration projects like the RIF/RGate turn data in to information
- How do you turn information into knowledge?
- Searches now return too much information for one person to analyse
- Need simple tools to share learning's
- Can the tools of Web2.0 help?

- **Web 2.0** refers to second generation of Web-based services that emphasize online collaboration and sharing among users.
- These services include
  - ◆ Social networking sites
    - <http://del.icio.us/>; <http://www.flickr.com/>;  
<http://www.myspace.com/>;
  - ◆ Wikis
    - <http://www.wikipedia.org/>
  - ◆ Communication tools - Blogs & RSS
  - ◆ Folksonomies – Tagging
- Key to the success of these services is their simplicity to use

- A **wiki** is a website that allows the visitors themselves to easily add, remove, and otherwise edit and change available content.
- This ease of interaction and operation makes a wiki an effective tool for mass collaborative authoring.

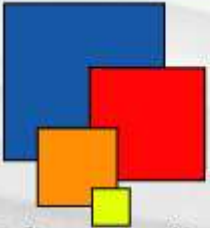
The screenshot shows a Windows Internet Explorer browser window displaying the Pfizerpedia wiki page for RGate. The browser's address bar shows the URL <http://pfizerpedia/index.php/RGate>. The page features a navigation menu at the top with links for 'article', 'discussion', 'edit', 'history', 'move', and 'watch'. The main content area is titled 'RGate' and includes a sub-header 'Research Gateway: Portal To Your Data' with the tagline 'Advancing drug discovery through better informed decisions'. The page text states: 'This page will be used to consolidate RGate's presence within Pfizerpedia. See the RGate TEUS page for information on this project.' Below this, there is an 'Announcements' section with a sub-heading 'Beta Test the New Chemical Structure Search Engine October 10-26' and a brief description: 'Structure Searching is Changing: How will it impact you? Find out. Give it a try before it goes live at the end of the month. The RGate stage environment has been opened to all users.' To the right of the main text is a 'Contents [hide]' table of contents listing various sections and their dates. On the left side of the browser window, there is a 'Pfizerpedia' logo and a navigation menu with links such as 'Main Page', 'News', 'Showcase', 'Community blog', 'Recent changes', 'Random page', 'Support', and 'Help'. A search bar is also visible at the bottom left of the page.

Easy of use is central

The screenshot displays the Pfizerpedia website interface. At the top left, there is a logo consisting of four overlapping squares (blue, orange, red, yellow) above the text 'Pfizerpedia'. Below the logo is a 'navigation' menu with links: Main Page, Community portal, Current events, Recent changes, Random page, Help, and Donations. A 'search' box with 'Go' and 'Search' buttons is located below the navigation menu. A 'toolbox' section at the bottom left contains a link for 'What links here'. The main content area features a 'List of Topics' section with tabs for 'article', 'discussion', 'edit', and 'history'. The 'article' tab is selected. The text below the tabs reads: 'This is a page where you can create new pages. The list is NOT comprehensive. To browse all of the pages in... To add a topic, just click the "edit" tab above, add an internal... on the new red link and start creating your new page.' Below this text is a list of topics, each preceded by a square bullet point: IPI-ConfEx, Toxicokinomics, Enzymology, Generic entry, Pairwise Analysis of Pfizer Data, Beyond Structural Alerts (BSA), CeN, Clinical Trial Simulation, Combination Therapies, Competitive Intelligence, and Computational chemistry.







**Pfizerpedia**

navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)
- [Donations](#)













search

toolbox

[article](#) [discussion](#) [edit](#) [history](#)

## Editing List of Topics

**Warning:** You are not logged in. Your IP address will be recorded

**B**     **A**          

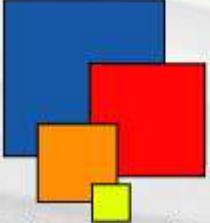
This is a page where you can create new pages.

The list is NOT comprehensive. To browse all of [\[\[Special:Allpages|here\]\]](#).

To add a topic, just click the "edit" tab above, topic (for example `<nowiki>[[My Topic]]</nowiki>` start creating your new page.

- \* [\[\[IPI-ConfEx\]\]](#)
- \* [\[\[Toxicokinomics\]\]](#)
- \* [\[\[Enzymology\]\]](#)
- \* [\[\[Generic entry\]\]](#)



  
**Pfizerpedia**

navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help
- Donations

search

















toolbox

- What links here
- Related changes
- Upload file
- Special pages

article discussion edit history

## Editing List of Topics

Warning: You are not logged in. Your IP address will be recorded.

**B**                

This is a page where you can create new pages.

The list is NOT comprehensive. To browse all pages, see [\[\[Special:Allpages|here\]\]](#).

To add a topic, just click the "edit" tab and add a topic (for example `<nowiki>[[My Topic]]</nowiki>`) to start creating your new page.

- \* [\[\[ICIC 2007\]\]](#) ←
- \* [\[\[IPI-ConfEx\]\]](#)
- \* [\[\[Toxicokinomics\]\]](#)
- \* [\[\[Enzymology\]\]](#)
- \* [\[\[Generic entry\]\]](#)
- \* [\[\[Pairwise Analysis of Pfizer Data\]\]](#)







**Pfizerpedia**

navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)
- [Donations](#)

search

toolbox

- [What links here](#)
- [Related changes](#)

[article](#) [discussion](#) [edit](#) [history](#)

## List of Topics

This is a page where you can create new pages.  
The list is NOT comprehensive. To browse all of the  
To add a topic, just click the "edit" tab above, add an  
on the new red link and start creating your new page

- [ICIC 2007](#)
- [IPI-ConfEx](#)
- [Toxicokinomics](#)
- [Enzymology](#)
- [Generic entry](#)
- [Pairwise Analysis of Pfizer Data](#)
- [Beyond Structural Alerts \(BSA\)](#)
- [CeN](#)
- [Clinical Trial Simulation](#)
- [Combination Therapies](#)
- [Competitive Intelligence](#)



  
**Pfizerpedia**

navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)
- [Donations](#)

search

toolbox

- [What links here](#)
- [Related changes](#)

[article](#) [discussion](#) [edit](#) [history](#)

## Editing ICIC 2007

Warning: You are not logged in. Your IP address will be recorded.

**B**     **A**       



127.0.0.1 talk for this ip log in / create account

article discussion edit history

## Editing ICIC 2007

**Warning:** You are not logged in. Your IP address will be recorded in this page's edit history.

**B**

The ICIC meeting covers trends in the field of scientific and professional information with an impressive line-up of significant figures in the information world. Among the topics covered and discussed by the 25+ speakers are:

- \* tools for intelligence and decision support, including mining chemical and biological data, visualisation, information and entity extraction - followed by a panel discussion.
- \* the new patent landscape, text and structure searching, free patent services, search and substructure searching - followed by a panel discussion.
- \* search engines and data integration.
- \* new business and information centre models.

==Who should attend & Why==

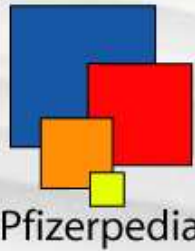
No other conference in Europe fulfils so precisely the need for the professional development and continuing education of Information Professionals.

The conference is of direct interest to information managers and patent searchers and is particularly noted for its coverage of leading-edge 'how-to' technical discussions and reviews of information resources. Commercially-available products and services, free patent databases and other unique resources available worldwide for intellectual property research all form part of the program, as well as issues generally affecting the industry. Additional workshops provide optional training/learning opportunities and a chance to brush-up on the latest searching techniques.

==When & Where==

21-24th October 2007 in Sitges (Barcelona), Spain at the





navigation

- Main Page
- Community portal
- Current events
- Recent changes
- Random page
- Help
- Donations

search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

[article](#) [discussion](#) [edit](#) [history](#)

## ICIC 2007

The ICIC meeting covers trends in the field of scientific and professional information with an impressive line-up of significant figures in the information world. Among the topics covered and discussed by the 25+ speakers are:

- tools for intelligence and decision support, including mining chemical and biological data, visualisation, information and entity extraction – followed by a panel discussion.
- the new patent landscape, text and structure searching, free patent services, search and substructure searching – followed by a panel discussion.
- search engines and data integration.
- new business and information centre models.

**Contents** [\[hide\]](#)

- 1 [Who should attend & Why](#)
- 2 [When & Where](#)
- 3 [Workshops](#)
- 4 [Speakers](#)
- 5 [Links](#)

### Who should attend & Why

[\[edit\]](#)

No other conference in Europe fulfils so precisely the need for the professional development and continuing education of Information Professionals.

The conference is of direct interest to information managers and patent searchers and is particularly noted for its coverage of leading-edge 'how-to' technical discussions and reviews of information resources. Commercially-available products and services, free patent databases and other unique resources available worldwide for intellectual property research all form part of the program, as well as issues generally affecting the industry. Additional workshops provide optional training/learning opportunities and a chance to brush-up on the latest searching



- A **blog** is a user-generated website where entries are made in journal style and displayed in a reverse chronological order.
- Blogs often provide commentary or news on a particular subject, such as food, politics, or local news. The ability for readers to leave comments in an interactive format is an important part of many blogs.
- Blog on a Target, Patent, Disease, Synthetic pathway, Business process.....
- Or even this conference

### lisbon's blog

#### Wednesday Morning

##### **Approaches to Information Integration**, Ben Gardner (Pfizer, UK)

Ben Gardner started off a day that was full of presentations and discussions about integrating information. As an internal data specialist, Ben was unique among the delegates at this conference who are mostly, if not entirely, external information experts. There was keen interest in hearing Ben's perspective, especially among the delegates from AstraZeneca, which has like Pfizer reorganized information scientists and data managers into a joint function. After describing the TA SIS group and its rationale, Ben illustrated approaches we have taken to data integration using the RIF and PharmaMatrix as examples of the integration of internal and external information, respectively. He then identified some types of data integration we currently targeting. He concluded with some suggestions as to what information suppliers could do to facilitate our efforts. His presentation stimulated a lively discussion and debate during the ensuing 50' panel discussion on Information Integration. One point that came up repeatedly was the need for human input, in addition to technology, to truly merge and integrate data/information.

[read more](#) | [4 comments](#) | [7 reads](#)

( categories: [Speakers](#) )

#### Tuesday Afternoon

Anne Marie Ashton (AstraZeneca, Sweden) presented on strategic alignment of information management. This interesting presentation highlighting what AZ is doing to align their services and demonstrate value-add. Many of the theme will not be a surprise to you:

- the need to embed information services into business processes
- the need to understand key decision points and what is required
- the need to be more assertive in demonstrating the strategic value information provided by IM can bring
- the need to demonstrate how IM contributes to the drug development cycle and increasing speed to decision.

[read more](#) | [add new comment](#) | [2 reads](#)

#### Tuesday Morning

1) Aalt van de Kullen (Solvay Pharmaceuticals) spoke on the topic 'Do end-users find the right information?' Aalt










# Social Bookmarking & Folksonomies


scuttle
bookmarks watchlist add a bookmark
log out

recent bookmarks
Sort by: Date Title URL

Search  for

---

[Stamford \(Metro-North station\) - Wikipedia, the free encyclopedia](#)  
 Station close to Gartner's research premises  
 09-10-2007 to gartner, stamford by revells - Copy

---

[ICWSM || International Conference on Weblogs and Social Media](#)  
 A conference devoted to blogs and wiki use for research. Now finished but links to stuff on line presumably  
 08-10-2007 to blogs, research, wikis by revells - Copy

---

[MediaWiki Password authentication](#)  
 Perhaps the start of a means of authenticating (and creating) Mediawiki logins from LDAP  
 08-10-2007 to ldap, mediawiki by revells - Copy


---

[Healthcare on the net - certification](#)  
 A means of certifying the credibility of your health related website  
 08-10-2007 to certification, health, hon, world wide web by revells - Copy

**recent tags**

blogs certification gartner google  
 health hon ldap **mediawiki**  
 mediawiki access control mediawiki  
 example regulatory research resource  
 rpi search semantic slates stamford wiki  
 wiki example wikis world wide web

Popular Tags →



- RIF – Enterprise level integration
  - ◆ Enables global sharing, leverage scale
  - ◆ Requires clean, curated data
- RGate– Internal/External data integration
  - ◆ Need to define industry wide standards
    - Common keys/ontology's
  - ◆ Need vendors to develop new pricing models
- Web2.0 – An opportunity?
  - ◆ Social networking to enable information to knowledge
  - ◆ Easy, quick, proven and here
    - Wiki's, Blog's, Social Bookmarking, Tagging



## Acknowledgements

