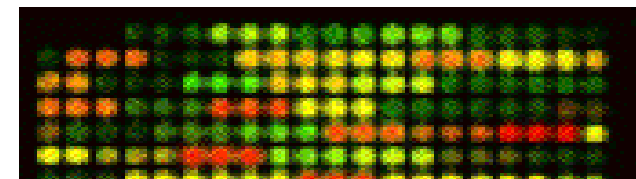


Use of Data Standards and Metadata in Information Exchange

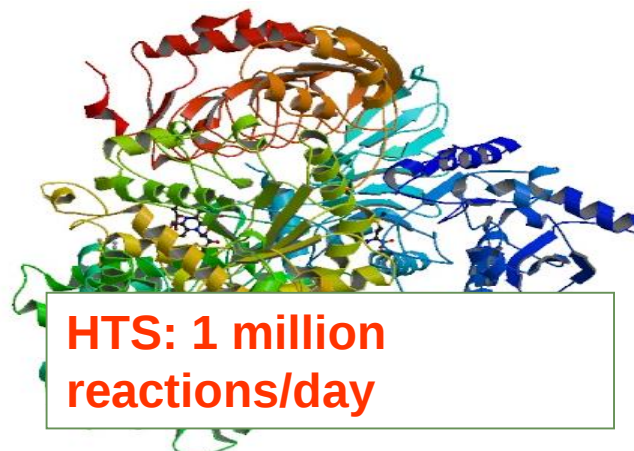
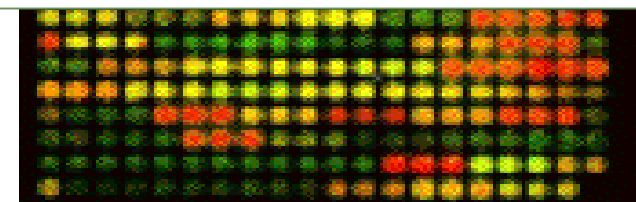
Rachel Uphill
From Big Data to Chemical
Information Meeting

-
- Understanding how the plethora of new data and improved analytical techniques can enhance future innovation and feed the drug-development pipeline
 - Implementing ontologies, standards, strategies and collaborations to enhance products and provide wider value
 - Discovering what your peers are prioritizing as part of their own Big Data strategies

Big Data in Drug Discovery

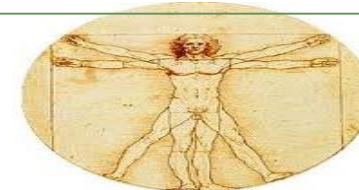


>1 million gene expression profiles

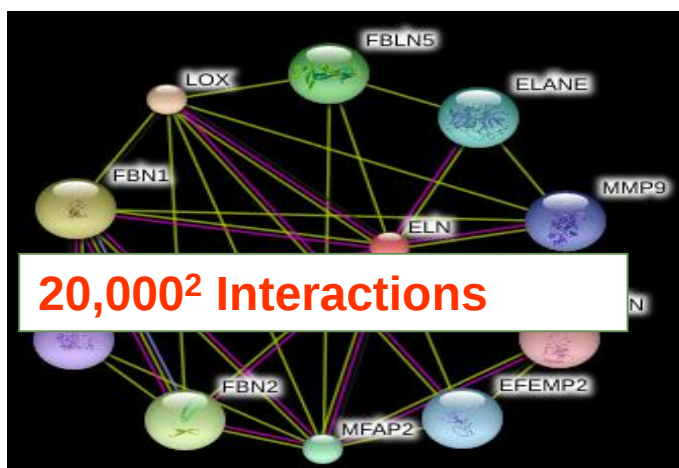


HTS: 1 million reactions/day

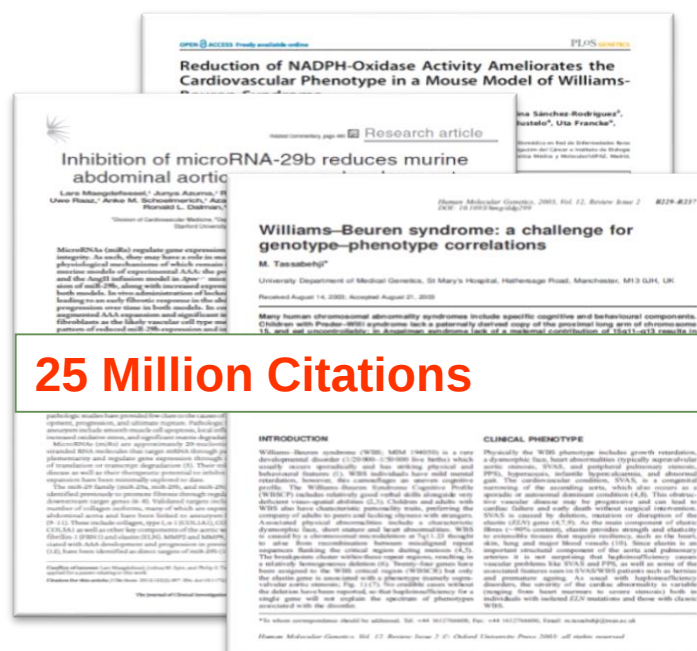
Over 40000 metabolites



Over 16,000 Organisms



20,000² Interactions



25 Million Citations



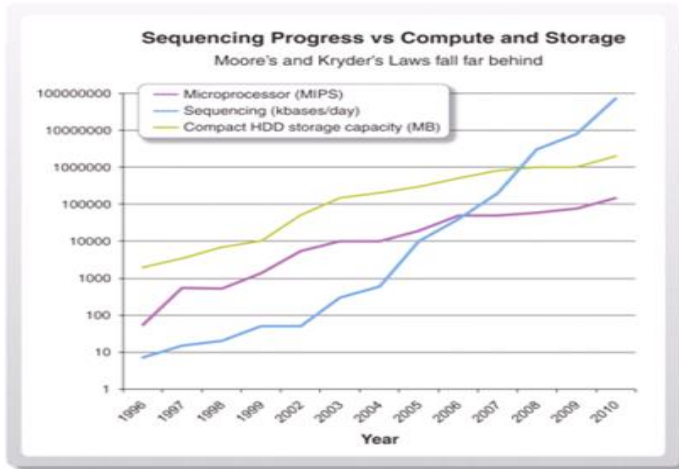
186,000 trials

ClinicalTrials.gov

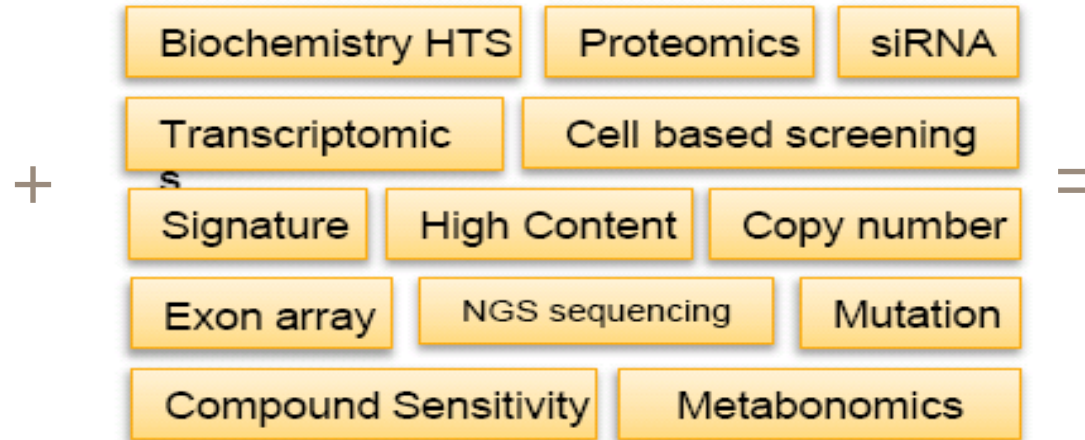
Complex and Disparate



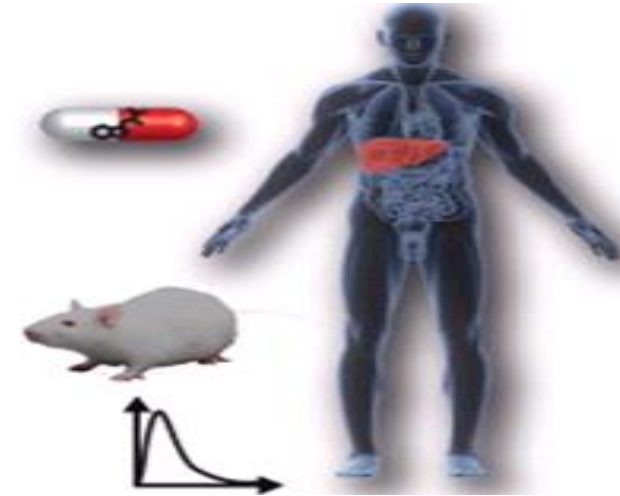
Increasing Data Size



Increasing Data Dimensionality

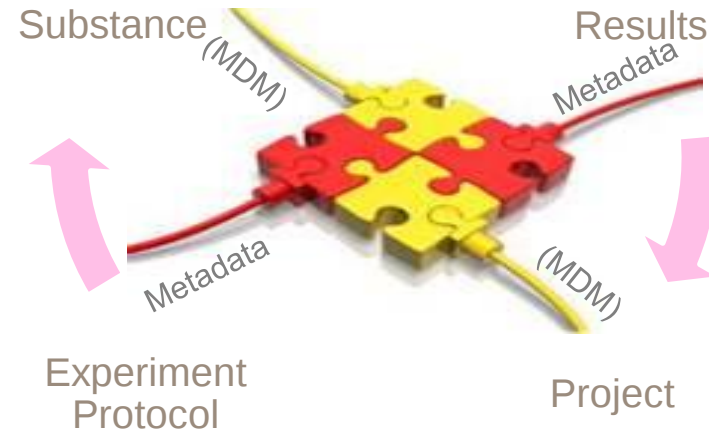


Increasing Complexity



- Data points are growing rapidly and endpoints are unclear
- Disparate datasets
- Data integration through increased use of Contract Research Organisations (CRO's)
- Complex analytics required on ever increasing data

Integrated systems and re-use of information



MDM Delivers one version of the truth across each of the integrated solutions, made easier through the use of standards and ontologies/metadata

Planning and designing our projects and study protocols will be simpler leading to more time to focus on scientific aspects

- Entering information once and making it more accessible across the organisation
- Facilitating enterprise working & decision-making
- Bringing the regions “closer” and collaborating further with our scientists

Vision → Strategy → Plan → Execute → Summary → Endpoint

-
- To maximise the value of Pharma R&D data there are some key disciplines that IT have been investing in:
 - Blue Printing, understand what you've got and what you need
 - If you were building a house you would first start with the plans
 - Stewardship & Governance, who it matters to in the business if it's wrong
 - You'd consult with the planning authorities to make sure everything was in acceptance, if it wasn't you'd get it corrected
 - Quality, how you know it's wrong
 - You wouldn't use a industrial boiler to heat your house
 - Master & Reference data management, single source of the truth
 - You'd check with the land registry to make sure the boundaries of the property were accurate before building an extension
 - Standards, understand what you have been asked for
 - You'd consult the building regulations and make sure all documents that are created containing the required data
 - Search and Analytics, You'd only trust your search results and analytical answers once you knew all these were correct
 - You'd build the chimney, once you'd completed the foundations and structure of the building
-

- **What is an Information Blueprint**

- Documentation of the information landscape illustrating the flow and structure of information at a level of detail appropriate for the audience.
 - Modelling the key business processes, understand the inputs and outputs of those processes and identifying the upstream suppliers of data and the downstream consumers
 - Modelling the high-level information concepts for a broad audience and the detailed data structures for a technical development and data governance audience.
 - Identifying transactional use of data as well as secondary reporting and analytical uses of data.
- Information Blueprints are collections of models created using a number of modelling techniques that document the business processes, data structures and systems landscape.

- **Major benefits of an Information Blueprint**

- Provide a business process based view of data usage and information requirements. Highlights gaps in information provision and enables impact analysis of changes to the information supply chain.
- Provides the framework for assessing and influencing business data quality.
- Facilitates communication within the business and between the business and IT.
- Encourages the development and use of common terminology by clarifying definitions & synonyms.
- Provides the common business understanding for use and re-use across IT systems

- Stewardship ... “formalizing accountability for the management of data resources for a subset of enterprise data.” (1)
- Data Governance ... “the execution and enforcement of authority over the management of data assets and the performance of data functions.” (1)
- Master Data Management ... the people, processes and technology that ensure a single managed view of critical enterprise information.

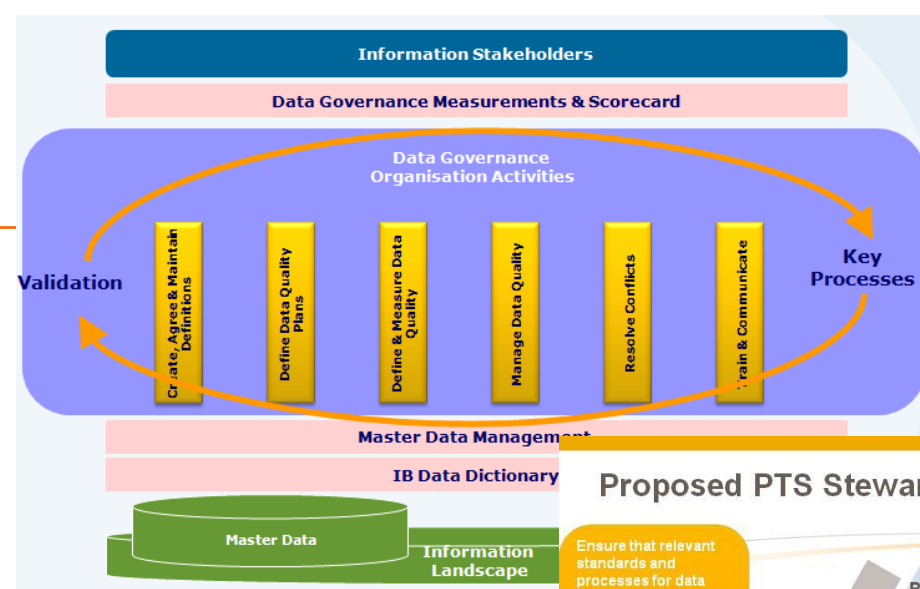
Information has value and purpose beyond a single use, system or report. We need to support this with elements of good Stewardship and Governance.

Data Governance and Stewardship

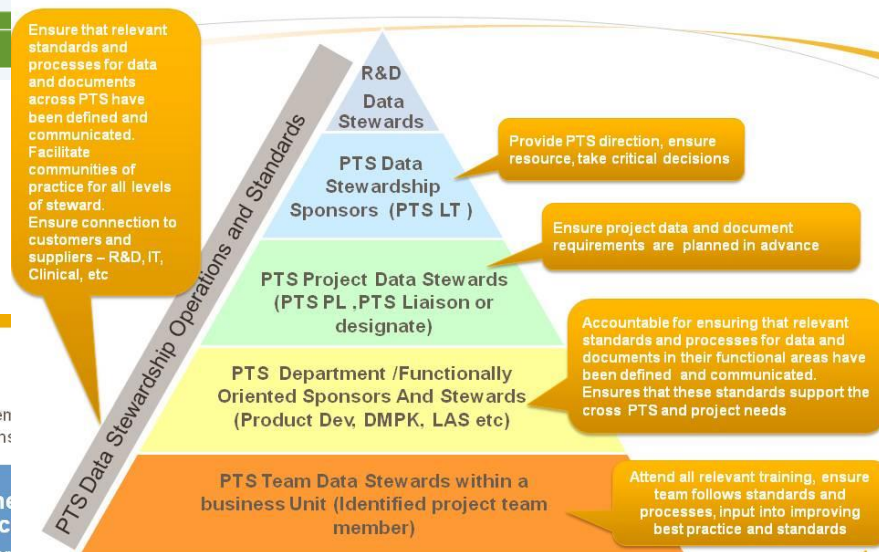


Good Data Stewardship and Governance encompasses:

- **Capability** to effectively find, understand and use data and document
- **Data standards** – the nature of our standards, how we wish to apply them in GSK
- **Business rules and processes** to support the creation and management of data and documents
- **Data disposition** – business rules for versions retained, storage, labelling, tracking of data and documents and **compliance** to these rules
- **Business model** to support data stewardship and increase the capability for data stewardship within the business lines – responsibilities, resource, utilities

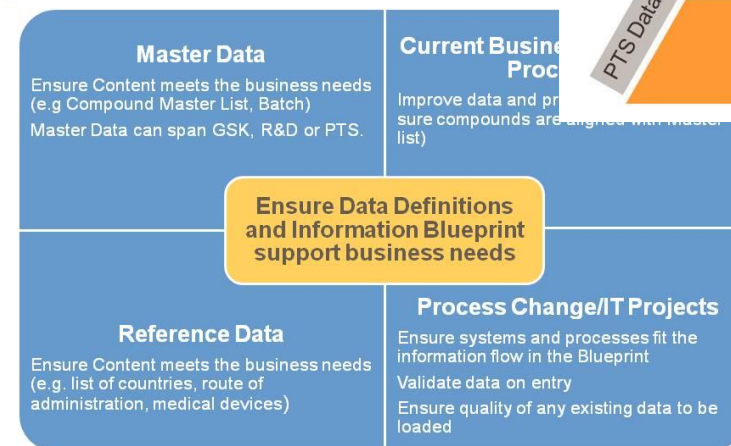


Proposed PTS Stewardship Framework



Data Governance Framework.

- Data foundation for the stewards to build on
- Some business resource needed to manage the framework: elen
- May require critical decisions from PTS Data Stewardship Spons





DQ Dimensions and Measures



- This shows dimensions and measures for assessing Data Quality of Master Data in R&D
- For transactional data the dimensions and measures may be different

One individual may have different roles (Family Dr. vs. Primary Investigator), different addresses, similar looking names recorded throughout GSK

Sales Force Rx data

Dr. John Jacobs
Manchester Family Practice
123 Main Street
Manchester, NH 13109

eTrack Investigator data

J Jacobs M.D.
NH University Hospital Clinic
Rx Trials Unit, Mail Stop 5456
Astoria, NH 14108

Informatica MDM

- Matches similar name, address(es)
- Merges names into 1, and in a preferred format - keeps two addresses



John Jacobs M.D.

Human intervention, governance sometimes required to ensure data quality + accountability

Why is this important?

❖ **HCP**: requires us to know we paid 1 MD (not 2) for all rolled-up payments.

❖ **Drug safety reports, trial supplies** are sent to the correct address

❖ GSK avoids making **duplicative payments**

❖ Clinical Study Investigators data are up to date, accurate for IND/NDA purposes.

Reference Data Management

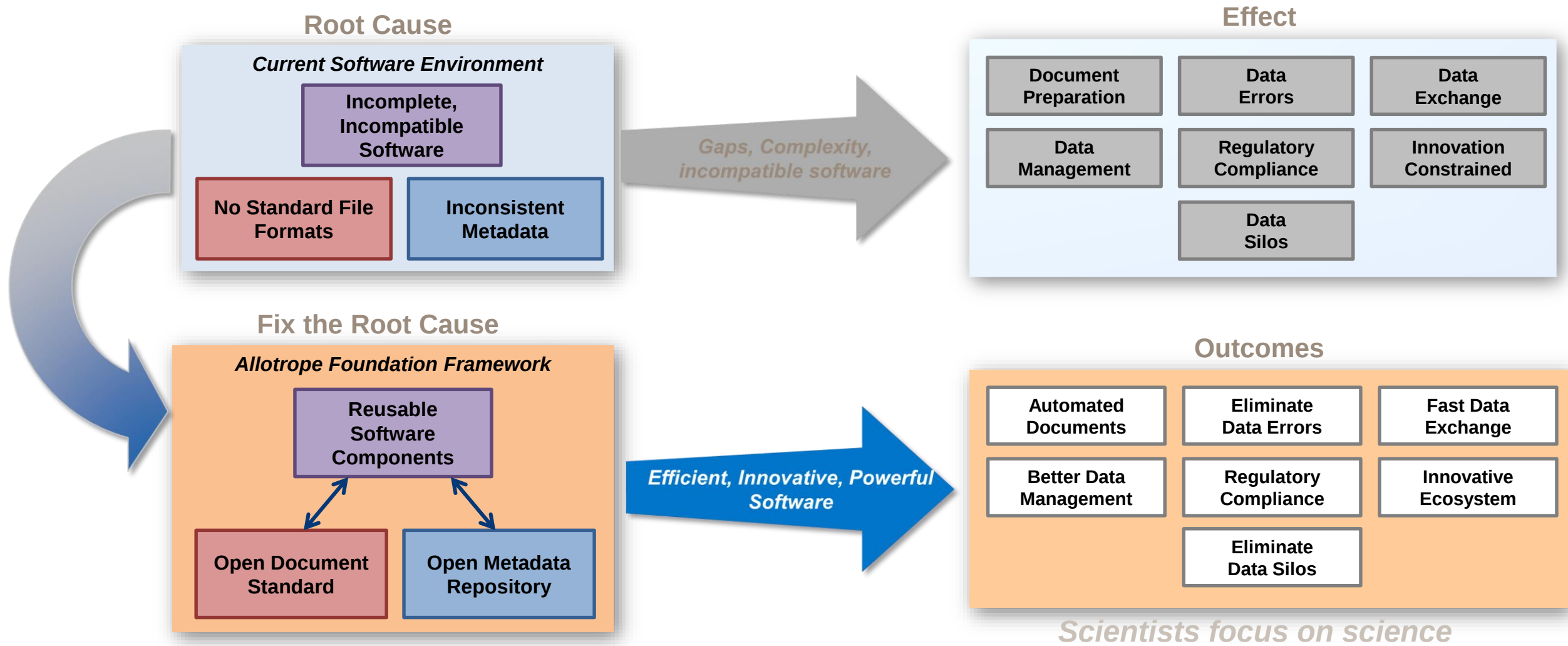
Vocabularies, Taxonomies and Ontologies



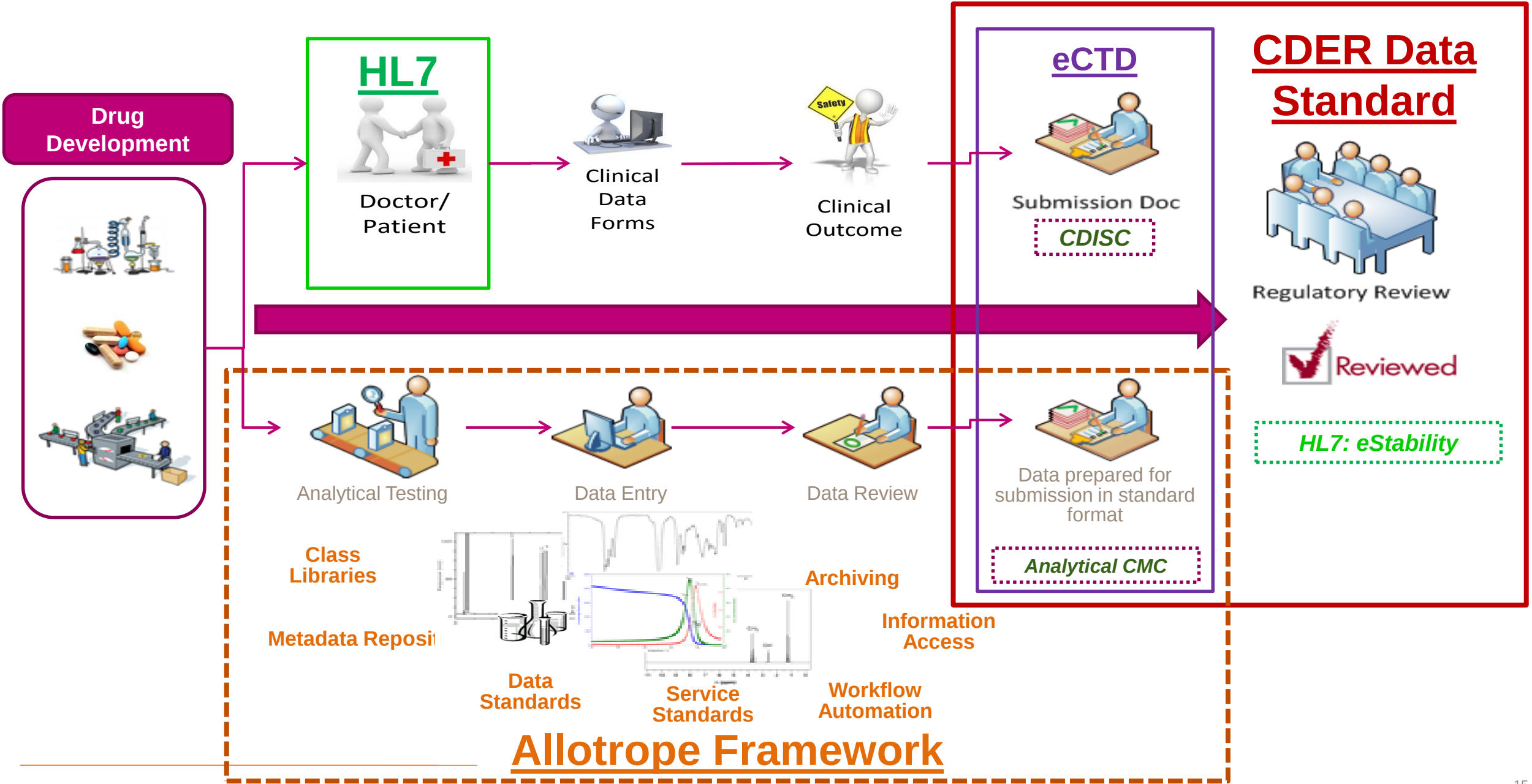
- An ontology is a set of related vocabulary/taxonomies specific to an area enabling Knowledge Management allowing us to create a hierarchy of related terms
 - When implementing ontologies need to understand the model, governance/best practice and quality of entries
 - Allow us to link disparate data sets together through related terminologies
 - Easily extensible and can be embedded within standards and frameworks
 - Allows for organisational mappings
 - Relevant Pharma Ontology Usages
 - [Reactome](#), available in an RDF format allowing organisations to link pathway information utilised in Metabolomics and relate disease, target and chemical information
 - Allotrope, affiliation of ontologies associated with the analytical chemistry space, for example equipment
-

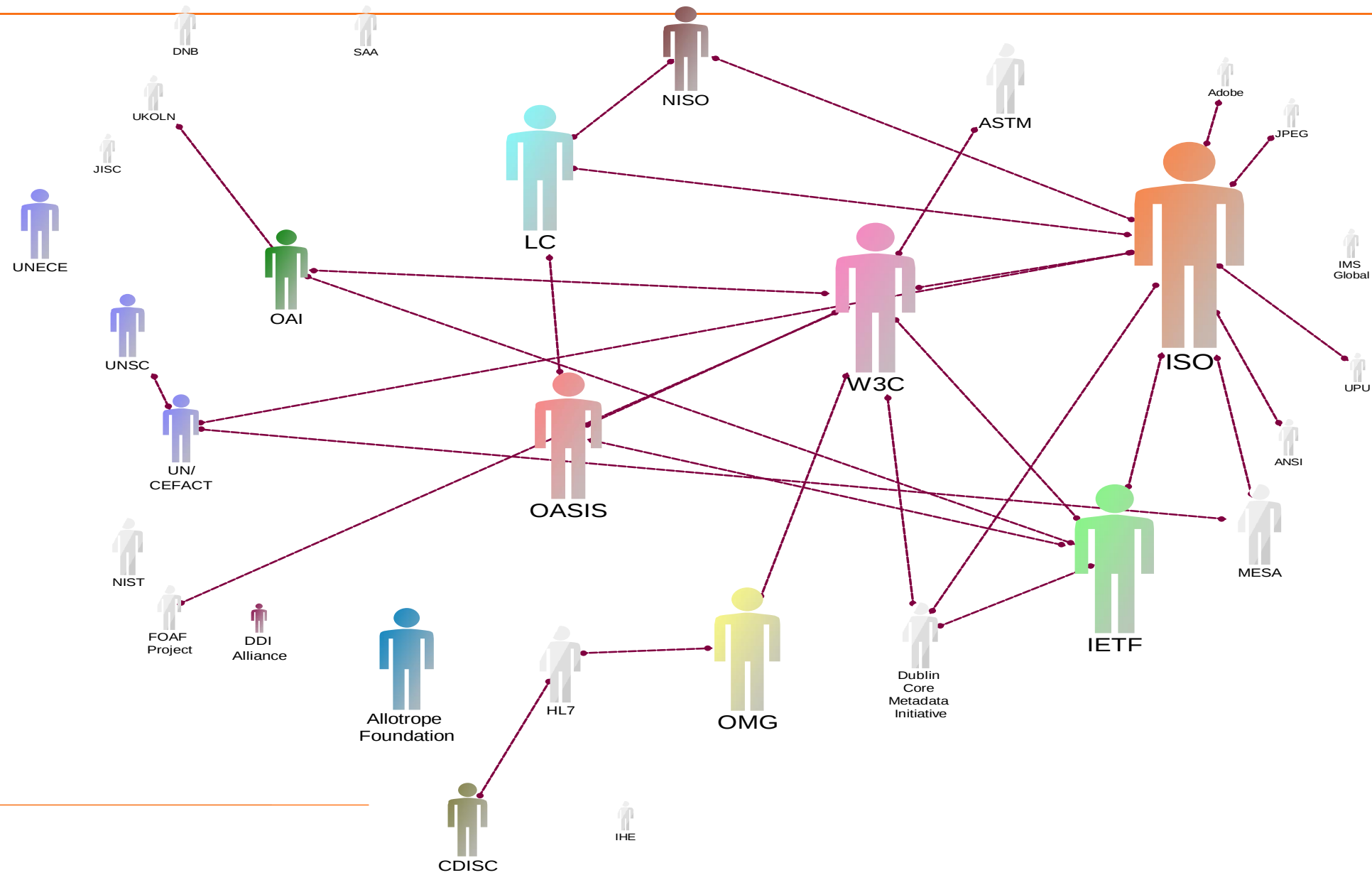
-
- For organisations to be able to exchange data not only do we need to agree on the structure but what is meant by the structure and the business rules associated with the data
 - To support this we can use the concepts we have already discussed Information Blueprints, Stewardship & Governance and Data Quality
 - The additional aspect and probably most important is that of our Master and Reference data where our Metadata will be held, the drop down lists in our everyday applications
 - To improve the ease associated with data exchange we need to agree on terminology and if we can't them support mappings/synonyms to enable us to integrate datasets at a later point that are then embedded within the use of the standard
 - In the growing world of information exchange and data types there is an ever increasing number of standards and frameworks to implement these
 - ISA-88 and ISA-95, associated .xsd formats BatchML and B2MML
 - Pistoia, HELM, Standard Metadata, Ontology, Standard Data Warehouse projects
 - Allotrope, Analytical Chemistry Framework
-

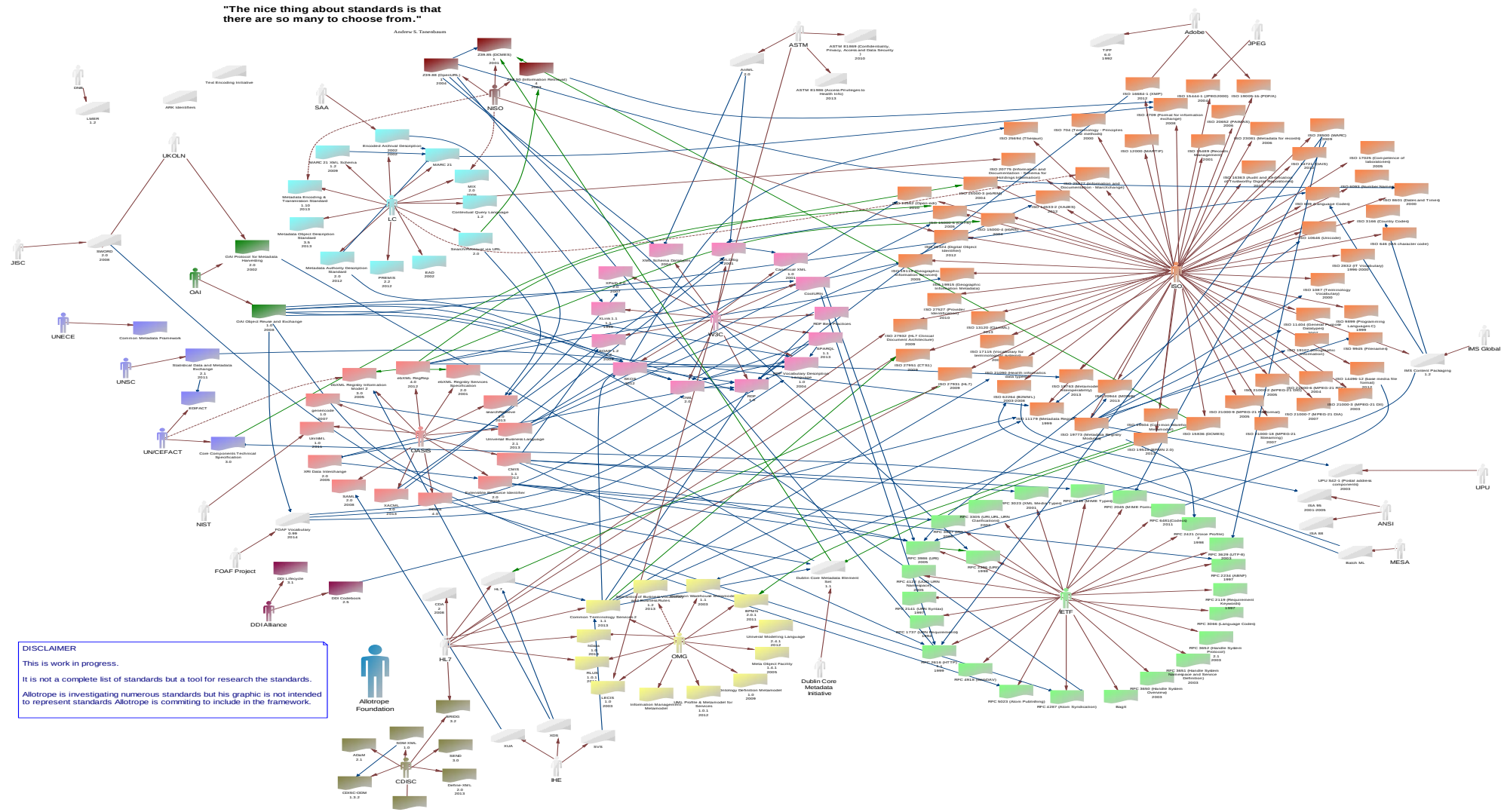
Allotrope : Advanced Data Design for Chemical R&D



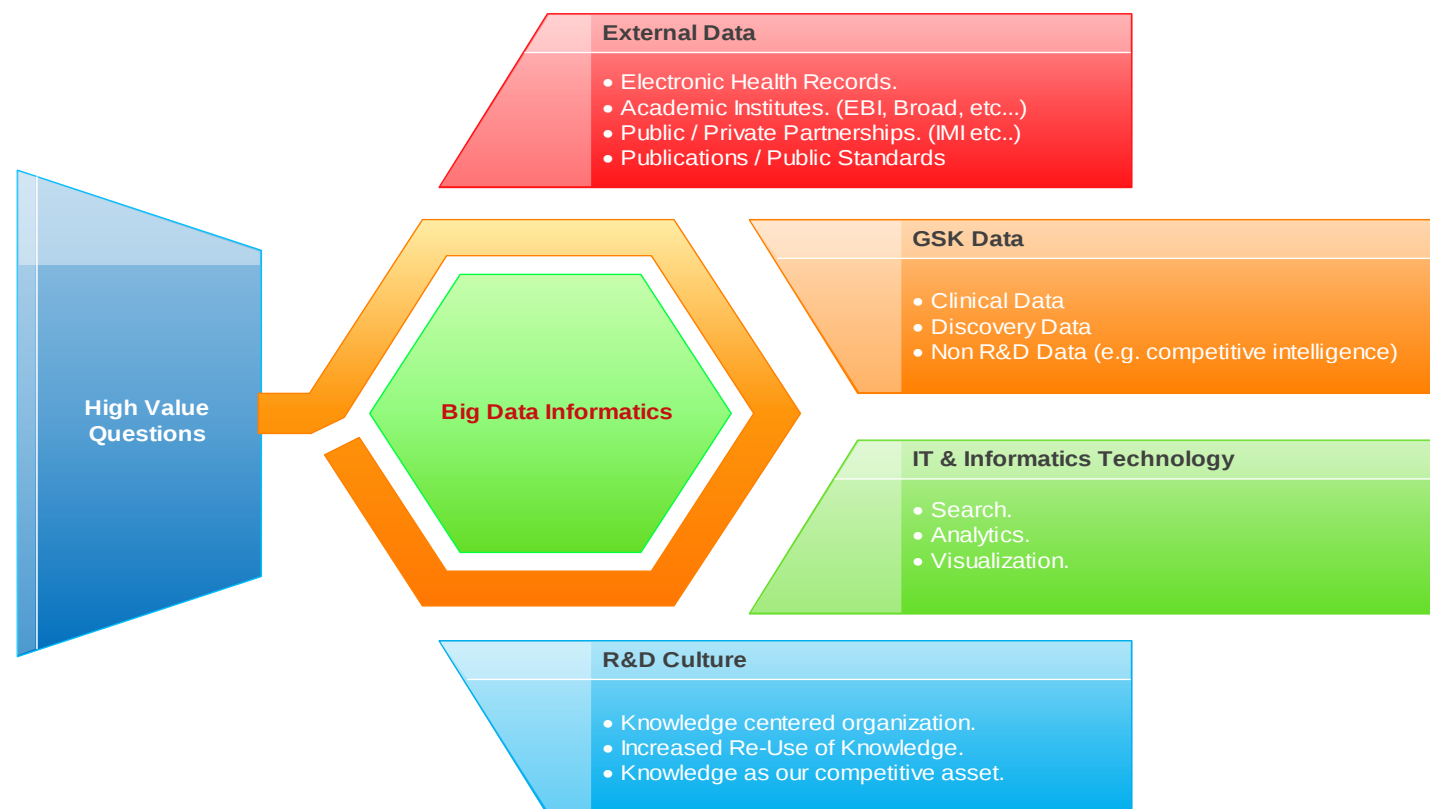
Allotrope Framework addresses the gap in standards for CMC analytical data







- Implement a strategy addressing the need for R&D to Big Data through Informatics to enhance decision making, to *Simplify of the Operating Model* and ensure efficiencies to *Deliver More Products of Value*
- **Goal: Maximize insight from the minimum necessary data**
 - Standards, MDM, Data Quality, Data Governance & Stewardship and Blue Printing are some of the foundational capabilities to do this



SOCRATES SEARCH



The image displays three overlapping screenshots of the SOCRATES search interface. The top-left screenshot shows a search for "eotaxin response" with filters for Internal Sources, Compounds, and Safety. The top-right screenshot shows a search for a chemical structure with the SMILES string CN1C=CC2=CC(NC)=CC=C21 and a chemical structure of a benzimidazole derivative. The bottom screenshot shows a search for "BRL-49653HG" with filters for Internal Sources, Compounds, and Safety, and a chemical structure of a benzimidazole derivative. The interface includes a search bar, filters, and a list of results with chemical structures and links to external data sources.



Scientific searching for R&D



Text and chemical searches



Internal and External data sources linked through ontologies

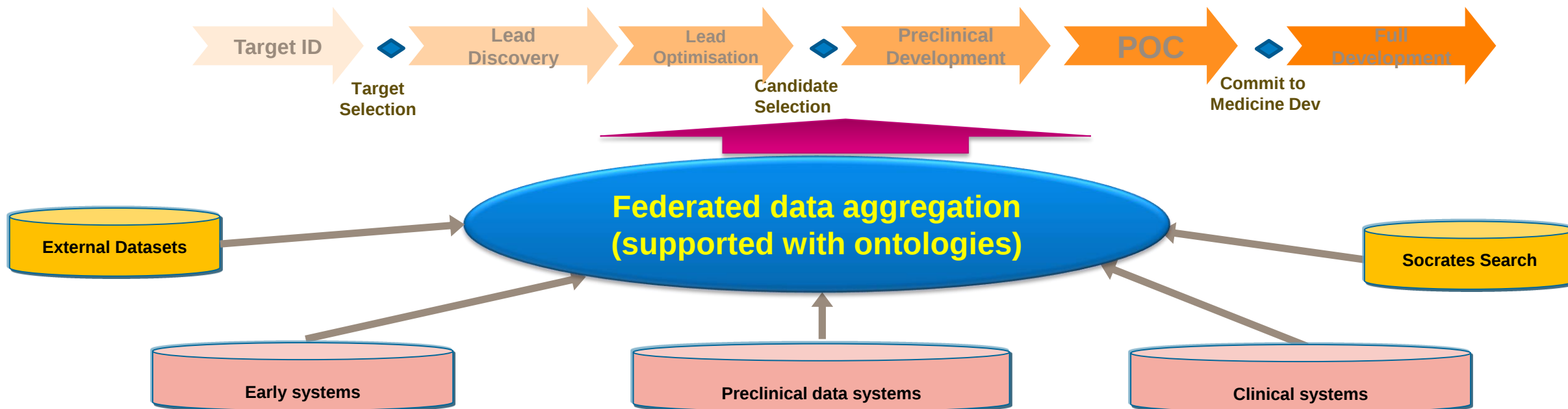


More access to R&D content

Phase II – Summarise GSK Experience of Compounds, Targets



- Provide overview of compounds using traditional drug Discovery-Development chevrons, allow drilldown into the raw data.
- Integrated into Socrates Search, whenever a compound or medicine is 'detected' in search results
- Targetpedia : existing system integrated and extended into Socrates



Targetpedia

Compoundpedia

Exporting to Excel

Whats next ? ...

