Precompetitive Informatics Initiatives in Drug Discovery

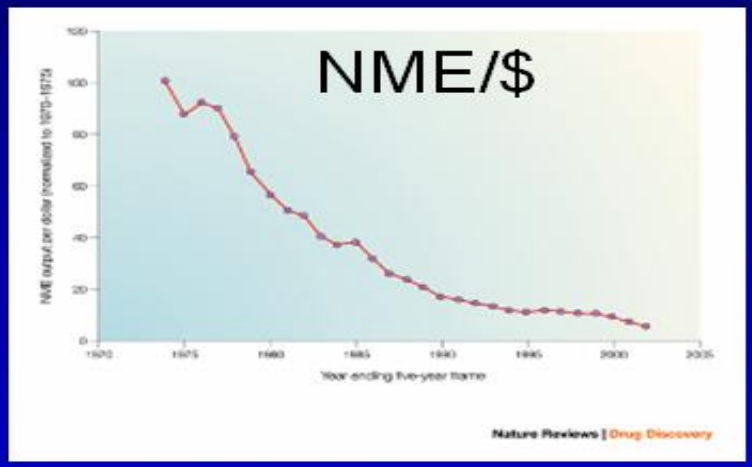
Bryn Williams-Jones
bryn.i.williams-jones@pfizer.com
Associate Research Fellow
Head of eBiology
Pfizer Sandwich Research Enabling Group

Why aren't we more productive?

SandwichResearch

Industry Productivity vs. Investment
The Challenge

Total R&D Investment ($ Billions)

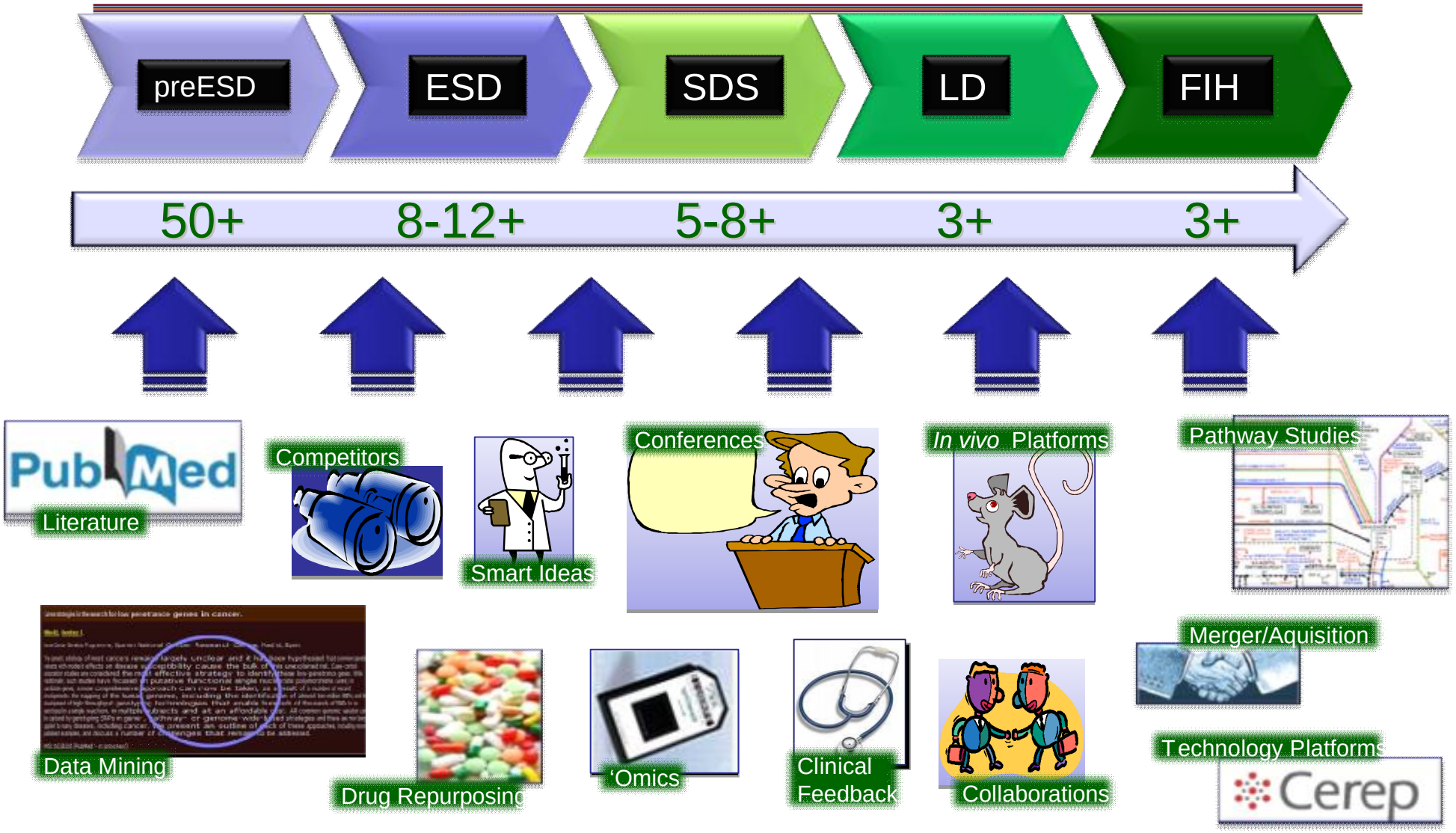


Source: PhRMA annual survey, 2000

Nature Reviews Drug Discovery 3, 451-456 (2004)

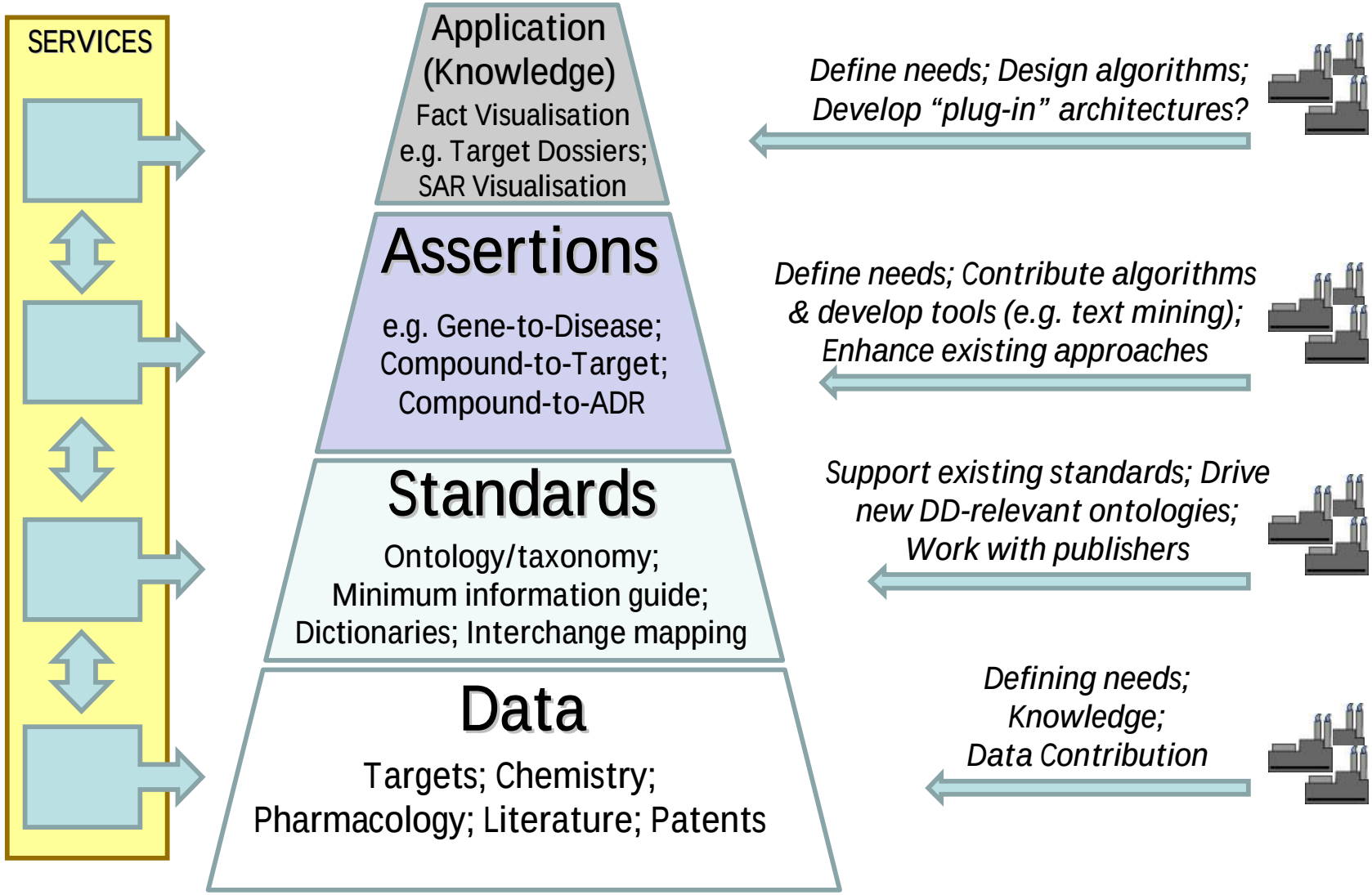
Drug Discovery Processes - The Scale Of The "Portfolio Reload" Need

SandwichResearch



The Technology Stack for Electronic Biology

SandwichResearch



SandwichResearch

Data tombs

TARGETS SURROUNDED BY INFORMATION

SandwichResearch

Pa

Chemical tool

Molecular Interacti

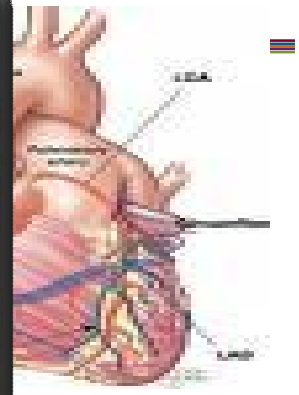
Protein Structure and Function

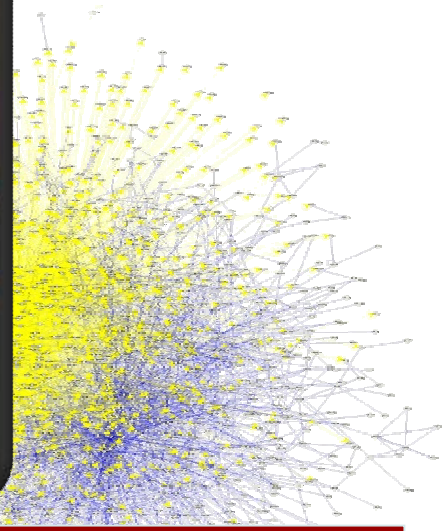
Sequence Variation

Splice Variants

SNPs and Pharmacogenomics

Phylogeny





Homologues and Mutagenesis

"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

SandwichResearch

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

Literature
Genbank
Patents
PubChem

Databases

Downloads

Data Integration

Data Analysis

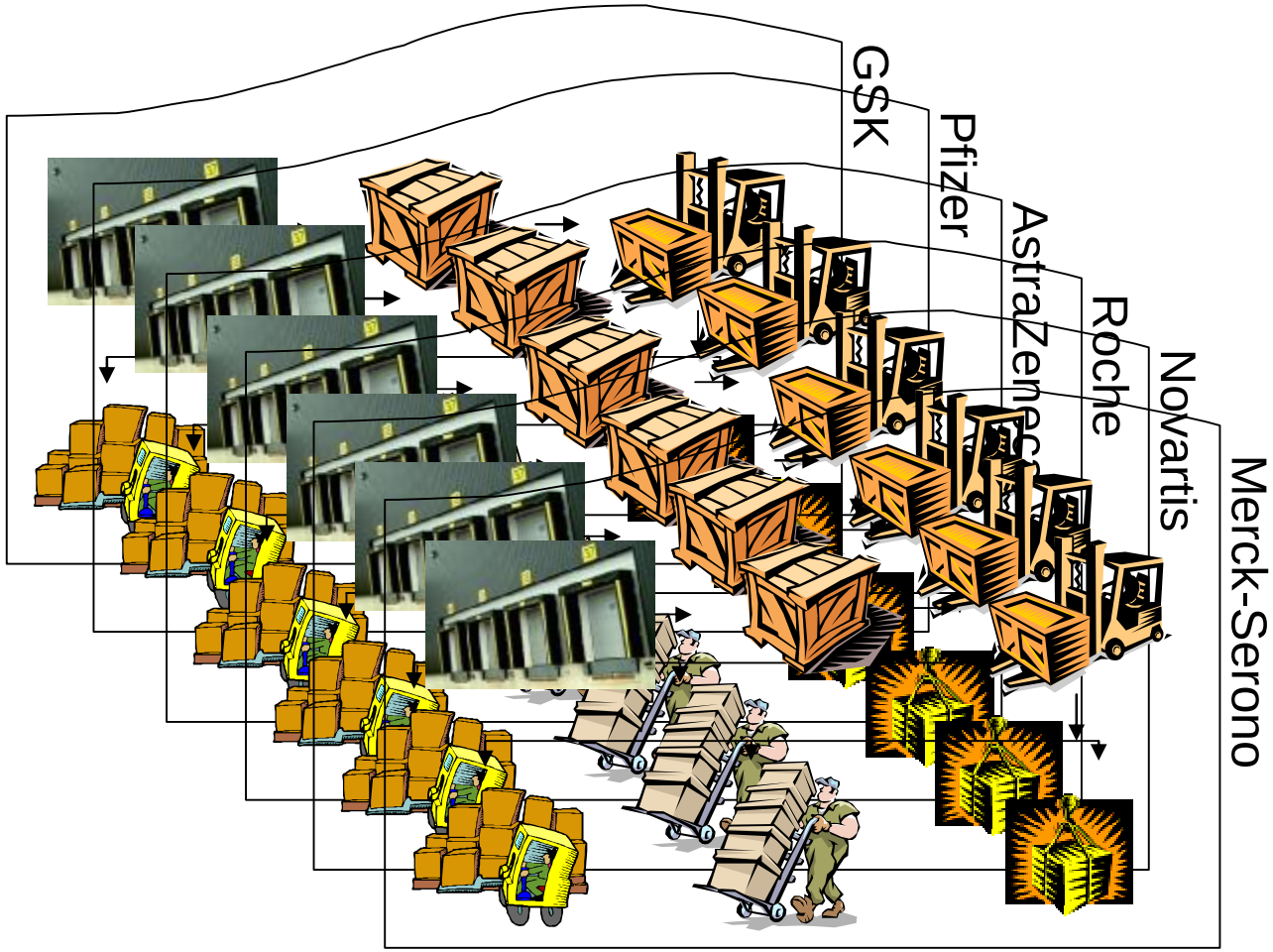
Firewalled Databases

Public Domain Drug Discovery Data

- The Current Situation

SandwichResearch

We are all doing this many times……



The changing landscape
Changes in R&D Information Strategies

SandwichResearch

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

SandwichResearch

PERSPECTIVES

OPINION

Lowering industry firewalls: pre-competitive informatics initiatives in drug discovery

Michael R. Barnes, Lee Harland, Steven M. Foord, Matthew D. Hall, Ian Dix, Scott Thomas, Bryn I. Williams-Jones and Cory R. Brouwer

Abstract | Pharmaceutical research and development is facing substantial challenges that have prompted the industry to shift funding from early- to late-stage projects. Among the effects is a major change in the attitude of many companies to their internal bioinformatics resources: the focus has moved from the vigorous pursuit of intellectual property towards exploration of pre-competitive cross-industry collaborations and engagement with the public domain. High-quality, open and accessible data are the foundation of pre-competitive research, and strong public–private partnerships have considerable potential to enhance public data resources, which would benefit everyone engaged in drug discovery. In this article, we discuss the background to these changes and propose new areas of collaboration in computational biology and chemistry between the public domain and the pharmaceutical industry.

Challenges such as declining productivity, patent expiries and a downward trend in drug pricing require the pharmaceutical industry to rapidly adapt to survive[1–3]. Moreover, while revenues are declining, the costs involved in meeting the safety and efficacy requirements of regulatory agencies are steadily increasing[4]. Against this backdrop, industry is also responding to the growing need for new drugs to treat rare diseases and diseases of developing countries[5–7]. These factors are causing pharmaceutical companies to continually re-assess ways of managing costs to direct the necessary investment towards later-stage pipelines while improving confidence in target and candidate drug selection.

Both internal and external influences are driving this change. Reduced internal funding is leading to an increased focus on data exploitation rather than data generation. With a growing body of high-quality data in the public domain, it is now recognized that the sheer volume of data or the computational platforms used to present it cannot in themselves transform drug discovery output. Instead, methods of data exploitation are crucial, leading to a shift from acquisition of internal platforms to investment in analysis and exploitation of data. Furthermore, over the past several years, understanding of how genome and related biology data can be used to inform target and compound choice has increased. This is leading to more selective investments in data and information platforms to ensure maximum impact. Technology developments (such as next-generation sequencing and high-content screening) have dramatically increased the sophistication of internal platforms required for managing public data. These costs are increasingly difficult to manage, and have limited competitive benefit — a trend that is set to continue.

The external factors behind these changes in the approach of the pharmaceutical industry to data include innovation with, and diversification of, data platforms in the public domain. These public data resources offer alternatives to internal data management and integration, and failure to adopt them could present an opportunity cost. In addition, over the past few years, there has been a maturation of services in the public domain. Historically, issues of robustness of public domain systems — including stability, accessibility and backwards compatibility, and a lack of structured release updates — has prevented the adoption of external tools within industry. However, as the public domain has matured and user bases have been firmly established, more stable and reliable services that meet all the needs of industry have become available.

As industrial drug discovery re-appraises its business model, the public-domain research community has become increasingly engaged in drug discovery efforts[8]. This has been supported at the level of regulators and governments by the Critical Path Initiative[9] of the US Food and Drug Administration (FDA), and the UK government's Cooksey Review of UK health research funding[8]. Both agencies recommended an increase in public–private partnerships as an alternative drug development model. Publicly funded research organizations such as the US National Institutes of Health (NIH) are already enthusiastically pursuing this route with their Roadmap Initiative and the establishment of the Molecular Library Screening Center Network (see Further information). A central component of this strategy has been the development of PubChem[10], a database of known small-molecule drugs that includes structure, bioassay and bioactivity data. Other public-domain drug discovery-focused systems are emerging, including DrugBank[11], ChEBI (Chemical Entities of Biological Interest)[12] and ChemBank[13].

These resources are already supporting public-domain drug discovery, although most are chemistry-focused and are not well integrated with public-domain target biology resources. However, much of the tacit biology and chemistry knowledge concerning 'best practice' for drug discovery — the complexities of target tractability, validation, safety, efficacy, and drug formulation and delivery — still resides primarily within industry, and the availability of information of this nature through public databases is limited at present.

Nature Reviews Drug Discovery **8**, 701-708 (September 2009) | doi:10.1038/nrd2944

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

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?

SandwichResearch

- 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

Internalisation Vs “Virtualisation”

- 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.

EBI Industry Programme

SandwichResearch

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
F. Hoffmann-La Roche
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-prog-index.html

An Emerging Cross Pharma Collaboration:

The Pistoia Alliance

http://pistoiaalliance.org

Pistoia Alliance

Pistoia Membership

updated: Jan 22, 2010

AstraZeneca

gsk
GlaxoSmithKline

NOVARTIS

Pfizer

Bristol-Myers Squibb Company

Roche

Lundbeck

chem/IT/ment
innovative solutions for life science applications

UPCO

KNIME

DeltaSoft Inc.

ChemAxon

accelrys®

Symyx

RESCENTRIS

GGA

BioXpr

Edge Consulting

MERCK
Be well

RSC | Advancing the Chemical Sciences

Boehringer Ingelheim

CDD COLLABORATIVE DRUG DISCOVERY

THOMSON REUTERS

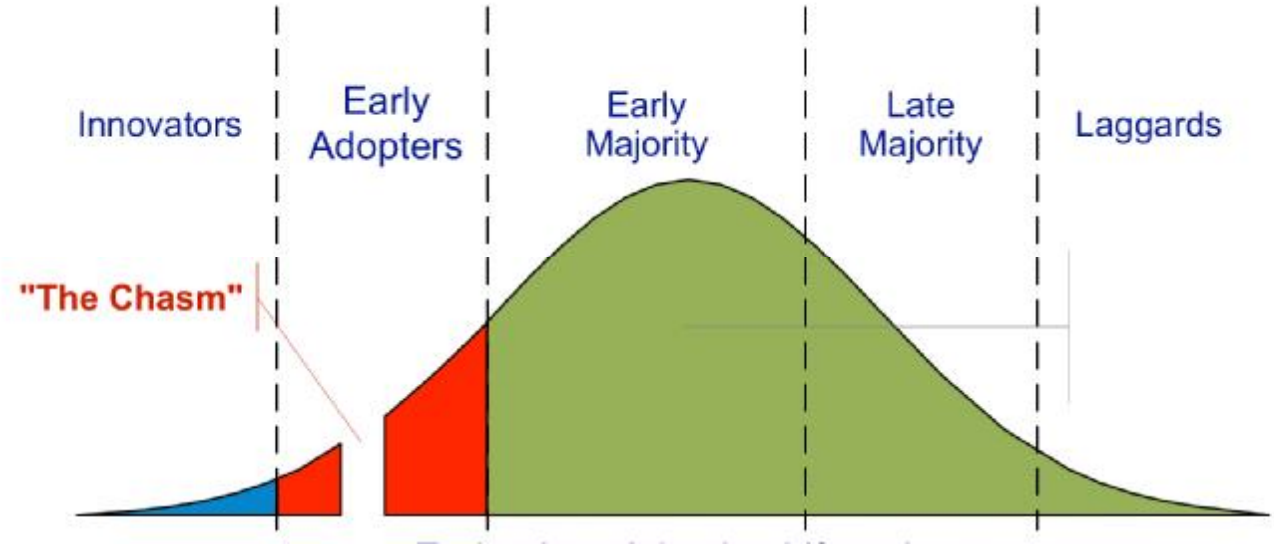
CambridgeSoft
Life Science Enterprise Solutions

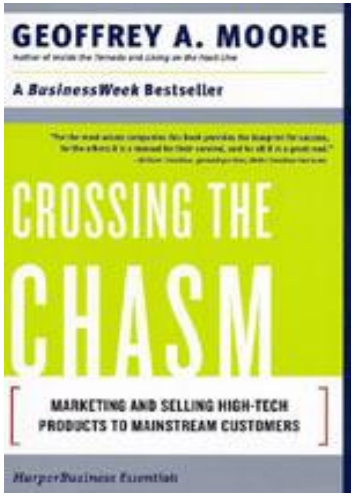
Infosys®

EMBL-EBI

Mission of Pistoia

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





The Pistoia Path Forward: Standardize, Simplify, Centralize

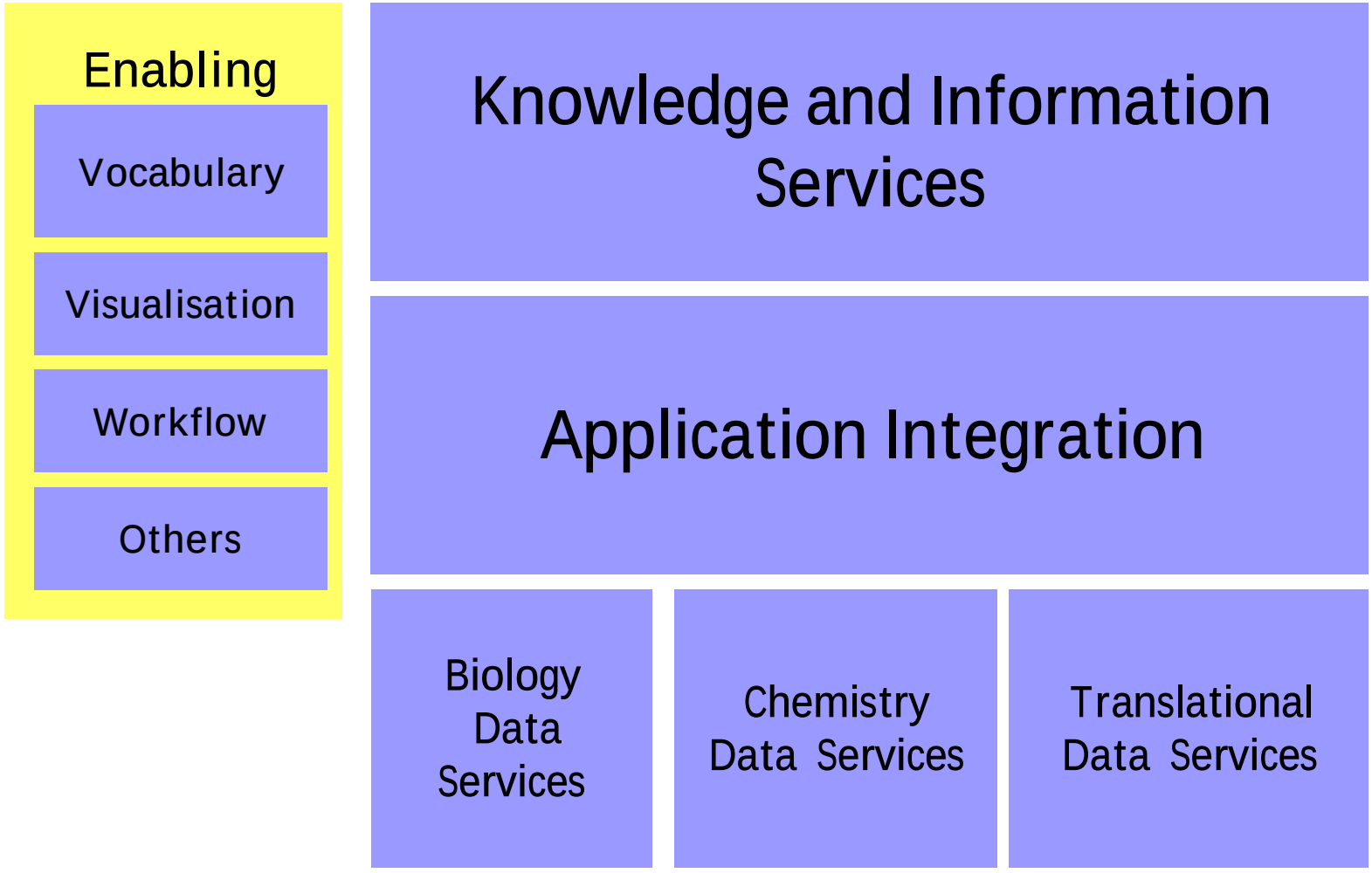
- Standardize our interfaces & messages
- Simplify our cross-industry architectures & support models
- Centralize services to reap economies of scale & scope

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



Innovative Medicines Initiative

SandwichResearch

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.

http://imi.europa.eu/index_en.html

The world's largest public private partnership

SandwichResearch

€2 Billion, multiple pillars, average project €20 million

Time for Change : Public Domain drivers

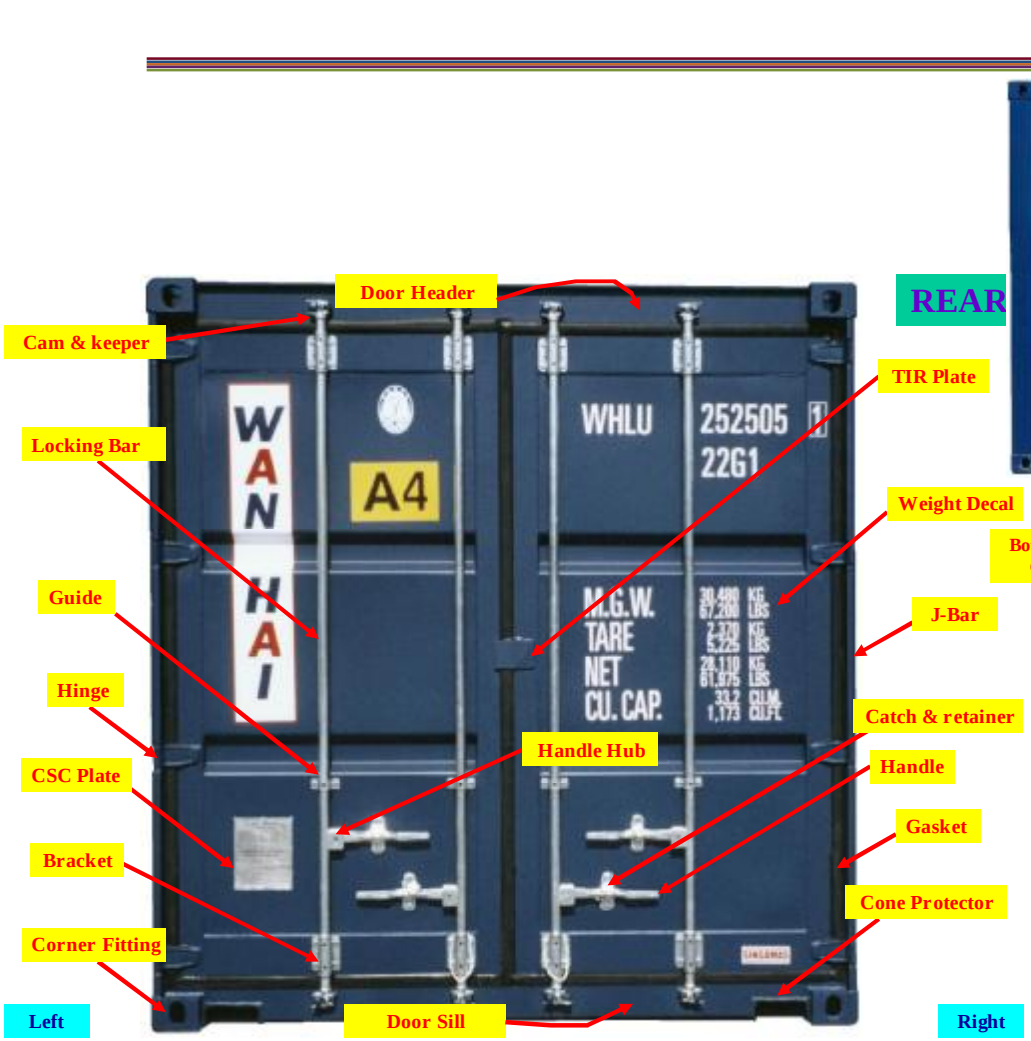
- Public Domain are engaging in Drug Discovery (DD)
- Public Domain Chemistry Resources are improving
  - NIH Roadmap Initiative
    - Molecular Library Screening Center Network (MLSCN)
    - PubChem (structure, bioassay and bioactivity data).
  - DrugBank, ChEBI, ChemBank and ChEMBL

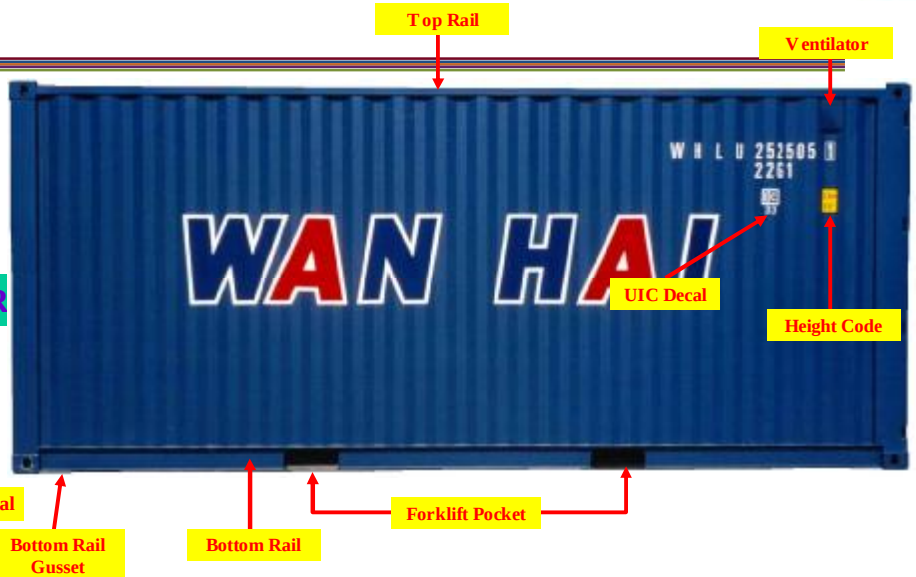
- Databases supporting biology-based DD are less apparent
  - Rich public domain resources for biology are not DD-centric

- 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

SandwichResearch



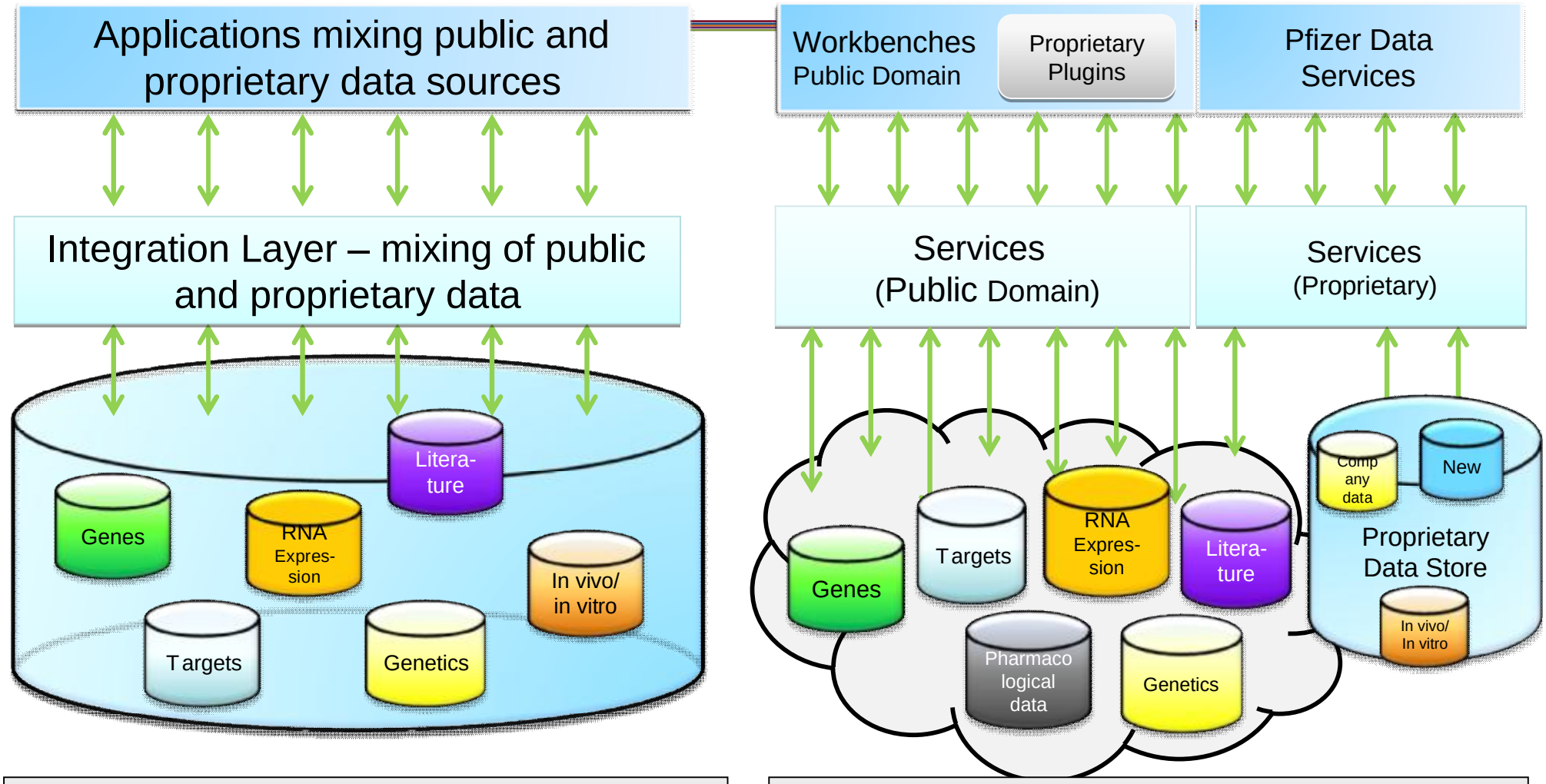


Why Standardise?

SandwichResearch

Future Architecture in Pre-competitive World

SandwichResearch



- 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

- 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'etre 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