The importance of chemical identifier standards in the pharmaceutical industry

Colin Wood

RSC Chemical Information and Computer Applications Group (CICAG)
21 October 2014

Abstract

The importance of chemical identifier standards in the pharmaceutical industry

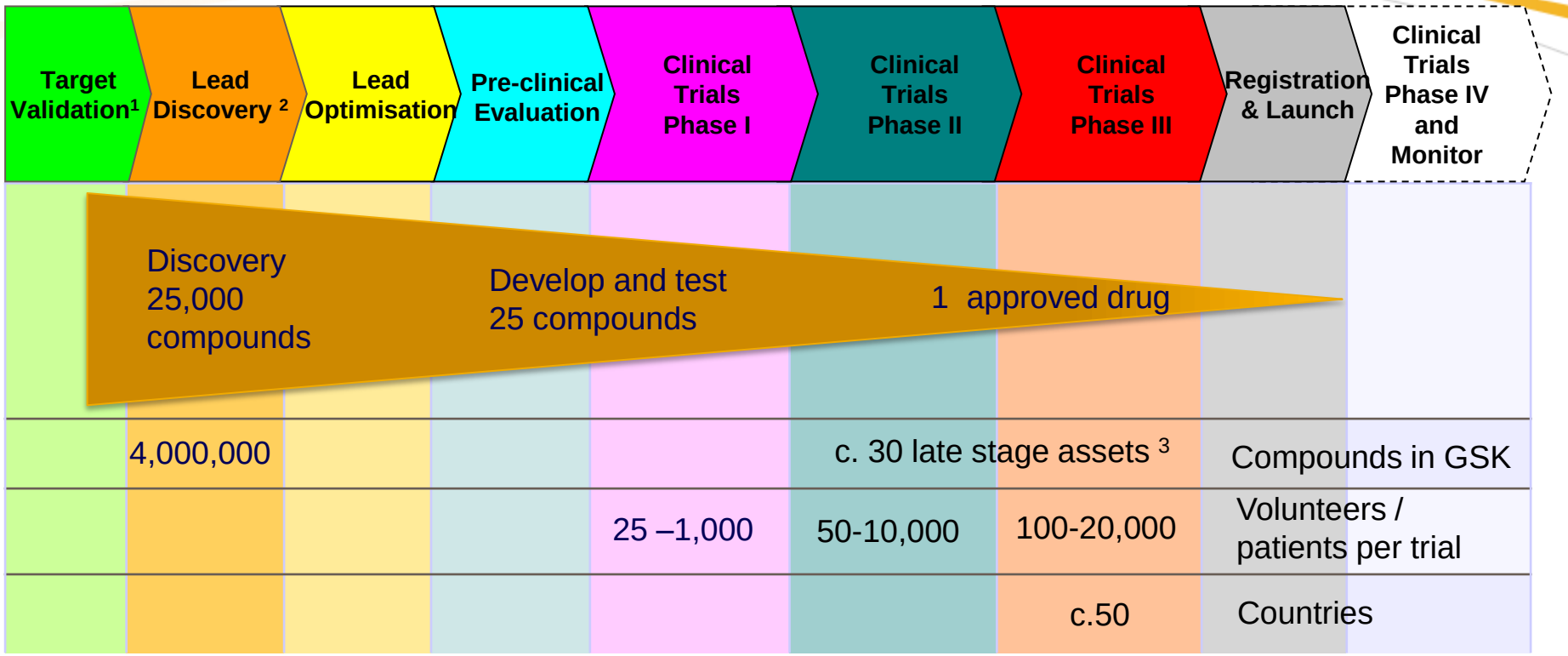
The use of specialised chemical registration systems to generate unique identity for molecules is well established within the Pharmaceutical industry, however there is little recognition that this is a form of master data management technology. Used appropriately these technologies can be used to provide unique identification of substances across the full lifecycle of a product. There are however many challenges, including human factors, technology issues and a lack of established standards across the industry. Colin will show how GSK has solved many of these challenges and will outline the gaps and issues that make this a complex and ongoing problem.

GSK - Setting The Scene

- **GlaxoSmithKline** is one of the world's leading pharmaceutical and healthcare companies.
- We have a challenging and inspiring mission to improve the quality of human life by enabling people to **do more, feel better and live longer**.
- With a firm foundation in science, we **discover, develop, manufacture and distribute prescription medicines**, vaccines and consumer healthcare products.
- Headquartered in the UK with major business operations in the US, we employ more than 99,000 people in 120 countries.
- The data generated within the **Research and Development** function can vary from high volume early screening data, to complex manufacturing methods or multinational clinical trial data.

The Drug Development Roadmap

Every hour we spend > £300,000 to find new medicines at our 24 research sites in 11 countries.



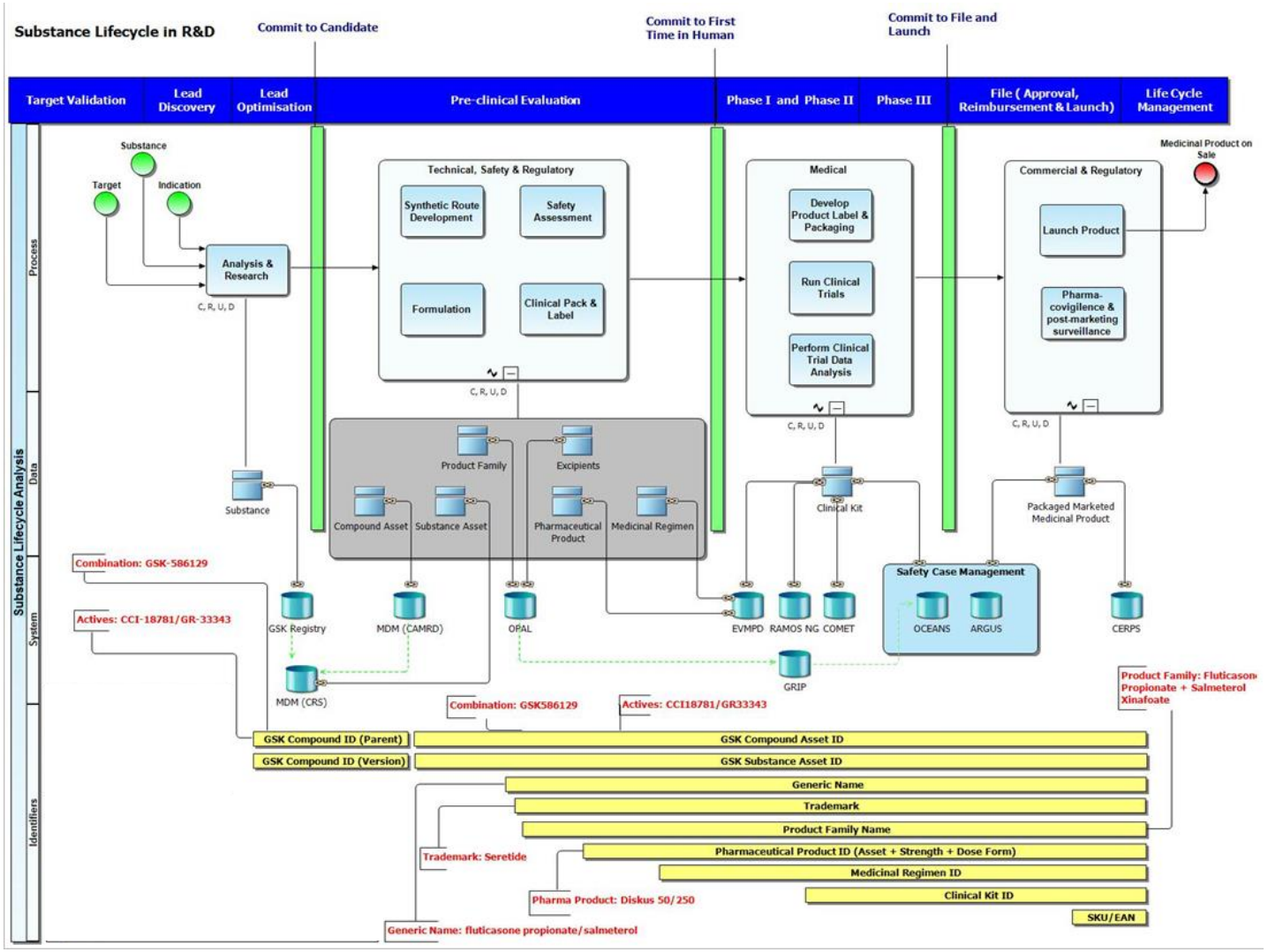
Development takes 10-15 Years and costs up to £1 billion per Drug[3]

Data needed for at least 30 Years

1. A target is a structure in the body that contributes to the development of disease or the symptoms of that condition, for example an enzyme.
2. Lead compound ; chemicals or natural compounds that might inhibit or enhance the target's activity.
3. From GSK annual report 2011.

Substance Lifecycle in Pharma R&D

Substance information spans the full R&D process

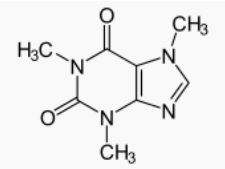


Master Data Management

Why Chemical Registration is a form of MDM Technology

- In *business*, **master data management (MDM)** comprises the processes, governance, policies, standards and tools that consistently define and manage the critical data of an organization to provide a single point of reference.

- In *computing*, an **MDM tool** can be used to support master data management by removing duplicates, standardizing data (mass maintaining), and incorporating rules to eliminate incorrect data from entering the system in order to create an authoritative source of master data. Master data are the products, accounts and parties for which the business transactions are completed *[Source: Wikipedia]*



Validate
- Standardise chemical representation
- Validate using business rules

Match
- Match to existing records
- Confirm matches

Register
- Generate Identifier
- Publish

Chemical Registration

Name : Dr A. Customer
Office Address: GSK Medicines Research Centre, Gunnels Wood Road, Stevenage
Phone: +44 1763 644000

Validate
- Standardise customer data
- Apply business rules

Match
- Match to existing records
- Confirm matches and merge

Master
- Master Record and Generate ID
- Publish

Master Data Management System

Semantics and Human Factors

Meaning is important when it comes to Master Data

- Depending on context the following terms could mean the same thing, or something entirely different
  - Molecule, Parent, Version, Salt, Compound, Substance, Active Substance, Asset, Drug Product, Drug Substance, API, Active Ingredient, Active Product Ingredient, Product, Pharmaceutical Product, Medicinal Product, New Therapeutic Entity, New Molecular Entity, New Biological Entity, New Chemical Entity, New Active Substance, NCE, NAS, NME, NTE, Hit, Lead Compound
- There is significant variation across the industry
  - Pistoia analysis on industry terminology April 2008

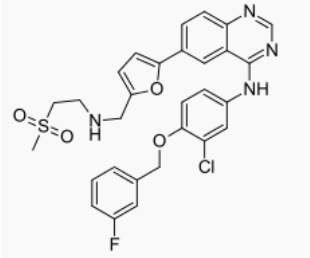
Domain Entity	AstraZeneca	GSK	Novartis	Pfizer	PubChem	FDA
Compound	Compound	Parent	Substance (Parent)	Parent	Compound	Molecular Entity
Substance (CompoundSubstance)	Substance	Version	Salt	Salt or Compound	Substance	Substance or Active Ingredient
Batch	Sample	Preparation or Lot	Batch	Batch	N/A	N/A
Lot	Sample	Sample	Sample		N/A	N/A

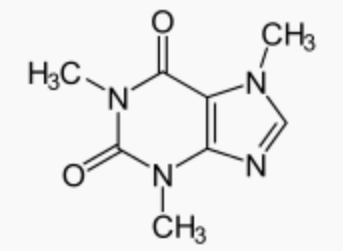
Semantics and Human Factors - 2

Even the choice and use of identifiers can create issues

- GSK has a very simple format for compound identifiers
  - Company prefix ,Numeric Identifier, Salt Code (No salt code for the parent)
  - E.g. GSK1234567A
- What do the prefixes mean?

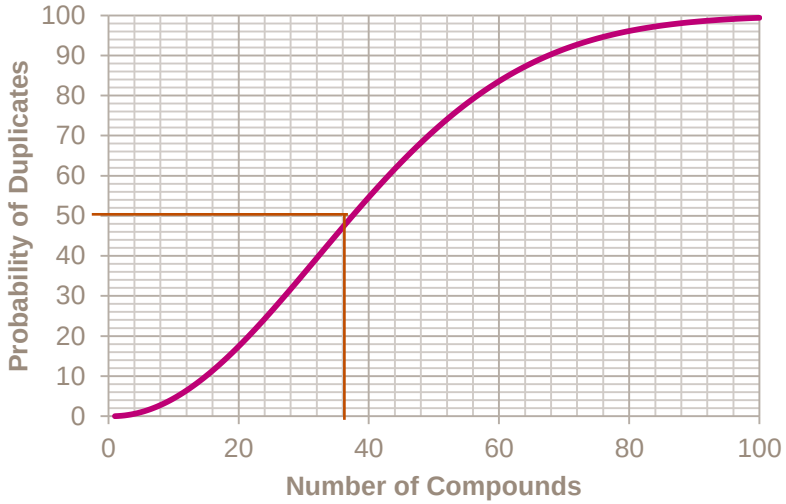
GW572016
(Lapatinib)





AH14925
(Caffeine)

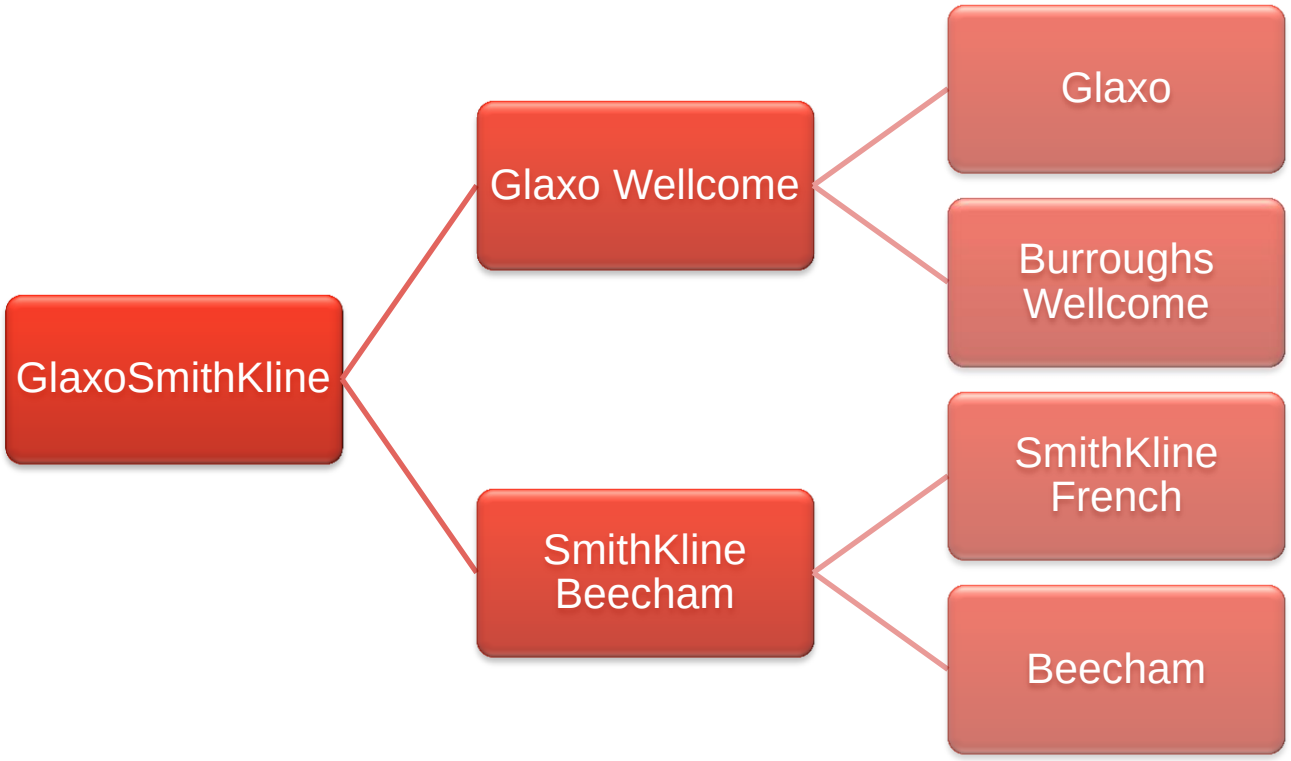
- GSK1234567 is lengthy and hard to remember
  - Is 567 sufficiently unique for informal communication?
  - If you have more than 38 compounds, then statistically no



History of Mergers and Acquisitions

We carry the history

- All pharma's have a long history of mergers



We accumulate compound identifiers spanning the history of the company

AH 14483
AH-14483
BRL-9073
CCI140
SKF-17507
SKF-882
SKF-90104

Paracetamol

AH-14925
BRL-7867
CCI3994
SKF-6053

Caffeine

We also continually acquire companies and establish collaborative arrangements.

What are we really identifying?

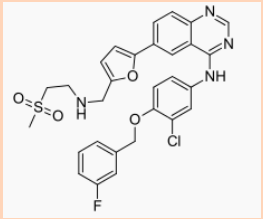
When do we change from a substance to a product?

- We understand what we are identifying when there is a single active
  - A single compound number which we can trace throughout the lifetime of a product
- What are we identifying in these cases?
  - GI265235/GR109714/509U81 (abacavir/lamivudine/zidovudine)
  - CCI18781/GR33343 (fluticasone propionate/salmeterol)
- At some point we start to use Trade or Brand names
  - Some products can have multiple names
- At some point in the development lifecycle we begin to identify “Product Families”
  - It’s convenient to use compound numbers to identify, but this is a different entity

We have started to adopt the following definitions

The following data entities span multiple R&D systems and processes

- Substance Asset
  - All medicinal products consist of substances; these substances can be active ingredients, excipients, or materials in packaging materials. The term asset indicates that the substance has been candidate selected or is used within one of our products.



- Product Family
  - A group of assets that share the same active ingredient or combination of active ingredients, primary large molecule, or delivery device. Any unique combination of active ingredients that can be used to create an actual product.

- Pharmaceutical Product
  - A finished dosage form of a specific product strength and/or dosage, that contains an active pharmaceutical ingredient generally, but not necessarily, in association with inactive ingredients. The term also includes a finished dosage form that does not contain an active ingredient but is intended to be used as a placebo and competitors' product used as a comparator.
    [IDMP - Pharmaceutical Product] Qualitative and quantitative composition of the pharmaceutical product as administered to the patient in line with the regulated product information.

- Medicinal Regime (or Medicinal Product)
  - One or more pharmaceutical products prescribed together to treat a single disease state. [IDMP – Medicinal Product] Medicinal Product] Any substance or combination of substances, which may be administered to human beings or animals for treating or preventing disease, with the view to making a medical diagnosis or to restore, correct or modify physiological functions. Medicinal products may relate conceptually to one or more "pharmaceutical products" as part of an authorised treatment regimen

IDMP = Identification of Medicinal Products

External Generic Naming Authorities

STATEMENT ON A NONPROPRIETARY NAME ADOPTED BY THE USAN COUNCIL:

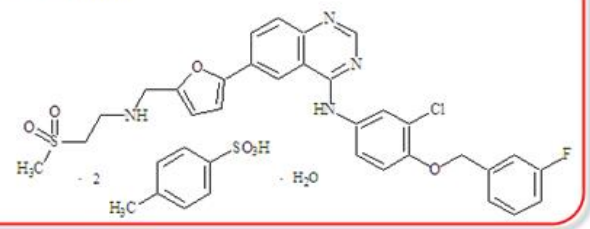
USAN	LAPATINIB DITOSYLATE
PRONUNCIATION	la pa' tin ib
THERAPEUTIC CLAIM	anti-neoplastic agent

CHEMICAL NAMES

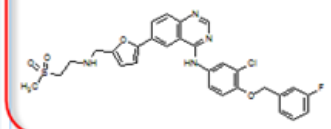
1) 4-Quinazolinamine, *N*-[3-chloro-4-[(3-fluorophenyl)methoxy]phenyl]-6-[5-[[[2-(methylsulfonyl)ethyl]amino]methyl]-2-furanyl]-, bis(4-methylbenzenesulfonate), monohydrate

2) *N*-[3-chloro-4-[(3-fluorobenzyl)oxy]phenyl]-6-[5-[[[2-(methylsulfonyl)ethyl]amino]methyl]furan-2-yl]quinazolin-4-amine bis(4-methylbenzenesulfonate) monohydrate

STRUCTURAL FORMULA



lapatinib



Latin	lapatinibum
French	lapatinib
Spanish	lapatinib
Russian	лапатиниб
Arabic	لابَاتِنِيب
Chinese	拉帕替尼
Phonetic	lapatinib
Molecular formula	C29H26ClFN4O4S
ATC Codes	

National commission(s)	Alternate name(s)
JAN	lapatinib tosilate hydrate

No data found on use of name by national commission(s).

IT Vendors

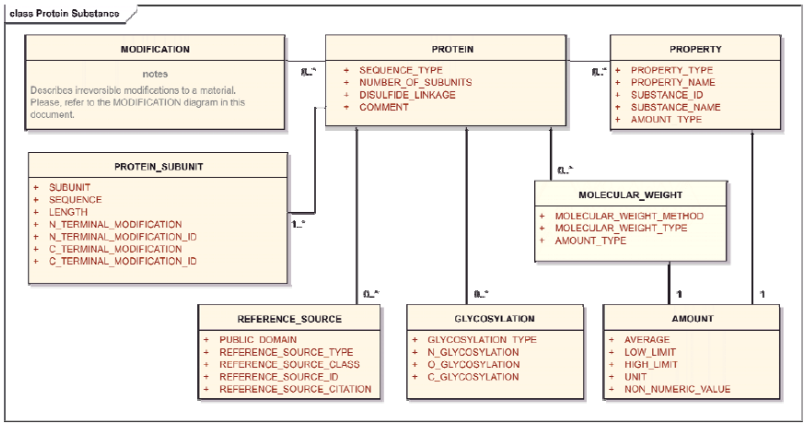
Unfortunately vendors make it hard

- R&D processes are supported by a significant number of specialised software products
  - Most large pharma organisations will use hundreds of applications that reference compounds or products
- Many software products supporting our industry provide a simplistic way of referencing substances or products
  - Simplistic definition of “Product” with perhaps only “Company Code” and “Generic Name” to describe
  - No built in facilities to integrate with external “master” sources
  - Any change to the above identified as an expensive customisation
- Lack of clear industry standardisation contributes to this

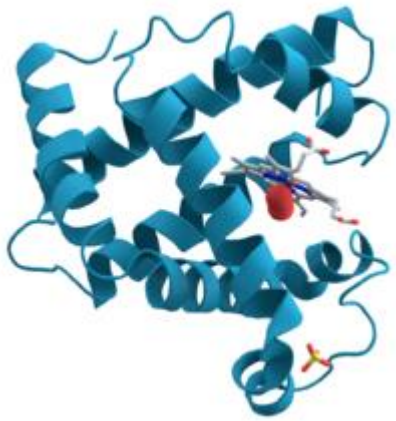
Biological Substances

Bio molecules are of growing importance

- Uniqueness is not a straightforward concept in Biology
  - Biologics cannot be fully defined by their structure. Must also describe parentage, how the substance was made. Genealogy/Lineage matters.
  - The rules for uniqueness change as science progresses

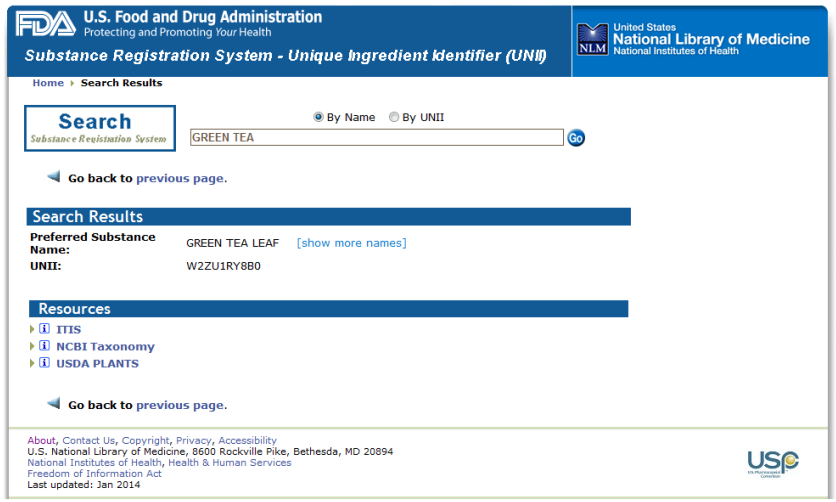


IDMP Models

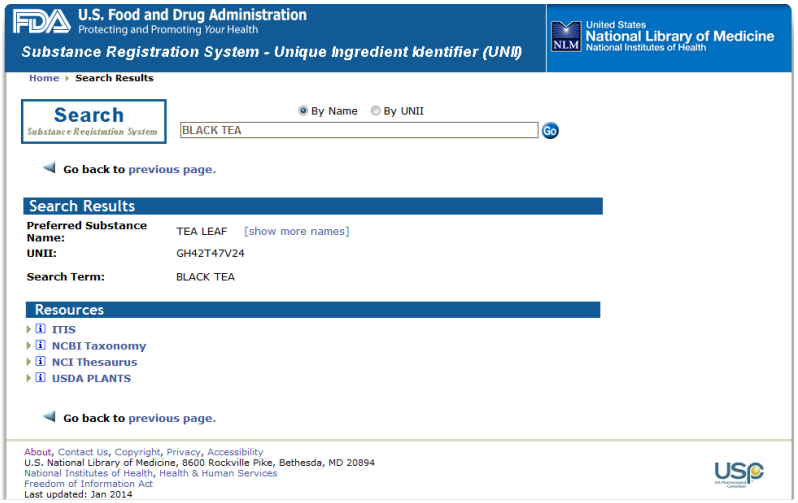


Structurally Diverse Substances

Many consumer products use ingredients such as Green Tea. How do you assign identity?



FDA have generated a substance identifier for Green Tea



Black Tea has a different identifier in SRS

Synonyms

- CAMELLIA SINENSIS LEAF
- CAMELLIA SINENSIS LEAF EXTRACT
- CAMELLIA SINENSIS LEAF EXTRACT [INCI]
- CAMELLIA SINENSIS LEAF POWDER
- CAMELLIA SINENSIS LEAF POWDER [INCI]
- CAMELLIA SINENSIS LEAF [INCI]
- CAMELLIAE FOLIUM
- CAMELLIAE SINENSIS FOLIUM
- CHA YE
- GREEN TEA
- GREEN TEA EXTRACT
- GREEN TEA LEAF EXTRACT
- GREEN TEA [VANDF]
- MATCHA TEA
- MATCHA TEA LEAF
- SENCHA TEA
- SENCHA TEA LEAF
- TEA EXTRACT
- TEA EXTRACT [FHFI]
- THEA SINENSIS [HPUS]
- WHITE TEA LEAF

These are regarded as synonyms

Synonyms

- AE-TEA
- ALLERGENIC EXTRACT- TEA CAMELLIS SINENSIS
- B1623 [LANGUAL]
- BLACK TEA
- BLACK TEA EXTRACT
- BLACK TEA EXTRACT [VANDF]
- BLACK TEA LEAF
- CAMELLIA SINENSIS LEAF, FULLY FERMENTED
- FOOD - PLANT SOURCE, TEA CAMELLIA SINENSIS
- ORANGE PEKOE TEA ALLERGENIC EXTRACT
- TEA
- TEA LEAF EXTRACT
- TEA [MI]

Black Tea has different Synonyms

Source FDA Substance Registration System: National Library of Medicine (NLM)

Common consumer substances

Inorganic compounds - Business rules for the identification of “Metals”

Identification of “Metals”

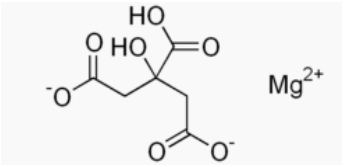
Mg

GI148744

Salts of Magnesium are assigned different identifiers

Other metals can be assigned similar identifiers

Magnesium Citrate



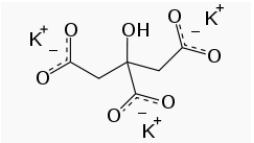
CCI23726AQ

Business rules are set up to recognise organic molecules. This molecule is therefore seen as the Magnesium Salt of Citrate. CCI23726 = Citric acid

Magnesium Hydroxide (Milk of Magnesia)

SKF20259

Potassium Citrate



CCI23726AR

Recognised as the Potassium salt of Citrate. From the same series as Magnesium Citrate!

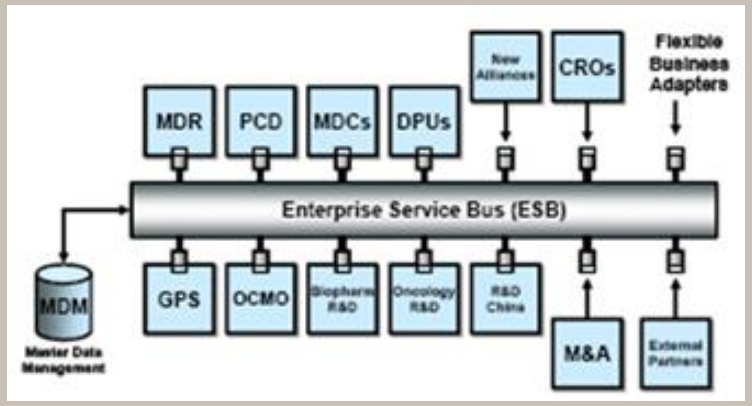
Compound Registry is able to provide unique identity, but the business rules are different

Potentially many different metal salts can be used as substances within Consumer Products

- Magnesium Aspartate (mineral supplement)
- Magnesium Chloride (nutritional supplement)
- Magnesium Lactate (mineral supplement)
- Magnesium Citrate (laxative)
- Magnesium Glycinate (mineral supplement)
- Magnesium Bromide (mild sedative)
- Magnesium Hydroxide (milk of magnesia)
- Magnesium Borate (antiseptic)
- Magnesium Stearate (used in formulation of tablets – common excipient in Rx)
- Magnesium Sulfate (laxative)
- Magnesium Malate (Oral magnesium supplement)
- Magnesium Orotate (Oral magnesium supplement)
- Magnesium Salicylate (antiseptic)

So how do we solve?

Integration and applied business processes



TECHNOLOGY

eLNB

Clinical Trial Tracking

Register Compound

Create Clinical Study

ESB

Compound Registry

MDM

All Compounds

Candidate selected compounds. MDM Validates id's via service calls.

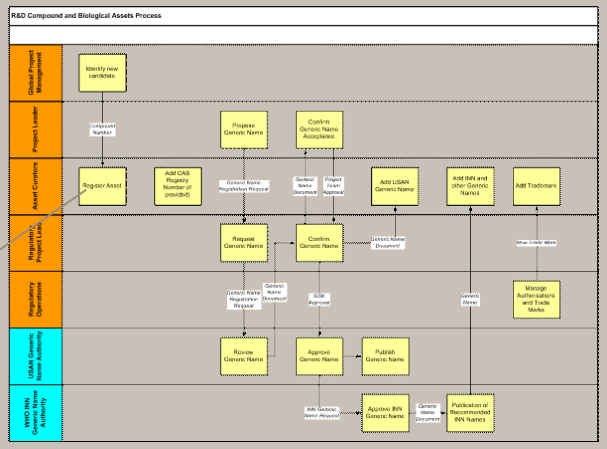
Manual curator process to add assets

Candidate Selection

In-licensing

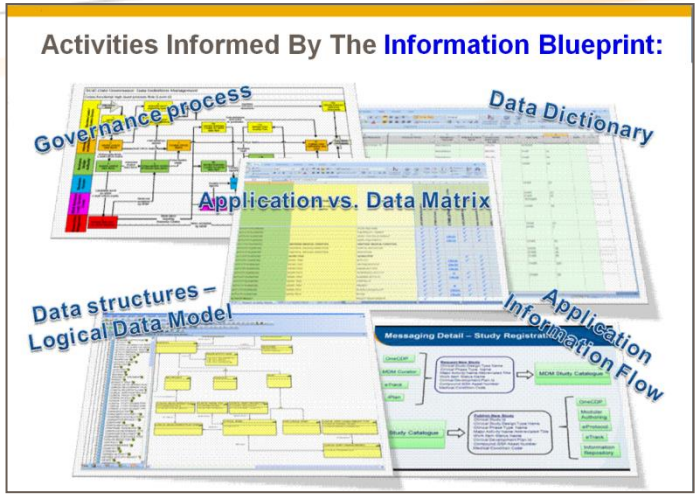
Out-licensing

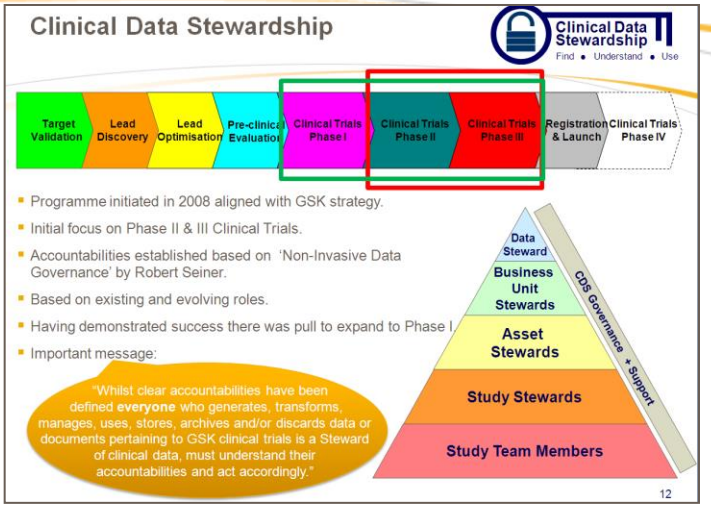
BUSINESS PROCESS

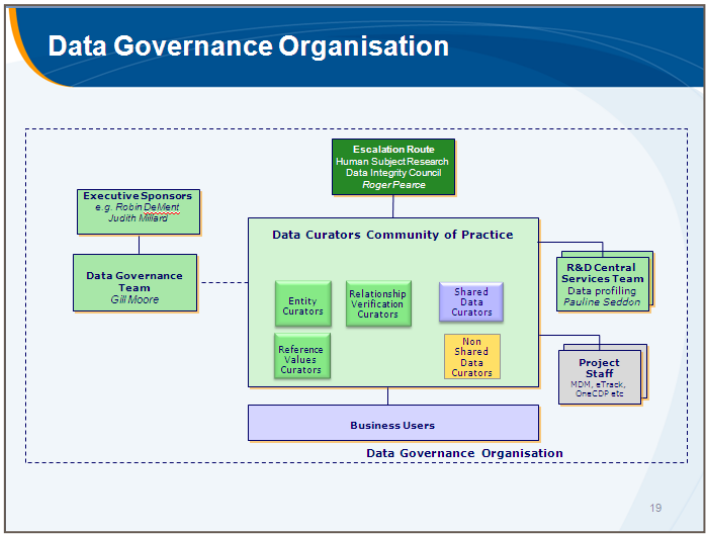


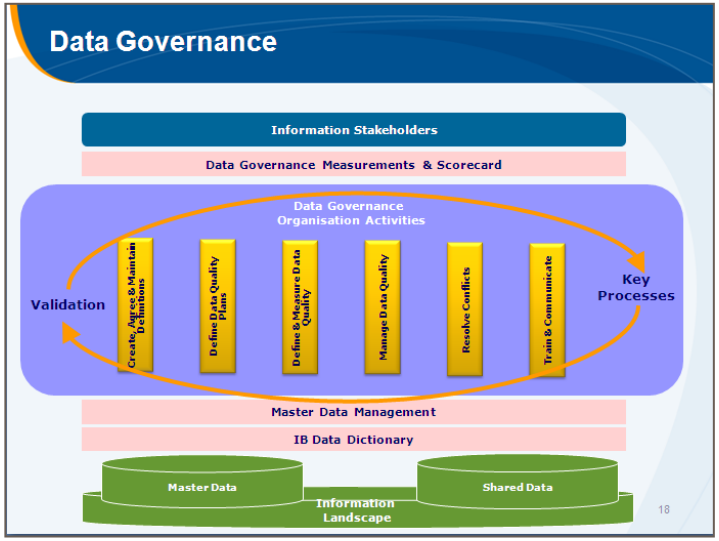
Processes governing creation of new "product families" defined

MDM Governance, Organisation and Information Life Cycle









ISO standards

As an industry we haven't defined or adopted standards

- As an industry we need standards
  - The standards have some issues and the implementation guidance is missing
  - European legislation also creates challenges
- ISO standards for the identification of medicinal products
  - ISO11615:2012 Health Informatics – Identification of medicinal products – Data elements and structures for the unique identification and exchange of regulated medicinal product information
  - ISO11616:2012 Health informatics – Identification of medicinal products - Data elements and structures for the unique identification and exchange of regulated pharmaceutical product information
  - ISO11238:2012 Health Informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on **substances**
  - ISO11239:2012 Health Informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated information on pharmaceutical dose forms, units of presentation, routes of administration and packaging
  - ISO11240:2012 Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of units of measurement

Thank you

do more
feel better
live longer