The Pistoia Alliance

A Construct for Pre-competitive Collaborations

Chris L. Waller, Ph.D.
Pfizer

Collaborative Drug Discovery Annual Meeting
San Francisco, CA
October 2009
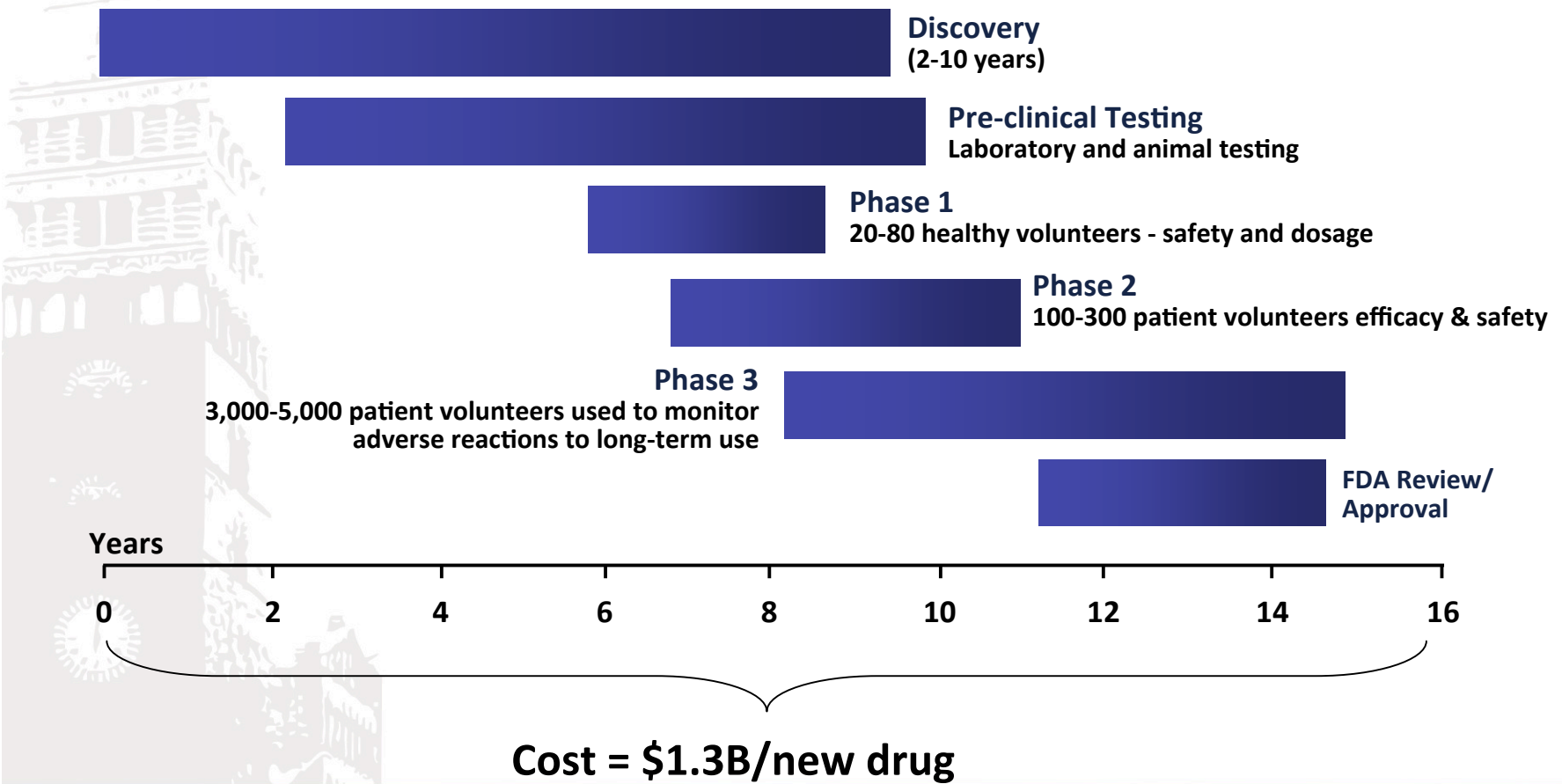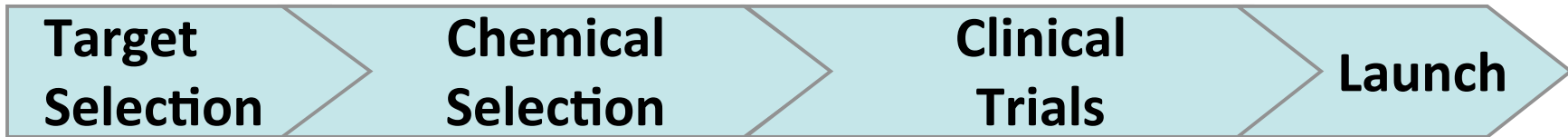
Pistoia
Alliance

Agenda

- Introduction
- Origins of Pistoia
  - History
  - Industry Drivers
  - Technology Trends
- Scope and Operations of Pistoia
  - Mission, Membership, Governance
  - Projects and Deliverables
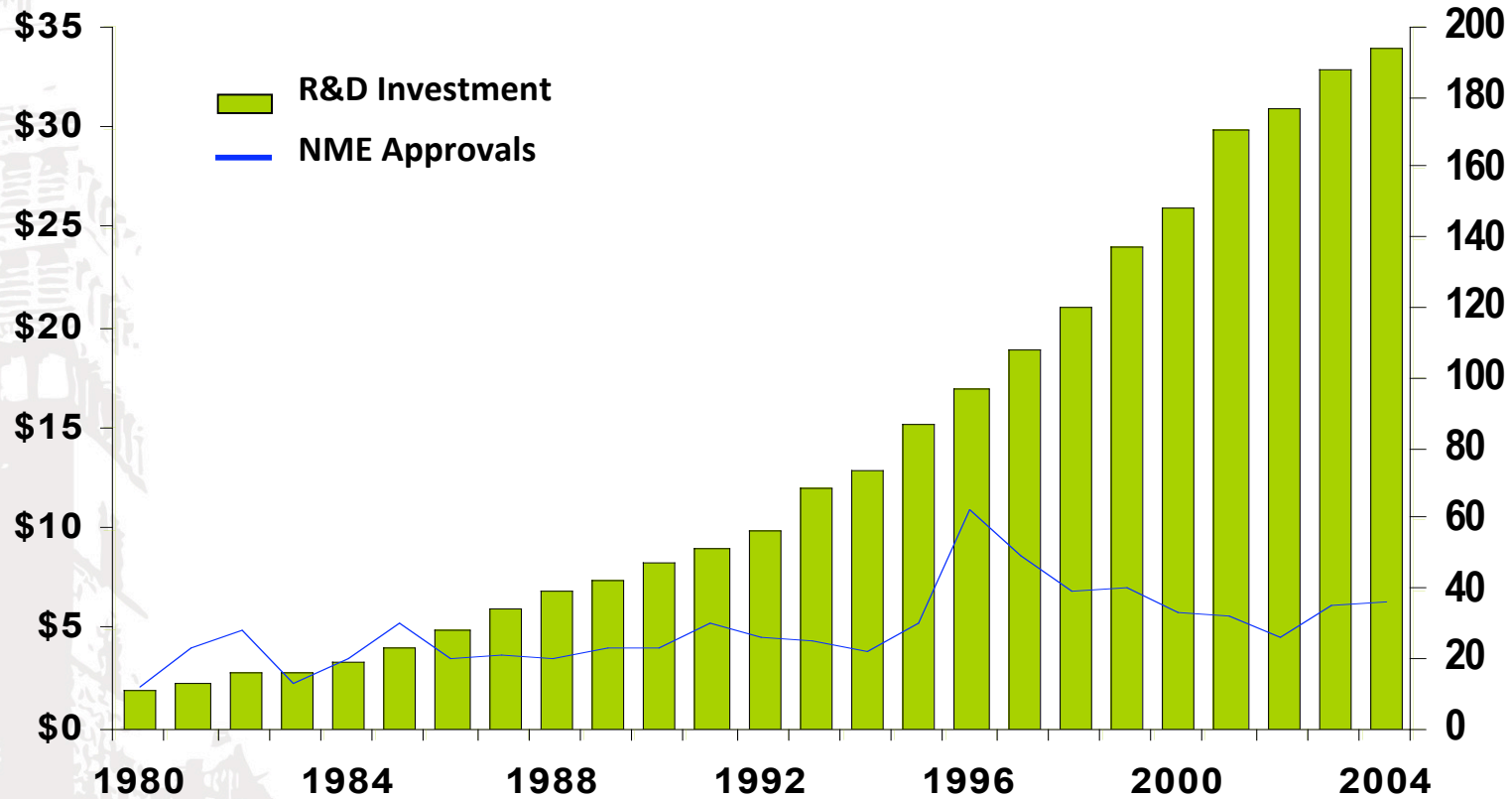- Discussion

R&D: Long, Expensive, and Risky

Pistoia Alliance

| Target Selection | Chemical Selection | Clinical Trials | Launch |

Discovery
(2-10 years)

Pre-clinical Testing
Laboratory and animal testing

Phase 1
20-80 healthy volunteers - safety and dosage

Phase 2
100-300 patient volunteers efficacy & safety

Phase 3
3,000-5,000 patient volunteers used to monitor adverse reactions to long-term use

FDA Review/ Approval

Years

0   2   4   6   8   10   12   14   16

Cost = $1.3B/new drug

Productivity is Decreasing

Chris's Mission Statement

- Foster collaborations between pharmaceutical, biotechnology, technology, academic, and government organizations in <u>pre-competitive</u> space to develop and promote the use of standards, identify partnerships, and transfer technology in order to drive greater process efficiency and lower costs.

Pre-competitive

- Refers to standards, data, or processes that are common across an industry and where the adoption, use, or prosecution of which provides no competitive advantage relative to peers.
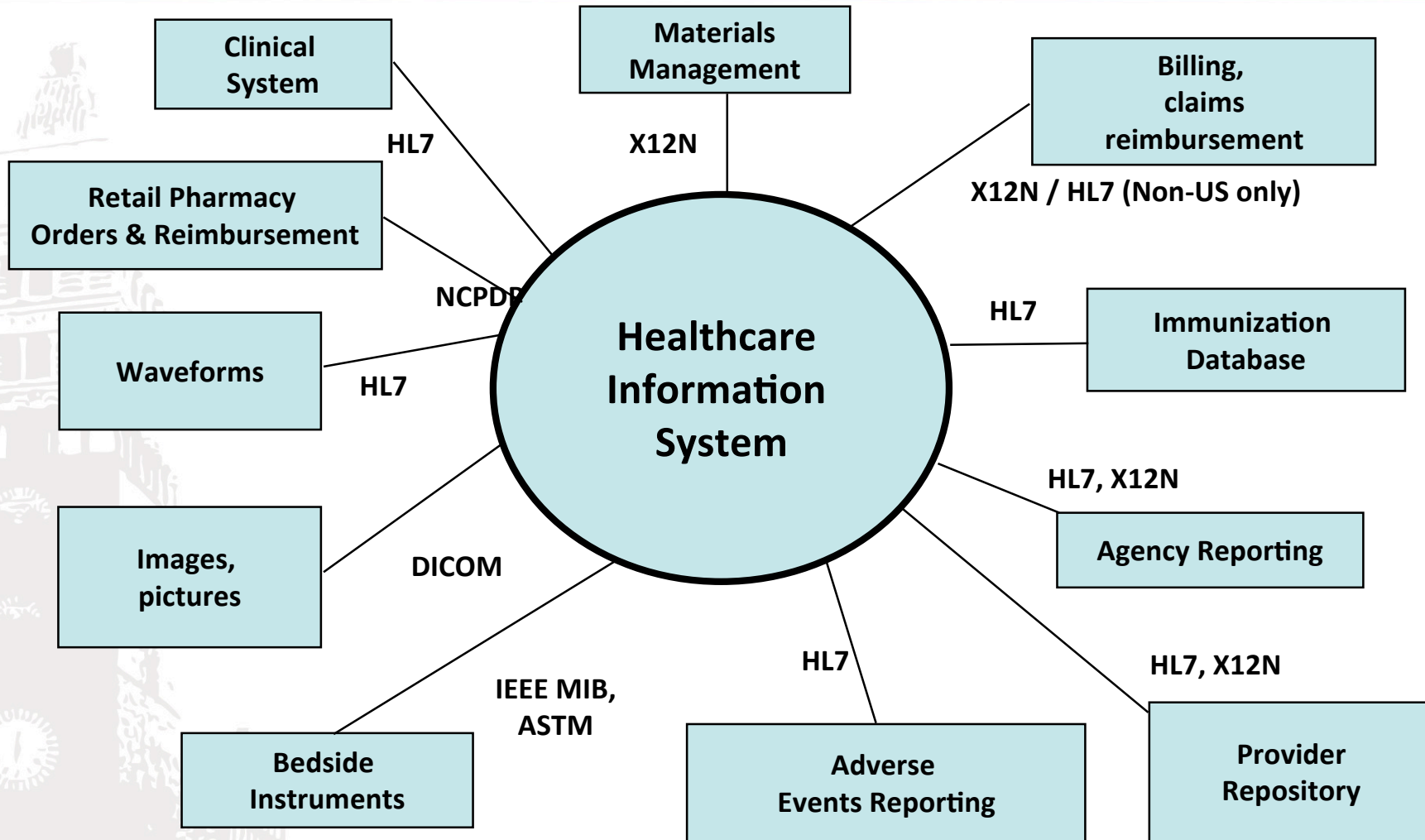
Leading Healthcare Data Exchange Standards

Pistoia Alliance

Clinical System

Materials Management

Billing, claims reimbursement

HL7

X12N

X12N / HL7 (Non-US only)

Retail Pharmacy Orders & Reimbursement

NCPDP

Healthcare Information System

HL7

Immunization Database

Waveforms

HL7

Images, pictures

DICOM

HL7, X12N

Agency Reporting

IEEE MIB, ASTM

HL7

HL7, X12N

Bedside Instruments

Adverse Events Reporting

Provider Repository

Image from Abdul-Malik Shakir

Background—How it all started

- In Pistoia, Italy

- Meeting of GSK, AZ, Pfizer and Novartis—identified similar challenges and frustrations in the IT/Informatics sector of Pharmaceutical Discovery
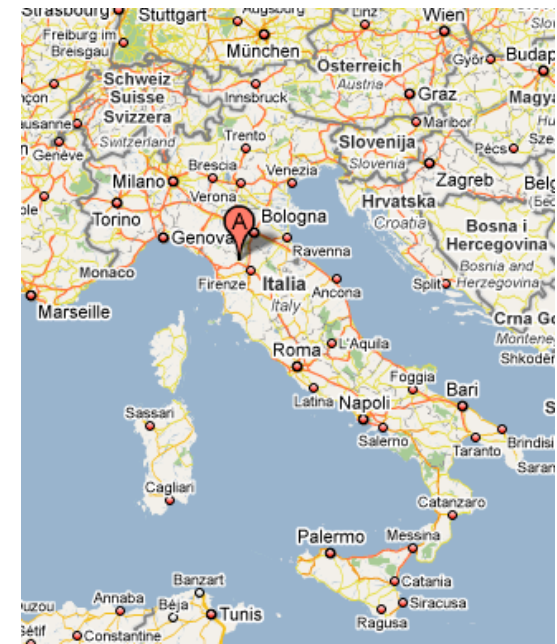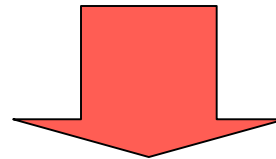
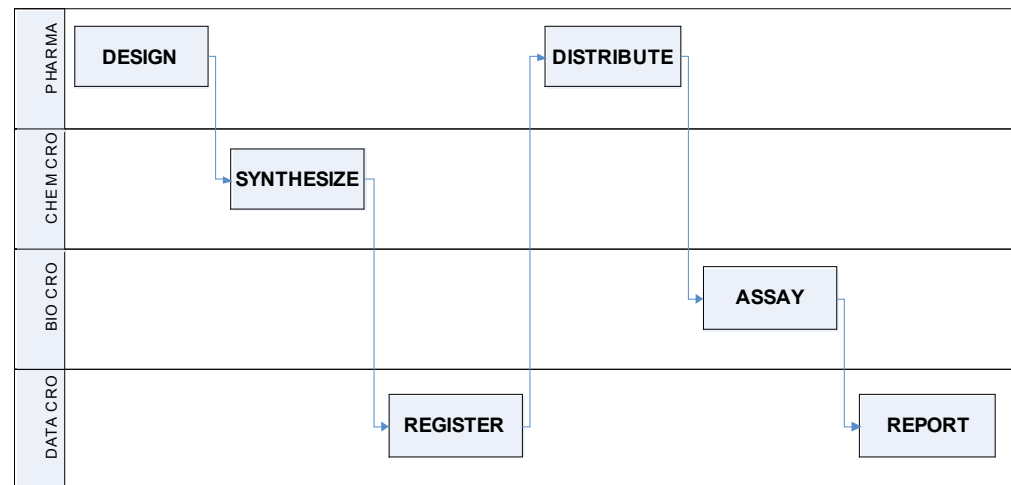# Industry Driver: Externalization

Cost pressures, disruptive technologies, and other forces often drive business processes to be externalized.
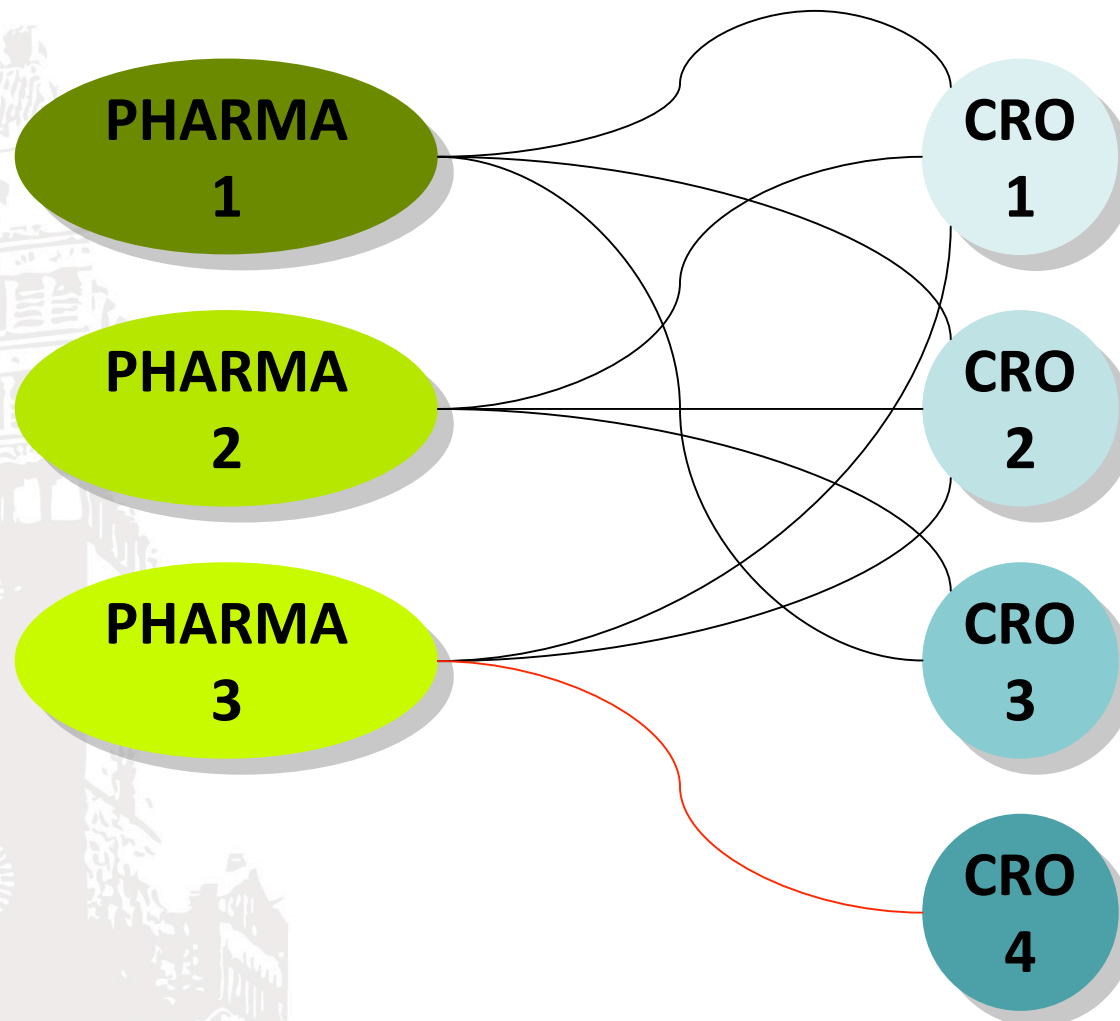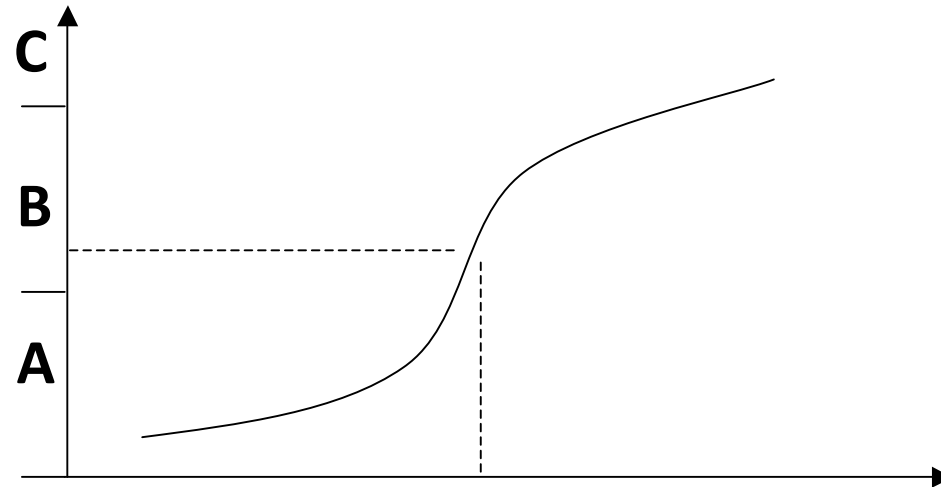
Fully Internal Model

PHARMA	DESIGN	SYNTHESIZE	REGISTER	DISTRIBUTE	ASSAY	REPORT

Selectively Integrated Model

PHARMA	DESIGN			DISTRIBUTE		
CHEM CRO		SYNTHESIZE				
BIO CRO					ASSAY	
DATA CRO			REGISTER			REPORT

PHARMA

CHEM

BIO

DATA

Greater Challenge: Unknown Semantic Collisions
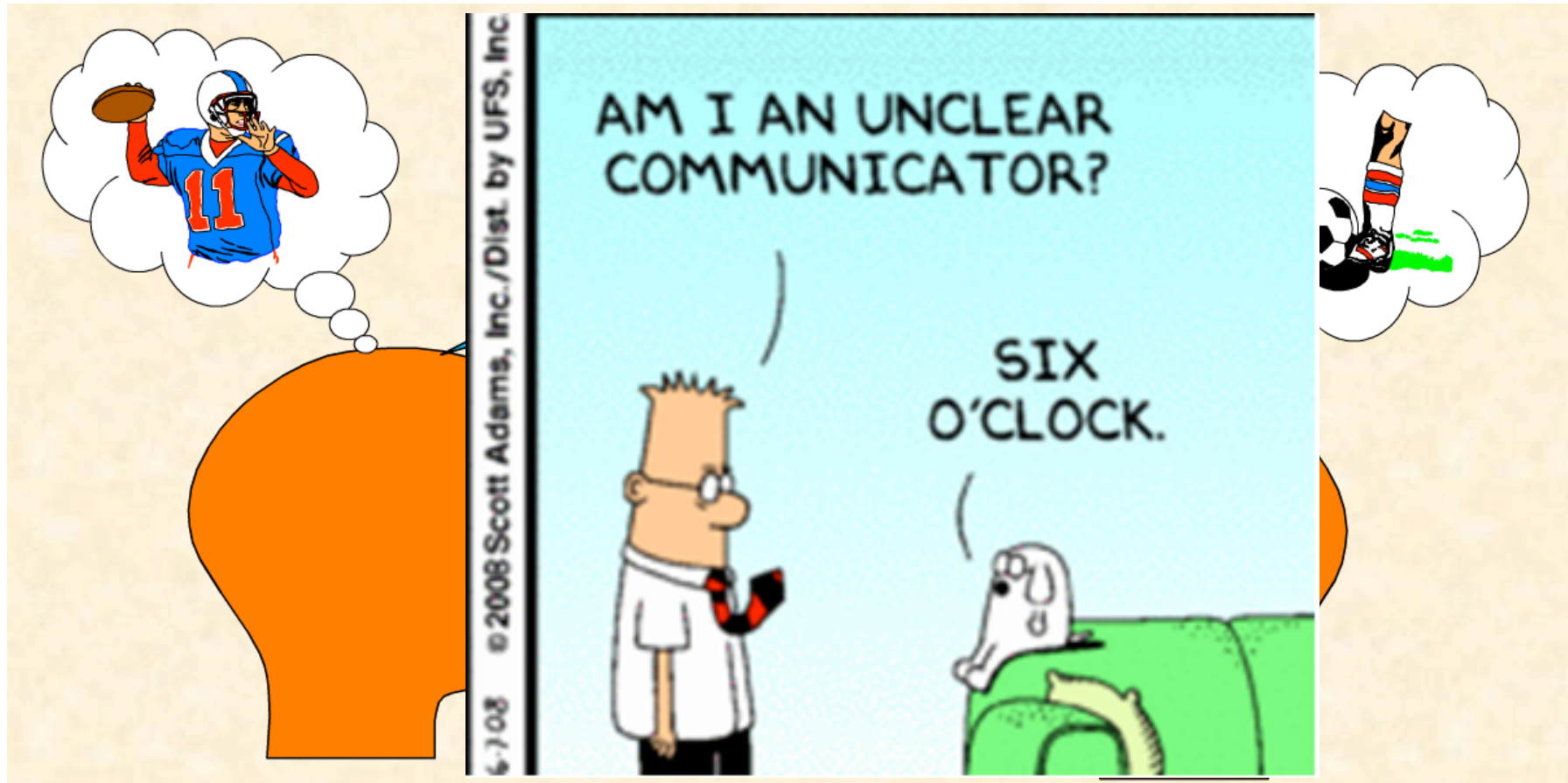
Revealing assumptions is an essential component of effective communication.

Image/quote from Abdul-Malik Shakir

Opportunity: Changing Tech Landscape

More Robust Technologies
- Web 2.0
- Services-Oriented Architecture
- Software-as-a-Service
- Open Source Initiatives

More Robust External Content
- Publicly available chem and bio sources
- Richer literature content
- Academic Sources of Tools and Data

The Path Forward:
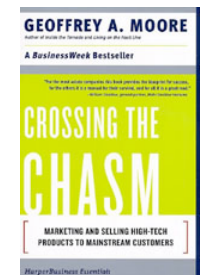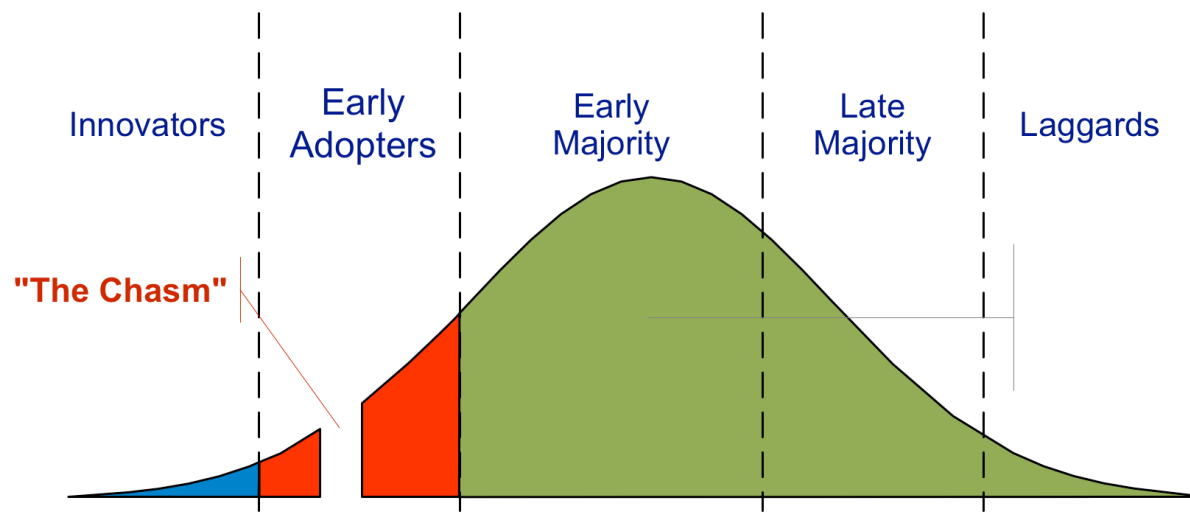## Standardize, Simplify, Centralize

- <u>Standardize</u> our interfaces and messages
- <u>Simplify</u> our cross-industry architectures and support models
- <u>Centralize</u> services to reap economies of scale and scope

Mission of Pistoia

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

Innovators | Early Adopters | Early Majority | Late Majority | Laggards

"The Chasm"

STANDARDIZE → SIMPLIFY → CENTRALIZE

PISTOIA → MEMBERS → SUPPLIERS

GEOFFREY A. MOORE

A BusinessWeek Bestseller

CROSSING THE CHASM

MARKETING AND SELLING HIGH-TECH PRODUCTS TO MAINSTREAM CUSTOMERS

Learn from Other Industries

Transportation

Banking

Geospatial

Retail

Clinical

Automotive

Healthcare

Pistoia Description and Purpose

Mission

To streamline pre-competitive workflow elements of pharmaceutical research and development by specifying common business terms, relationships and processes

Goal

- Develop taxonomies and vocabularies, application interface specifications, data dictionaries, data models, etc.
- Establish standards that will be embraced by producers and consumers of pre-competitive workflows

Pistoia Membership

Pistoia Alliance

>65 Individuals from 18 member organizations

Pistoia Alliance Organisation

Board of Directors

Develops roadmap
Final approval of technical standards
Voting Members (pharma, biotech)
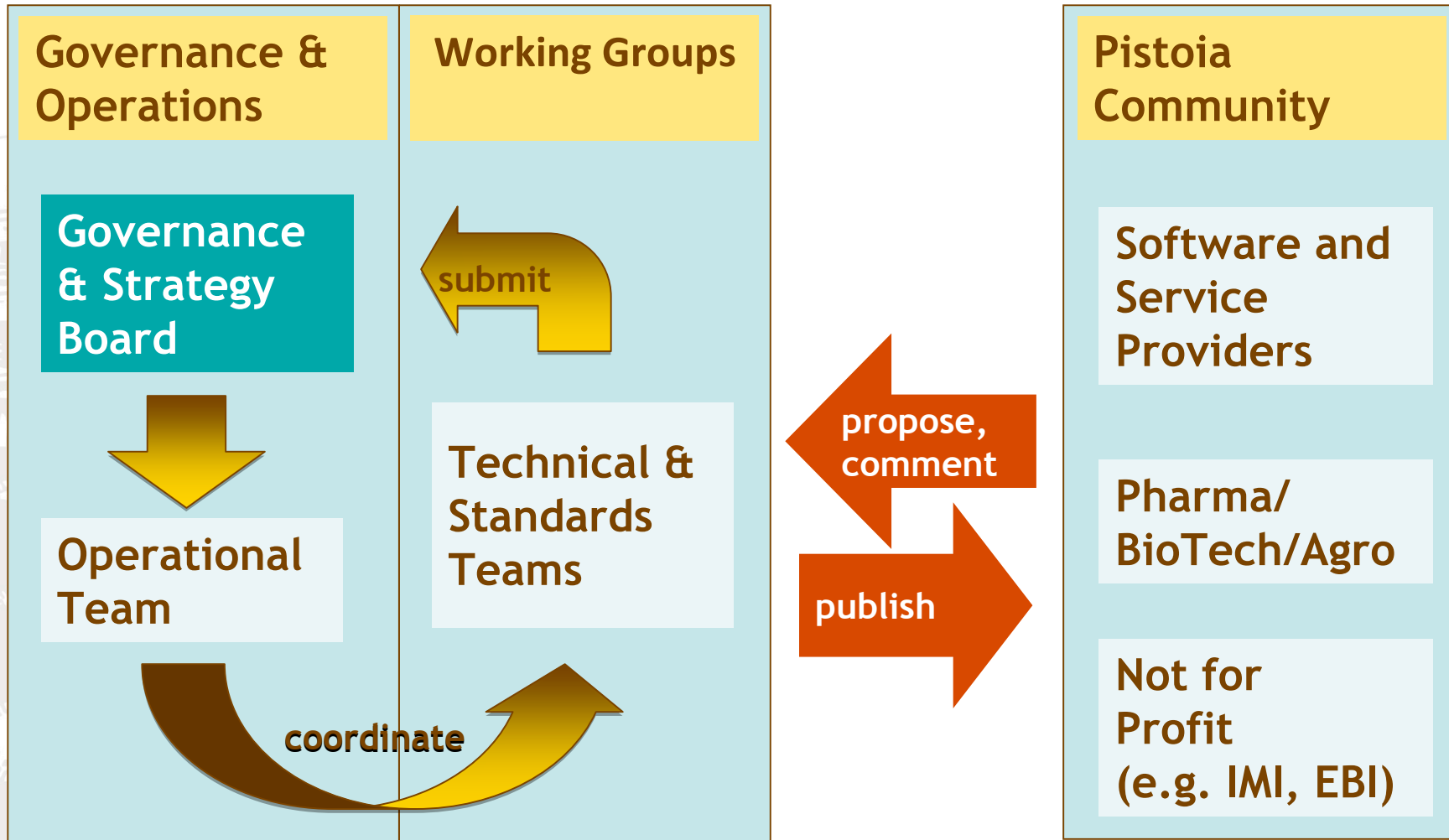Non-Voting Members (vendors, academics)

Officers

President, Secretary, Treasurer, PM, Communications

Technical Committee

Develop standards
Submit standards to Gov & Strat Board
Finalize standard and publish
Observe use of standard and evolve if necessary

Pistoia Standards Process

Pistoia Alliance

Governance & Operations	Working Groups
Governance & Strategy Board	submit
Operational Team	**Technical & Standards Teams**

coordinate

propose, comment

publish

Pistoia Community

Software and Service Providers

Pharma/ BioTech/Agro

Not for Profit (e.g. IMI, EBI)

Pistoia Domains & Working Groups

Technical Committee

•Approves WG Standards publication
•Ensures consistency of standards across Pistoia

Domain Steering Groups

•Steers Working Groups
•Creates remit for Working Group
•Consistency across Science domain

Pistoia Science Domains

Enabling Domains

Chemistry Data Services	Biology Data Services	Knowledge Services
A B	C D	E F

Working Groups WGs

A= ELN Query Services

C= Sequence Services
....

Application Integration

G   H

Vocabulary

Visualisation

Workflow

Others - TBD

Pistoia Alliance
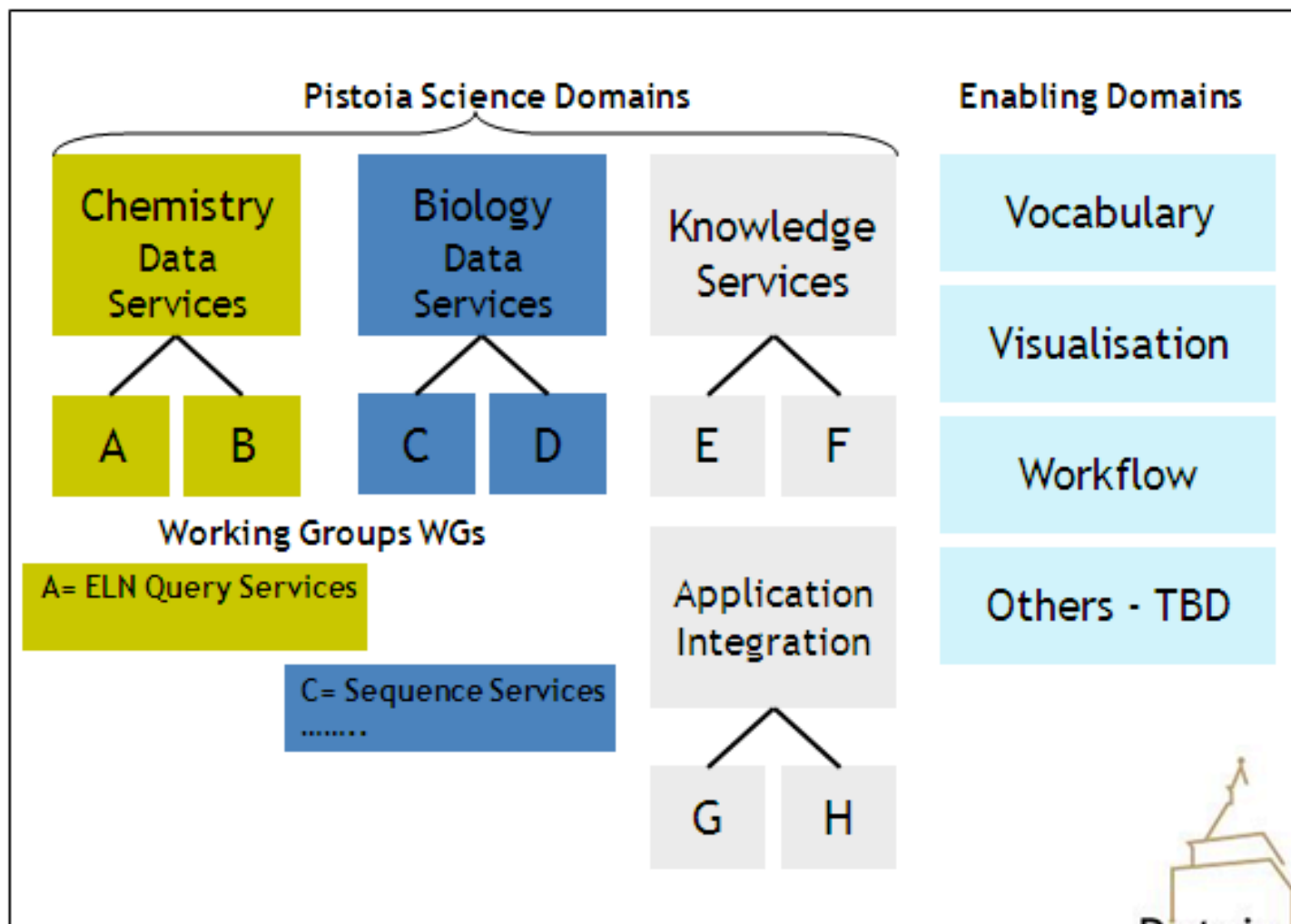
Pistoia Domains
Focused on business workflows/supply chains

Enabling

Vocabulary

Visualisation

Workflow

Others

Knowledge and Information Services

Application Integration
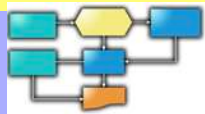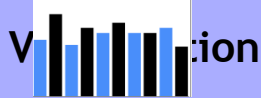
Biology Data Services

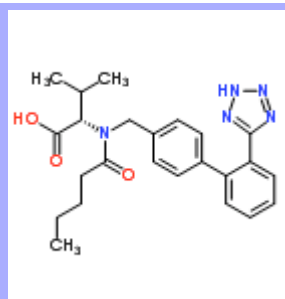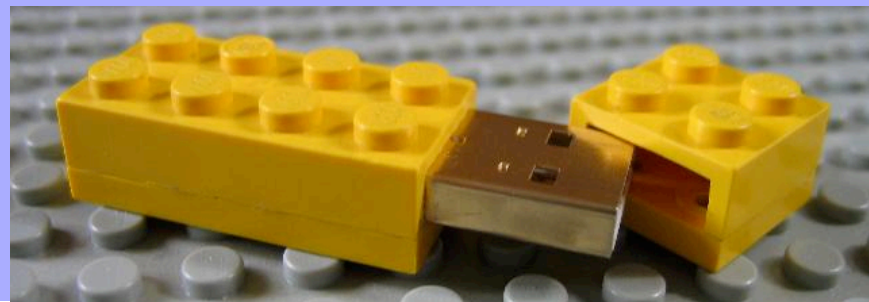**Chemistry Data Services**

Translational Data Services

Pistoia Alliance

Pistoia Domains
Focused on business workflows/supply chains

Enabling

Knowledge **formation**

Capturing Demand using Domain Framework

Pistoia Alliance

Need a disease ontology

Enabling

How to enrich literature with semantic tags

Vocabulary

Knowledge and Information Services

Visualisation

What is this raw data? How did you calculate the summary?

Workflow

How to show chemical structures in a wiki

Application Integration

Biology Data Services

Chemistry Data Services

Translational Data Services

How do I describe the sample I'm shipping to you?

How do I link to clinical systems?

Capturing Demand using Domain Framework

Pistoia
Alliance

Need a disease ontology

Need a disease ontology

How to enrich literature with semantic tags

How to enrich literature with semantic tags

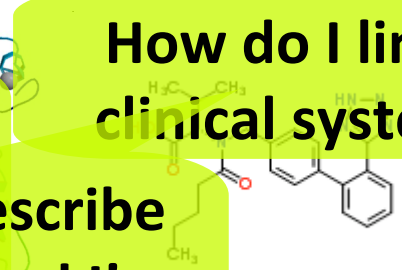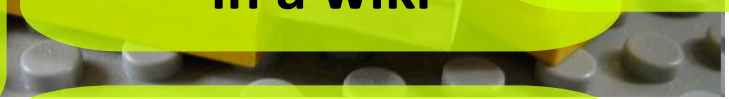What is this raw data? How did you calculate the summary?

What is this raw data? How did you calculate the summary?

How to show chemical structures in a wiki

How to show chemical structures in a wiki

How do I describe the compound I'm shipping to you?

How do I describe the compound I'm shipping to you?

How do I link to clinical systems?
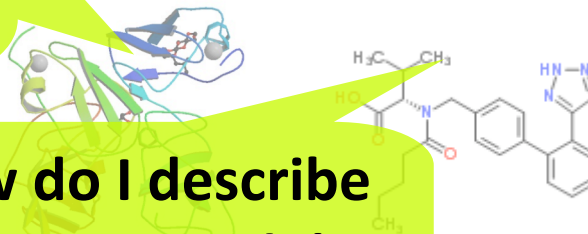
How do I link to clinical systems?

Pistoia Current/Emerging Activities - plotted
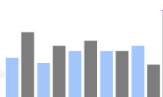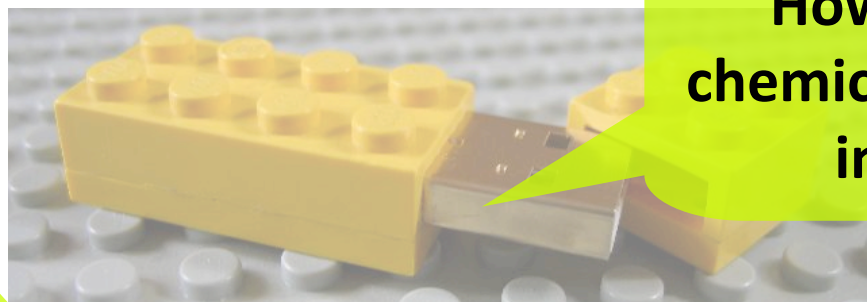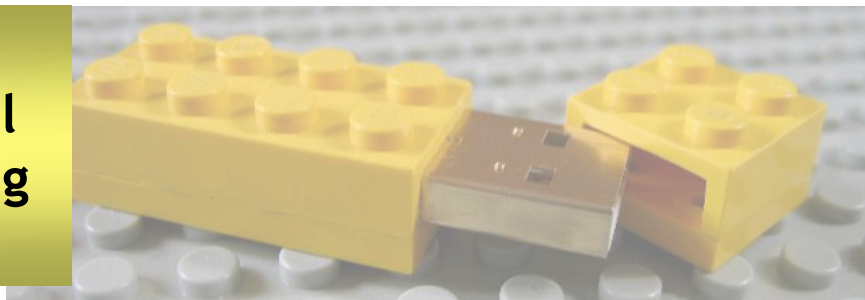
Vocabulary Services

Disease Knowledge Services (SESL)

Chemical Rendering

Sequence Services

ELN Query Data Services
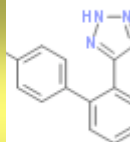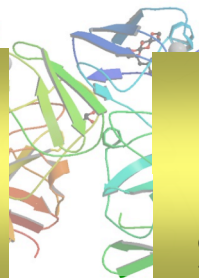
Domain Example: Initiatives in Biology Standards

Pistoia – Curzon Initiative
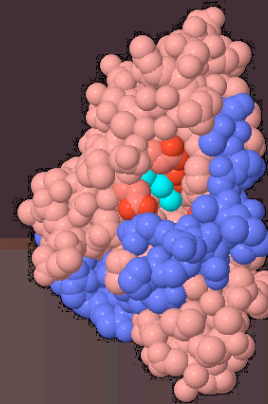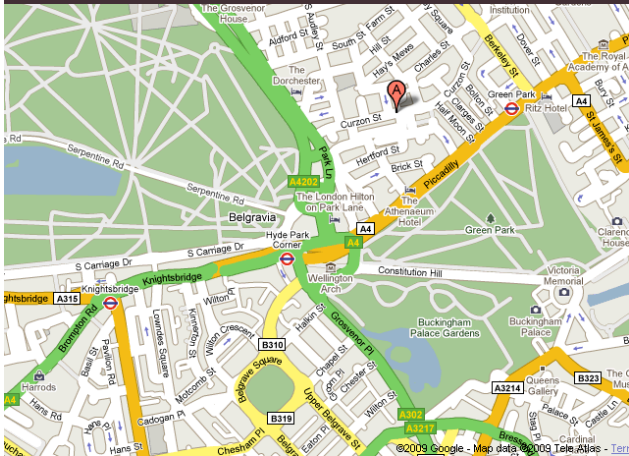
Pistoia – Curzon Initiative

- Biomedical Knowledge Integration
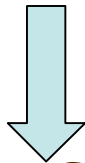
- External Bio-Services Standards

AstraZeneca	GSK	Pfizer
Ian Dix (AZ)	Mike Barnes (GSK)	Cory Brouwer (Pfizer)
Niklas Blomberg (AZ)	Chris Larminie (GSK)	Bryn Williams-Jones (Pfizer)
Nick Lynch (AZ)	Ashley George (GSK)	Lee Harland (Pfizer)

2 key needs in Biomedical KM

1. Semantic standards in information supply chain (commercial & academic)
   - Standardisation of analysis tool interfaces
2. Push model for knowledge access
   - Information awareness, access & mapping is key activity for internal information professionals
   - Can we move to a model where content owners/generators can 'advertise' content assets?

Building an open services infrastructure

SERVICES

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

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

Standards
Ontology/taxonomy;
Minimum information guide;
Dictionaries; Interchage mapping

Data
Targets; Chemistry; Pharmacology;
Literature; Patents

*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*

Pistoia Alliance

CURZON STREET STATION

Biomedical Knowledge Service Framework

Multiple Consumers

Target Dossier

Compound Dossier

Disease Dossier

Network Viz

Knowledge Applications

Open Stds

Consumer Firewall

Supplier Firewall

Service Layer

Assertion & Meta Data Mgmt

Transform / Translate

Integrator

Std Public Vocabularies

Business Rules

Common Service Broker

Content Suppliers

Db 2

Corpus 1

Db 3

Db 4

Corpus 5

Effort required to fit DBs to service layer

Pistoia Alliance

Pistoia Alliance:
Chemistry Domain

WORKING GROUP UPDATE:
ELN QUERY SERVICES

Pistoia Alliance

ELN Query Services Working Group

- ## Purpose
  - To increase the quality of scientific decision-making through superior exploitation of integrated cross-domain knowledge

- ## Goals
  - To define a standard for an ELN data mart and its query services, encompassing information from multiple domain sources, whether internal or externally derived, and independent of any one technology provider
  - Enable knowledge mining and alerting, including a published API to enable data exploitation by external systems

ELN Query Services: Progress

- Kick off meeting April 2009
  - Scope identification
  - Requirements definition and consolidation

- Progress
  - Biweekly meetings held for Interface Discussions
  - User story collection and consolidation
  - Active use of Ning Site for communication
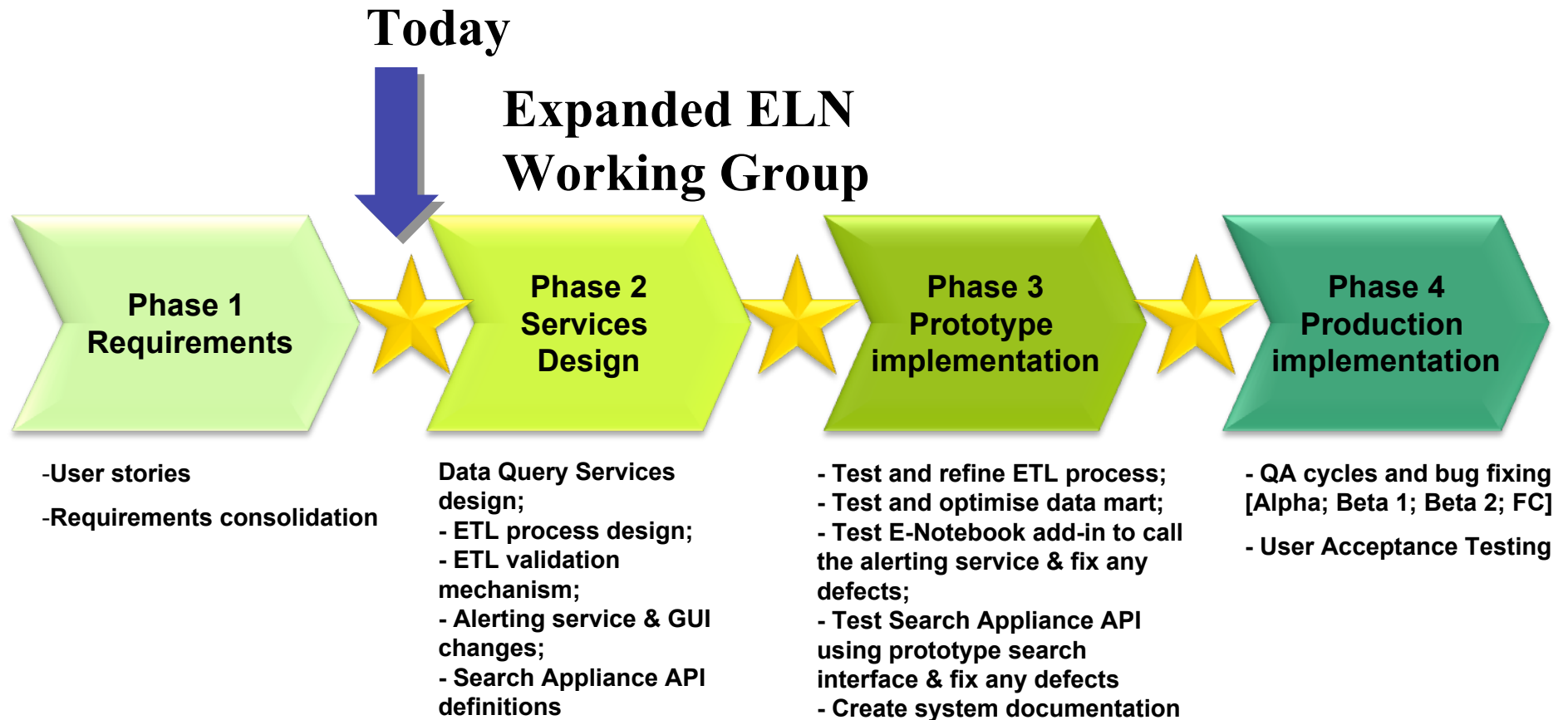
ELN Query Services: Scope

- Define and publish high-level foundation design principles for an ELN data mart and its query services

- Design and implement a prototype version of a synthetic chemistry ELN data mart using information from the Discovery Chemistry workflow

- System independent query interface & tools accessing via a published API

- Expand the data model to incorporate data from ELN sources across the life science space - biology, pharmaceutical sciences and analytical sciences
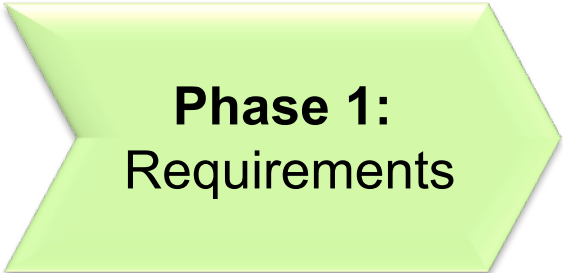
ELN Query Services:
Logical breakdown of project

Today

Expanded ELN Working Group

| Phase 1 Requirements | Phase 2 Services Design | Phase 3 Prototype implementation | Phase 4 Production implementation |

-User stories

-Requirements consolidation

Data Query Services design;
- ETL process design;
- ETL validation mechanism;
- Alerting service & GUI changes;
- Search Appliance API definitions

- Test and refine ETL process;
- Test and optimise data mart;
- Test E-Notebook add-in to call the alerting service & fix any defects;
- Test Search Appliance API using prototype search interface & fix any defects
- Create system documentation

- QA cycles and bug fixing [Alpha; Beta 1; Beta 2; FC]

- User Acceptance Testing

Defined tollgates between phases

ELN Query Services: Requirements

Phase 1:
Requirements

- To provide the ability to rapidly and flexibly search ELN experiment information
- Readily accessible from within the ELN and also to external systems or services
- Enable comprehensive reporting based on ELN data
- Alerting capabilities to provide added-value information and to optimise decision making
- To be able to segment the data services based on security roles
- Allow drill-down to experiment details from resultant hit set
- Built on flexible service-oriented architecture
  - Web services interfaces for querying and retrieval

ELN Query Services: Phase 1
Personas - Roles

Name	Persona
Anna	Chemist works in lab 80% of time preparing compounds
Bert	CRO chemist (external)
Cathy	Lab Manager works in lab 20% of time.
Dave	Alliance/Collaboration manager
Enid	Stores person
Fred	Biologist
Grace	General user
Hector	Analyst (physical chemistry, spectroscopist)
Irene	Department head, does not do any laboratory work
John	Records Manager

ELN Query Services: Phase 1
Story Types

Story type	Purpose	Goal
Ad hoc chemistry query	An ELN user submits a query using an ELN search interface. The ELN Data Mart returns the results which are rendered in the results table. A non-ELN user submits a query using a web-search interface. The ELN Data Mart returns the results which are rendered into HTML.	To provide faster searching of the Med Chem ELN via its built-in search interface To provide rapid searching of the Med Chem ELN data for non-ELN users
Interface derived query (Search API)	An external system or service submits a query to the ELN data mart API. The query is executed and the required data extracted. The results are returned to the external system or service	To provide access to ELN data from an external system or service.
Metrics/Reporting	A manager executes a stored query (providing variable parameters where necessary). The ELN Data Mart returns the results which are then rendered using a pre-defined template to generate a formatted report. Could include both on demand and batch based reporting.	To provide on demand reports for monitoring targets, measuring value and planning future strategy
Live Alerting	A user action triggers a stored query within the E-Notebook based on the data entered. The ELN Data Mart returns related data which is displayed to the chemist.	To provide added-value information during the day-to-day use of the Med Chem ELN to optimise decision making

ELN Query Services: Phase 1
User Story Collection

- Working group → Consolidation of Stories

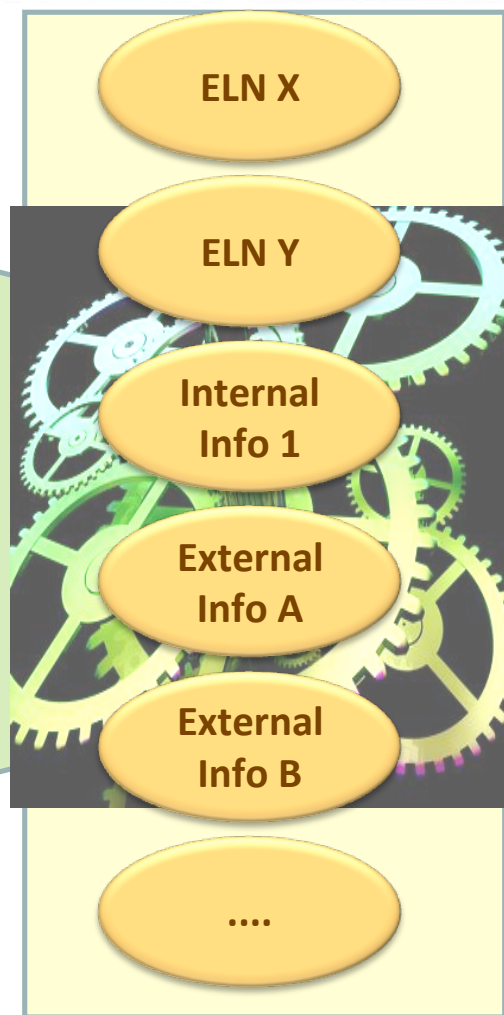| Type of Story	Indication of story such as ad hoc query, interface drive, metrics/reporting, live alerting, security
User Story detail	The business based description of the process the user wishes to initiate.
User story input detail	The inputs generated for successful outcome
User story output detail	The results of the data mart action returned to the user
Requirements detail	Extraction of specific data mart requirements based on the stated use input and system output
Other implications	Things that need to be considered further or need investigating to enable clear requirements to be stated.
Data entities input list	The primary data entities the user story will call on in standardized form.
Data entities output list	The primary data entities the user story will call on in standardized form.
User feedback on story (includes non functional stories)	What next once the user receives results, and how is that data consumed and used

ELN Query Services: Phase 1
User Story Collection
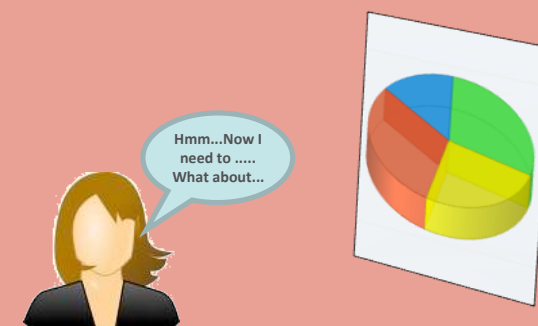
ELN Query Services: Phase 1
User Story Example: US_AHQ_001

User Story detail	Anna wishes to find reactions and data for a particular conversion
User story input detail	Anna enters a reaction/transformation of interest and selects either required retrieval fields or expects to retrieve all the experiment
User story output detail	User receives a list containing all experiments that fit the conversion along with a selected set of fields from the mart in an appropriately sorted sequence (user defined sort) with the selected data available
Data entities input list	reaction:structure transformation:structure (list of fields to be retrieved)
Data entities output list	reaction:structure transformation:structure reaction_id:integer yield: float scientist:text site/location:text experiment_detail:text (other fields requested)

ELN Query Services: Phase 1

User Story Example: US_IDQ_003

User Story detail	Enid is using stock control system needs to identify all reagents that have not been used in the past five years in her annual audit.
User story input detail	Enid enters a list of ids of reagents. The system will need to expand this id into structure and all possible synonyms than use these to search exhaustively for these as used in any reaction over the indicated time period. Both structure and experiment fields will likely need to be searched
User story output detail	list of id's as input with summary usage stats showing how often used and when. Enid id unlikely to wish to drill down to understand usage, she is only interested in finding those reagents that have not been used. If a reagent has a single use close to her time interval she may want to drill down to the scientist to ask them for their opinion on keeping the reagent.
Data entities input list	reagent_id:string date:reange
Data entities output list	reagent_id:string reagent_name:string date:date aggregated usage

ELN Query Services
Next Steps

- **Next Steps**
  - Pistoia Publicaton process
  - Checkpoint and onto Phase 2
  - Active use of Ning Site for communication and collaboration

ELN Query Services

- Search, Performance, Reporting...
- Are you actively and effectively using your ELN data?
- How are you accessing your data?
- Querying across data sources:
  - Internal vs. external data
  - Data driven decision making
  - Balance of data usage
  - Accessibility
  - Combined searches, independent
  - ...

For More Information

- ## Pistoiaalliance.org
  – Vision, Overview, How to Join

- ## Ning Site
  – Pistoia Alliance Members only

An Intentional Segue into Technology Transfer

- Grant or License Access to Pfizer-developed assets to support pre-competitive components of research, development, and medical operations.

- Philanthropic and capitalistic (revenue generation) opportunities.

- Examples:
    - Biological Protocol Definitions/Models
    - Electronic Notebooks (IP Traceability) for Chemistry
    - Computational Surrogates for ADME/T Assays
    - Research Outsourcing/Bidding/Tracking Tools

Discussion

Additional Information on Pistoia

Pistoia Roadmap - Overview

2008	2009

Strategy

Early Build

Biology Awareness

Network/Pharma Awareness

Externalisation

Portfolio build

Extended Strategy Build

Setup

Governance Model

Governance Board and Officers

Enrolment - priority

Normal

Legal Setup

Finance setup & budget planning

Operating processes

Communications

Programme

Lhasa

Biology/Disease Area Programme

ELN Data Mart

Chemical Renderer

Decision to create Pistoia as Not for profit

Official Launch

Review

Open Collaboration—Process

Problem	Demand	Request		Develop	Solution Delivery

Collaborating Companies input

↓

Share outside company walls	→	Elaborate Common Needs	→	Decide to collaborate	→	Decide how to develop	→	Decide how to develop together

Prepare 3rd party → 3rd party hosting

Prepare tender/RFP

Open Collaborations opens opportunities earlier

Open Collaboration Interface ← **Open Collaborations opens opportunities earlier** ← Traditional Company Interface

Pistoia Alignment

Industry Standards Groups

•Do we need an advisory board? CDISC

Members

Advisory Board?

Board of Directors

Pistoia Alliance

Technical Committee

Domain Classification

Domain Steering Groups

Industry Challenges

Disease

Pharmacology

DMPK

Safety

Enabling

Vocabulary

Visualisation

Workflow

Others

Knowledge and Information Services

Application Integration Standards

Biology Data Services

Chemistry Data Services

Translation Data Services

Working Groups