Precompetitive Informatics Initiatives in Drug Discovery

Bryn Williams-Jones
bryn.i.williams-jones@pfizer.com
Associate Research Fellow
Head of eBiology
Pfizer Sandwich Research Enabling Group
Why aren’t we more productive?

Industry Productivity vs. Investment
The Challenge

Total R&D Investment ($ Billions)

Source: PhRMA annual survey, 2000
Drug Discovery Processes - The Scale Of The “Portfolio Reload” Need

preESD | ESD | SDS | LD | FIH

50+ | 8-12+ | 5-8+ | 3+ | 3+

PubMed

Literature

Competitors

Smart Ideas

Conferences

In vivo Platforms

Pathway Studies

Data Mining

Drug Repurposing

Omics

Clinical Feedback

Collaborations

Merger/Aquisition

Technology Platforms
The Technology Stack for Electronic Biology

**Data**
- Targets; Chemistry; Pharmacology; Literature; Patents

**Standards**
- Ontology/taxonomy; Minimum information guide; Dictionaries; Interchange mapping

**Assertions**
- e.g. Gene-to-Disease; Compound-to-Target; Compound-to-ADR

**Application (Knowledge)**
- Fact Visualisation: e.g. Target Dossiers; SAR Visualisation

**SERVICES**

- Define needs; Design algorithms; Develop “plug-in” architectures?
- Define needs; Contribute algorithms & develop tools (e.g. text mining); Enhance existing approaches
- Support existing standards; Drive new DD-relevant ontologies; Work with publishers
- Defining needs; Knowledge; Data Contribution
Data tombs
TARGETS SURROUNDED BY INFORMATION

“Too much data”   “Too many applications”
“This doesn’t apply to me”   “You need a PhD in IT to use this stuff”
“What does this really mean to my project?”
Public Domain Drug Discovery Data
- The Current Situation

Pharma are accessing, processing, storing & re-processing public domain data

- Literature
- PubChem
- Genbank
- Patents

Data Integration → Databases → Downloads

Data Analysis

Firewalled Databases
Public Domain Drug Discovery Data
- The Current Situation

We are all doing this many times……
The changing landscape
Changes in R&D Information Strategies

Then

• Internalisation of external content
• Extensive Internal software build
• Vertical Application Development
• Internal application management
• End-to-end Ix service delivery internally
• Inflationary Budgets
• Limited assay types & content volume
• Little Collaboration

Now

• Increased push for services
• Reduced capability for Internal software build
• Externalisation of application management
• Increasingly reliant on sourcing external Ix services
• Flat Budgets
• Huge data/content volumes
• Increasing collaboration
Lowering the firewalls

Pre-competitive collaboration

- Must change the way we do things
- Complexity is increasing, budgets are decreasing
- We can’t afford to try to invent better wheels
- Quality of public sources increasing
- Even with one of the worlds largest R&D budgets, >95% of science research done outside

Nature Reviews Drug Discovery 8, 701-708 (September 2009) | doi:10.1038/nrd2944
Computational Chemistry and Biology are not the same!

- Drivers for change are universal but domain needs differ
- Compare the desktops of comp chemists to comp biologists
  - **CompChem:** Commercial or proprietary in-house data and tools
  - **CompBiol:** commercial and in-house tools sit alongside (and in many cases are based upon) a vast selection of public domain resources.
- Consider the nature of the data which they analyse
  - **CompChem:** highly competitive data (e.g. novel small molecules).
  - **CompBiol:** largely public data much earlier in the discovery pipeline
- Distinctions present different challenges
  - **CompChem:** need to foster development of public domain tools
    - Chemistry resources need security and client-side interfaces
  - **CompBiol:** Almost the extreme opposite - large numbers of public domain resources that companies are struggling to utilise for DD
    - Biological resources need re-focusing for DD
What is pre-competitive?

• In simple terms – activities that don’t offer a significant competitive advantage

Competitive
• Candidate compound
• First crystal structure
• ‘Omics data from difficult to obtain clinical samples

Pre - Competitive
• 10th crystal structure in a family
• 3 year old molecular profiling data
• A gene dictionary built from public resources
Gene reference databases: a pre-competitive paradigm?
- In most pharma, no public gene index has been “quite right” for internal use due to issues such as:
  - accessibility, stability, integration rules, comprehensiveness or scope
- Ironically, pharma have each (re)created individual views of gene information using public data
  - Ultimately this is self-perpetuating; continually requiring maintenance and development to meet the perceived “unique needs” of Pharma

What do Pharma really need?
- Can we do this once, ‘properly’, somewhere that everyone can access?
- Can we use services from a public resource?
The Data Standards & Analysis Challenge

- In many areas of science our ability to generate data is outstripping our ability to curate, analyse and understand the data.

- We are all now generating data on a truly industrial scale.

- The analysis of scientific data also needs to become as industrialised.

- In order to accomplish this will require data standards.
Where are pre-competitive activities taking place?

**EBI Industry Programme**
- The EBI industry programme hosts pre-competitive quarterly meetings with 16 member companies working in the field of pharmaceutical and biotechnology R&D informatics. Members co-organize intensive workshops that focus on key informatics issues encountered during drug discovery and development.

**Pistoia**
- Pistoia is an initiative to streamline non-competitive elements of the pharmaceutical drug discovery workflow (chemistry, biological screening and logistics) by developing open standards for common business terms, relationships and processes.

**Innovative Medicines Initiative**
- IMI is a public–private partnership between the European Federation of Pharmaceutical Industries and Associations, and the EU. The goal of IMI is to share knowledge from biopharmaceutical sector by pooling competencies and resources to forge public–private collaborations.
SCOPE: The Industry Programme:

- provides a forum for interaction between the EBI and our users in industry
- provides training for our commercial users
- informs ‘industrial users’ of EBI’s plans
- feeds industry requirements into the EBI’s planning
- provides a neutral meeting place for inter-company interactions on bioinformatics
- coordinates workshops on topics decided by programme – gathering expert speakers (industrial and academic)
- initiates ‘special projects’ at the EBI with targeted collaborative funding
- liaises as appropriate with other industry initiatives

Member Companies

- AstraZeneca
- Bayer Schering Pharma
- Boehringer Ingelheim
- Galderma
- GlaxoSmithKline
- Eli Lilly and Company
- J & J Pharmaceutical R&D
- Merck Serono S.A.
- Nestlé Research Centre
- Orion Pharma
- Philips Research
- Pfizer Ltd
- Syngenta
- Sanofi-Aventis
- Unilever

http://www.ebi.ac.uk/industry/ind-program-index.html
An Emerging Cross Pharma Collaboration: The Pistoia Alliance

http://pistoiaalliance.org
Pistoia Membership
updated: Jan 22, 2010

AstraZeneca
GlaxoSmithKline
Novartis
Pfizer
Bristol-Myers Squibb
Roche
UPCO
Lundbeck
chem/IT/ment
DeltaSoft Inc.
Knime
Rescentris
GGAChemAxon
accelrys
Symyx
BioXpr
CDD
MERCK
RSC Advancing the Chemical Sciences
Boehringer Ingelheim
EMBL-EBI
Thomson Reuters
CambridgeSoft
Infosys
Mission of Pistoia

- Pistoia is the BRIDGE to cross the chasm to a more agile pre-competitive environment

---

STANDARDIZE → SIMPLIFY → CENTRALIZE

PISTOIA → MEMBERS → SUPPLIERS
The Pistoia Path Forward: Standardize, Simplify, Centralize

- **Standardize** our interfaces & messages
- **Simplify** our cross-industry architectures & support models
- **Centralize** services to reap economies of scale & scope

**Mission**

To streamline pre-competitive workflow elements of Pharmaceutical R&D by specifying common business terms, relationships & processes
Pistoia Domains
Focused on business workflows/supply chains

Enabling
- Vocabulary
- Visualisation
- Workflow
- Others

Knowledge and Information Services

Application Integration

- Biology Data Services
- Chemistry Data Services
- Translational Data Services
The Innovative Medicines Initiative is a unique Public-Private Partnership (PPP) between the pharmaceutical industry represented by the European Federation of Pharmaceutical Industries and Associations (EFPIA) and the European Communities represented by the European Commission.

IMI's overall goal is to make Europe again the world leader in pharmaceutical research for the benefit of the economy and society, by removing research bottlenecks in the current drug development process.

The world’s largest public private partnership

€2 Billion, multiple pillars, average project €20 million
Time for Change: Public Domain drivers

- Public Domain are engaging in Drug Discovery (DD)
- Public Domain Chemistry Resources are improving
  - NIH Roadmap Initiative
    - Molecular Library Screening Center Network (MLSCN)
    - PubChem (structure, bioassay and bioactivity data).
  - DrugBank, ChEBI, ChemBank and ChEMBL

- Databases supporting biology-based DD are less apparent
  - Rich public domain resources for biology are not DD-centric

- Pharma spends over $50 billion p.a. on R&D
  - How much of this knowledge/information is in the public domain?
  - How much knowledge is tacit? e.g. druggability?
  - How much is truly competitive?
“Virtualisation” needs standards
Why Standardise?
Future Architecture in Pre-competitive World

Applications mixing public and proprietary data sources

Integration Layer – mixing of public and proprietary data

- Historical resource costly bespoke solution mirroring and integrating public data with proprietary
- Resource focussed on integration with less available for innovation
- Most Pharma companies replicate this pipeline

Workbenches
Public Domain

Proprietary Plugins

Pfizer Data Services

Services
(Public Domain)

Services
(Proprietary)

Proprietary Data Store

- Future stable high quality public resources can be taken directly, proprietary data and services being overlaid
- Substantially less resource needed on integration if common standards are implemented
- Pharma and public share higher quality stable resources
Benefits for Everyone

Academics
• More powerful and integrative access to more data
• Making own data more accessible
• Potential for innovative discovery
• A framework to spinout Academic SMEs

SME
• A mandated standard that will make tools/systems immediately usable by any customer (the raison d'être for Pistoia I think?)
• Potential to leverage public domain data more easily
• An opportunity to become "VHS compatible" rather than betamax (maybe that's a bit too controversial)

Pharma
• More powerful and integrative access to more data
• Opportunity to switch off internal systems
• Potential for Innovative Discovery
• Secure client-side access to public data
Some questions particular to today…

- Connections with other organisations
  - CDISC, W3C
  - How open are others prepared to be?

- Science is global
  - Pre-competitive work is not geographically limited
  - How do we better influence US funders, Institutions, Consortia, Disease Societies etc?

- Working in disease areas that are not oncology
  - How can we develop generically applicable workflows that can be pressure tested in data rich areas?
Acknowledgements

Pfizer

• Lee Harland, Cory Brouwer, Ian Harrow, Enoch Huang, Rob Hernandez, Stephen Campbell, Phil Verdemato, Markella Skempri, Chris Waller, Kevin Hebbel, Ted Slater, Cathy Marshall, Dave Burrows, Nigel Wilkinson, Jerry Lanfear

External

• Ian Dix (AZ), Niklas Blomberg (AZ), Nick Lynch (AZ)
• Mike Barnes (GSK), Matt Hall (GSK), Chris Larminie (GSK), Ashley George (GSK), Steven Foord (GSK), Malcolm Skingle (GSK)
• Members of the Pistoia Alliance
• EFPIA IMI collaborators
• Members of the EBI Industry Programme