Mashing Up Drug Discovery

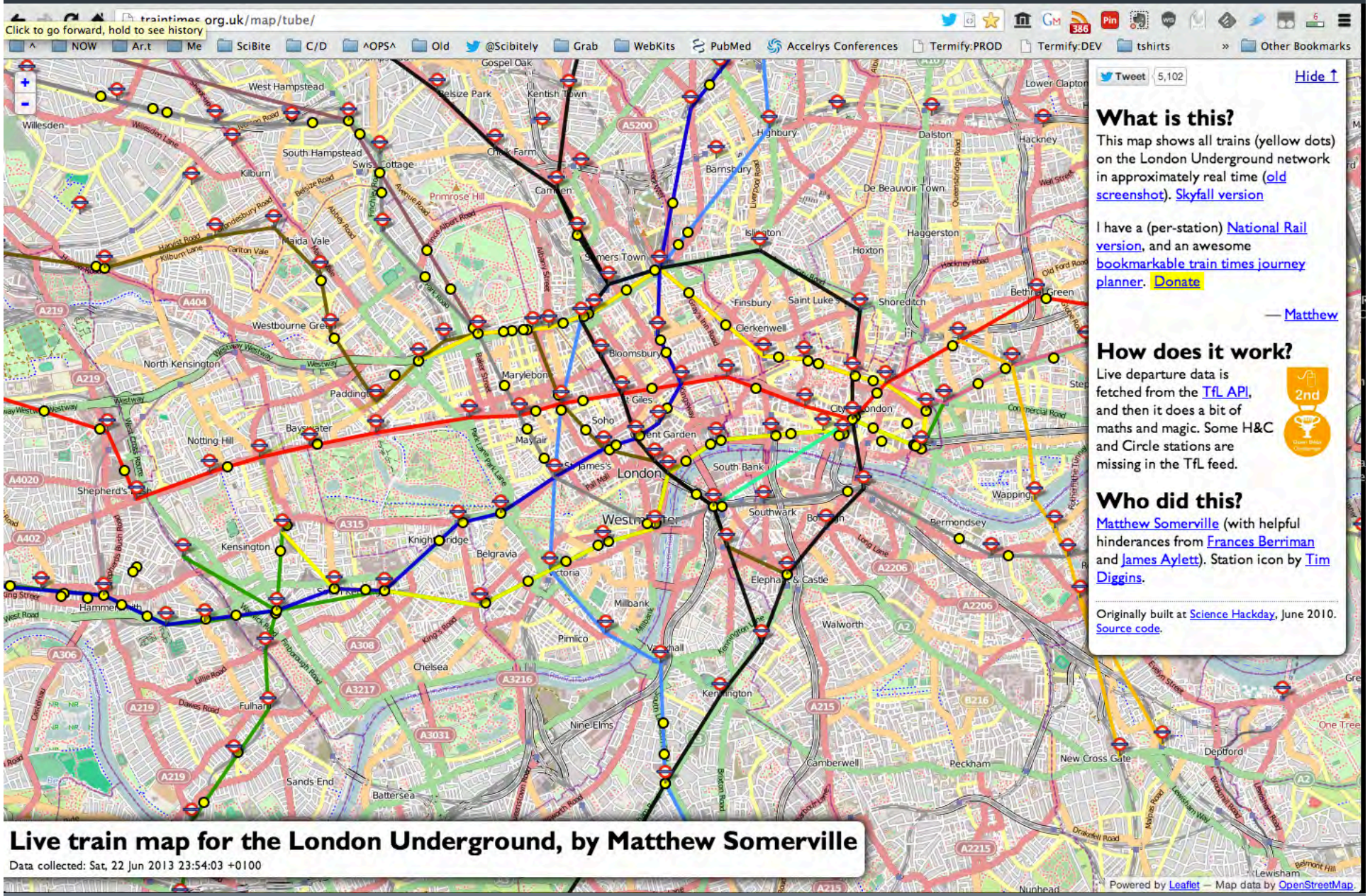
There's an API for that….

Lee Harland July 2013
Open PHACTS http://openphacts.org
SciBite http://scibite.com
ConnectedDiscovery http://connecteddiscovery.com
@Scibitely

What does an Application Program Interface look like?

```
<application xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://wadl.dev.java.net/2009/02" xmlns:ops="http://ops.epo.org"
xmlns:xs="http://www.w3.org/2001/XMLSchema" xsi:schemaLocation="http://wadl.dev.java.net/2009/02  ..\schema\wadl.xsd">
        <grammars>
                <include href="../schema/ops.xsd"/>
        </grammars>
        <resources base="http://ops.epo.org/3.0/rest-services">
    <!-- GET Published data search -->
    <resource path="published-data/search/{publishedConstituents}/" id="publishedDataSearch">
      <doc xml:lang="en" title="Published document search"/>
      <param name="publishedConstituents" type="xs:string" required="true" style="template" default="biblio">
        <option value="biblio"/>
        <option value="abstract"/>
        <option value="full-cycle"/>
        <option value="abstract,full-cycle"/>
        <option value="biblio,full-cycle"/>
      </param>
      <method name="GET" href="#wpdSearchGET"/>
      <!-- search query -->
    </resource>
                <!-- GET Published data biblio -->
                <resource path="published-data/{ref-type}/{ref-format}/{number}/{publishedConstituents}" id="publishedDataRetrieval">
                        <doc xml:lang="en" title="Published document retrieval GET interface"/>
                        <param href="#ref-type"/>
                        <param href="#ref-format"/>
                        <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
      <doc xml:lang="en">Reference number</doc>
      </param>
      <param name="publishedConstituents" type="xs:string" required="true" style="template" default="biblio">
        <option value="biblio"/>
        <option value="abstract"/>
        <option value="full-cycle"/>
        <option value="abstract,full-cycle"/>
        <option value="biblio,full-cycle"/>
      </param>
                        <method name="GET" href="#wpdGETPOST"/>
                </resource>
    <!--Published data fulltext availability inquiry GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/fulltext" id="publishedDataFulltextRequest">
      <doc xml:lang="en" title="Published document fulltext availability inquiry"/>
      <param href="#ref-type"/>
      <param href="#ref-format"/>
      <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
        <doc xml:lang="en">Reference number</doc>
      </param>
      <method name="GET" href="#wpdGETPOST"/>
    </resource>
    <!--Published data claims retrieval GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/claims" id="publishedDataClaimsRequest">
      <doc xml:lang="en" title="Published document claims"/>
      <param href="#ref-type"/>
      <param href="#ref-format"/>
      <param name="number" type="xs:string" required="true" style="template" default="EP1000000">
        <doc xml:lang="en">Reference number</doc>
      </param>
      <method name="GET" href="#wpdGETPOST"/>
    </resource>
    <!--Published data description retrieval GET -->
    <resource path="published-data/{ref-type}/{ref-format}/{number}/description" id="publishedDataDescriptionRequest">
```

http://traintimes.org.uk/map/tube/



REPORTS CONTENT

- Avian flu home
- News
- Research
- Google earth
- Timeline
- FAQs

REPORTS LINKS

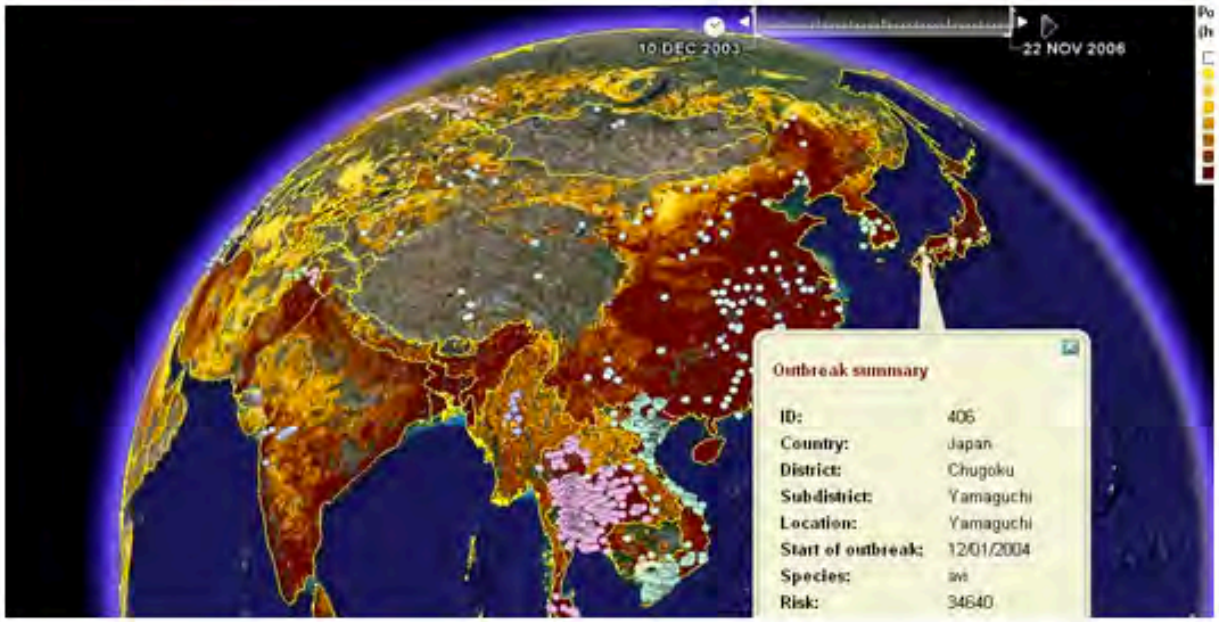
- Stem cells
- Climate change

NPG resources

- Nature
- Nature Cell Biology
- Nature Medicine
- news@nature.com
- Nature Reviews Drug Discovery
- Naturejobs

google earth

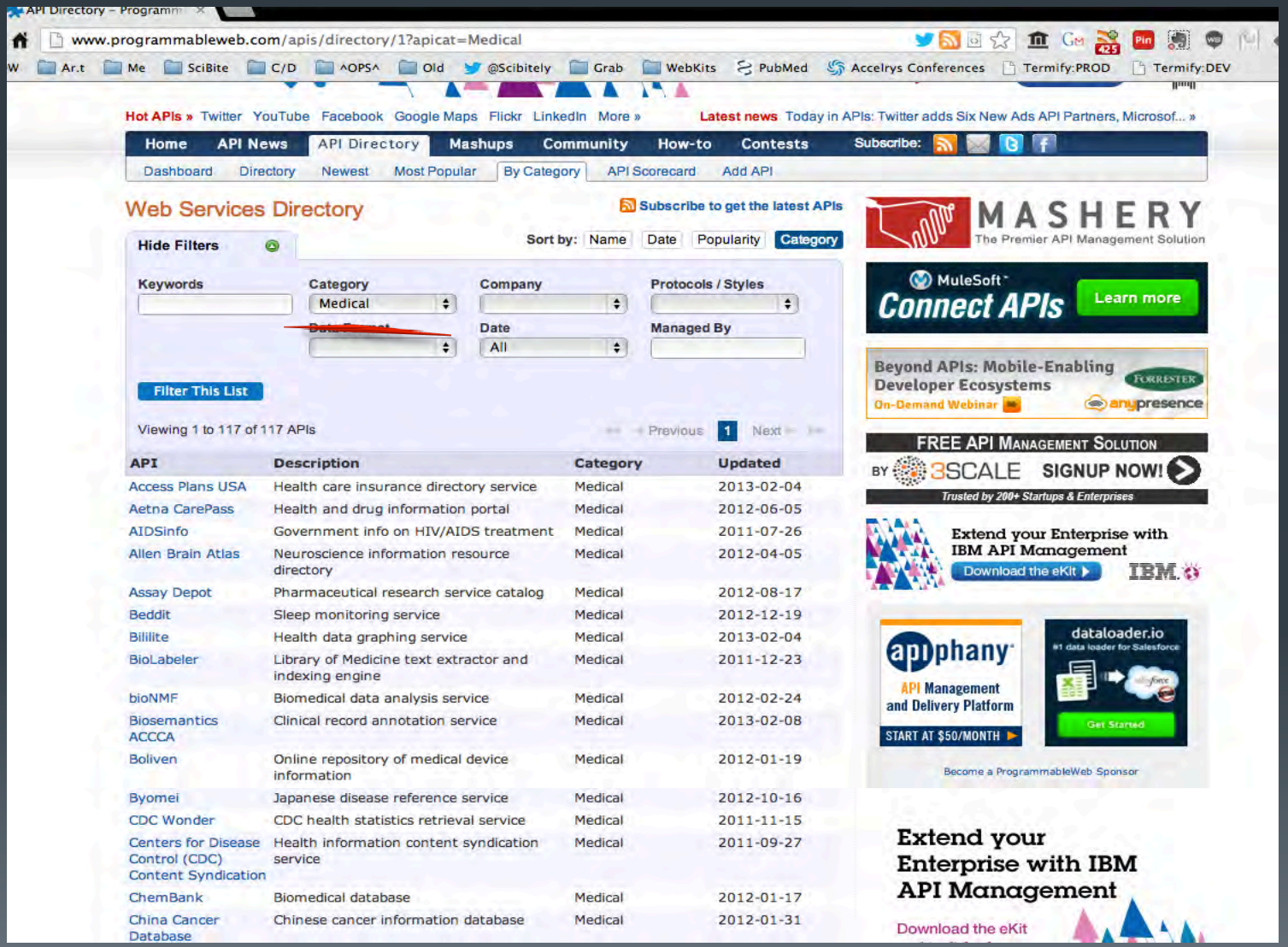
Track the spread of the H5N1 virus worldwide since 2003 with our award-winning animated Google Earth Mash-Up.



View the time-enabled maps


rolsen
3 Metre
10feet
rolsen
rolsen
rolsen
rolsen
rolsen

Are There Any Bio/Chem/Med?



http://programmableweb.com/

Good Examples

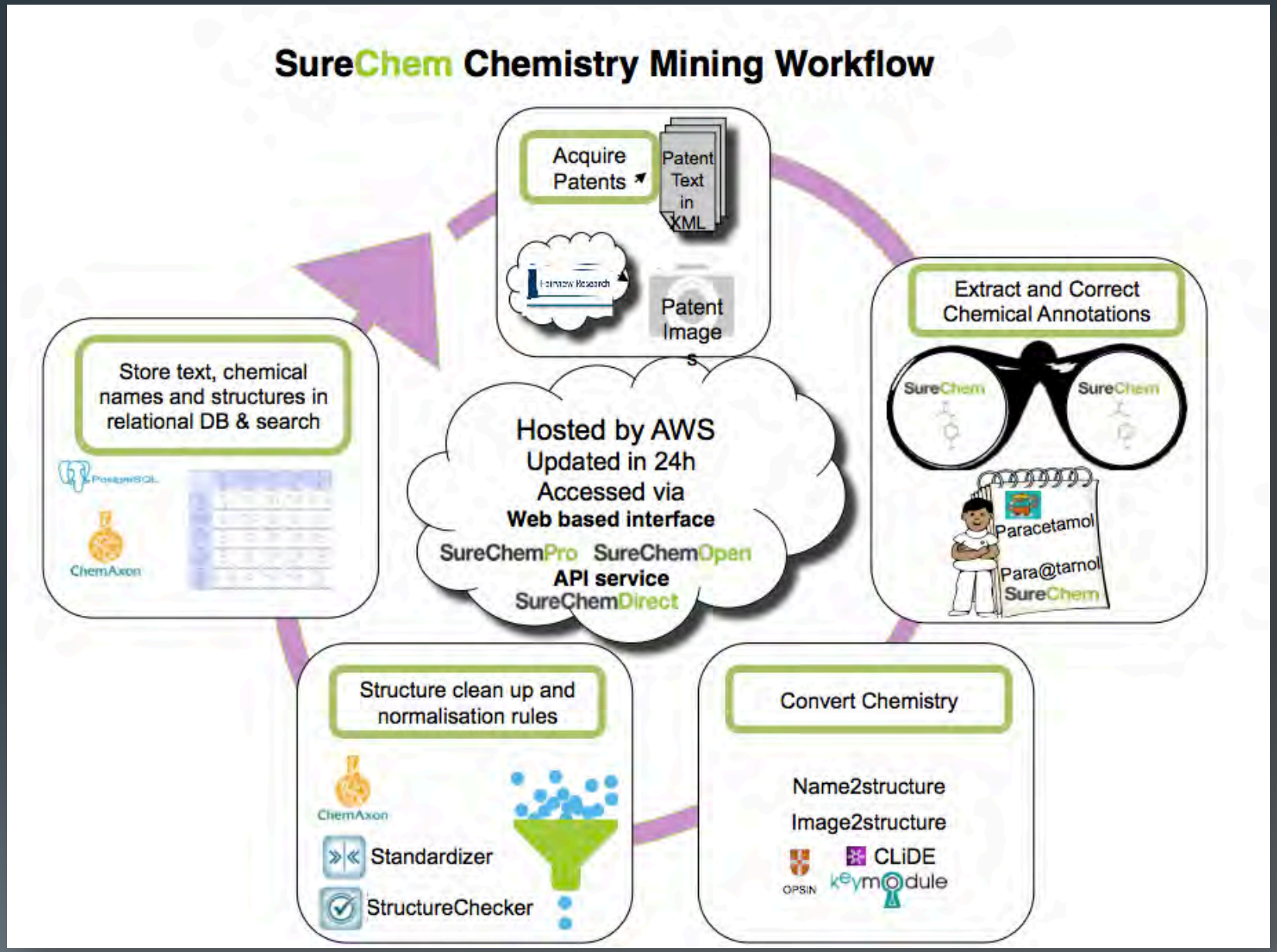
- NLM/NCBI/PubChem - http://www.ncbi.nlm.nih.gov/
- EBI (inc ChEMBL,ChEBI etc) – http://ebi.ac.uk
- NCI Cactus - http://cactus.nci.nih.gov/
- ChemSpider - http://www.chemspider.com/
- ChemAxon – e.g. Chemicalize http://www.chemicalize.org/
- OpenTox - http://www.opentox.org/dev/apis/api-1.2/structure
- Altmetric & ImpactStory
- .. Many more web services out there…


Free
vs
Accessible

Case Studies

1. SureChem

Patent Chemistry


SureChem Chemistry Mining Workflow
Acquire Patents
Patent Text in XML
Patent Image s
Extract and Correct Chemical Annotations
SureChem
SureChem
Paracetamol
Para@tamol
SureChem
Store text, chemical names and structures in relational DB & search
ChemAxon
Hosted by AWS
Updated in 24h
Accessed via
Web based interface
SureChemPro SureChemOpen
API service
SureChemDirect
Structure clean up and normalisation rules
ChemAxon
Standardizer
StructureChecker
Convert Chemistry
Name2structure
Image2structure
OPSIN
CLiDE
keymodule

SureChemOpen

- **Free** patent chemistry searching for **all**
- **Early** access to newly published structures
- **Easy navigation** through the relevant chemistry in the full text of patent documents
- Search, View, Print
- Links to ChemSpider and RSC journals
- A free viewer for SureChemDirect users

Secured by 128-bit encryption providing same level of security as online banking and other proprietary chemistry searching tools!

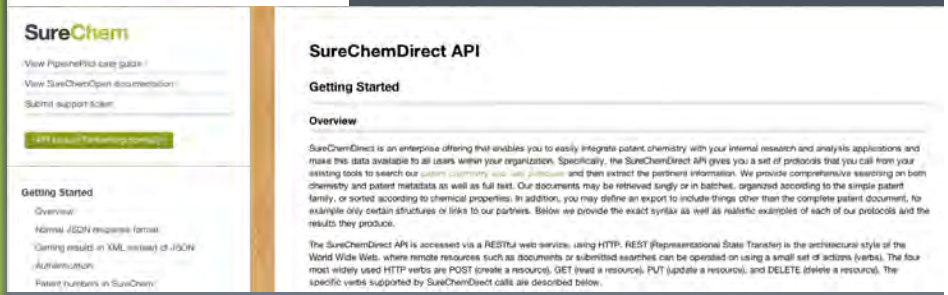
SureChemDirect

API

- Direct access **to chemistry & full text** patent data
- Data and search hosted on SureChem AWS instance
  - Dedicated AWS instances available
  - Secure https line, 128-bit encryption
- **Next day indexing** of chemistry
- **Integration** with internal databases and workflow tools (SharePoint, Seurat, SpotFire etc)
- **Rapid deployment** using **Pipeline Pilot package**
- Easy manipulation of large volumes of patent chemistry
- Full support and documentation
  - http://docs.surechem.com/direct/

Data Feed

- Direct access to **chemistry & patent metadata in csv** format
- Access via secure ftp
- Batch screening & analysis behind your firewall
- Daily updates



2. SciBite

Drug Discovery, Right Now.

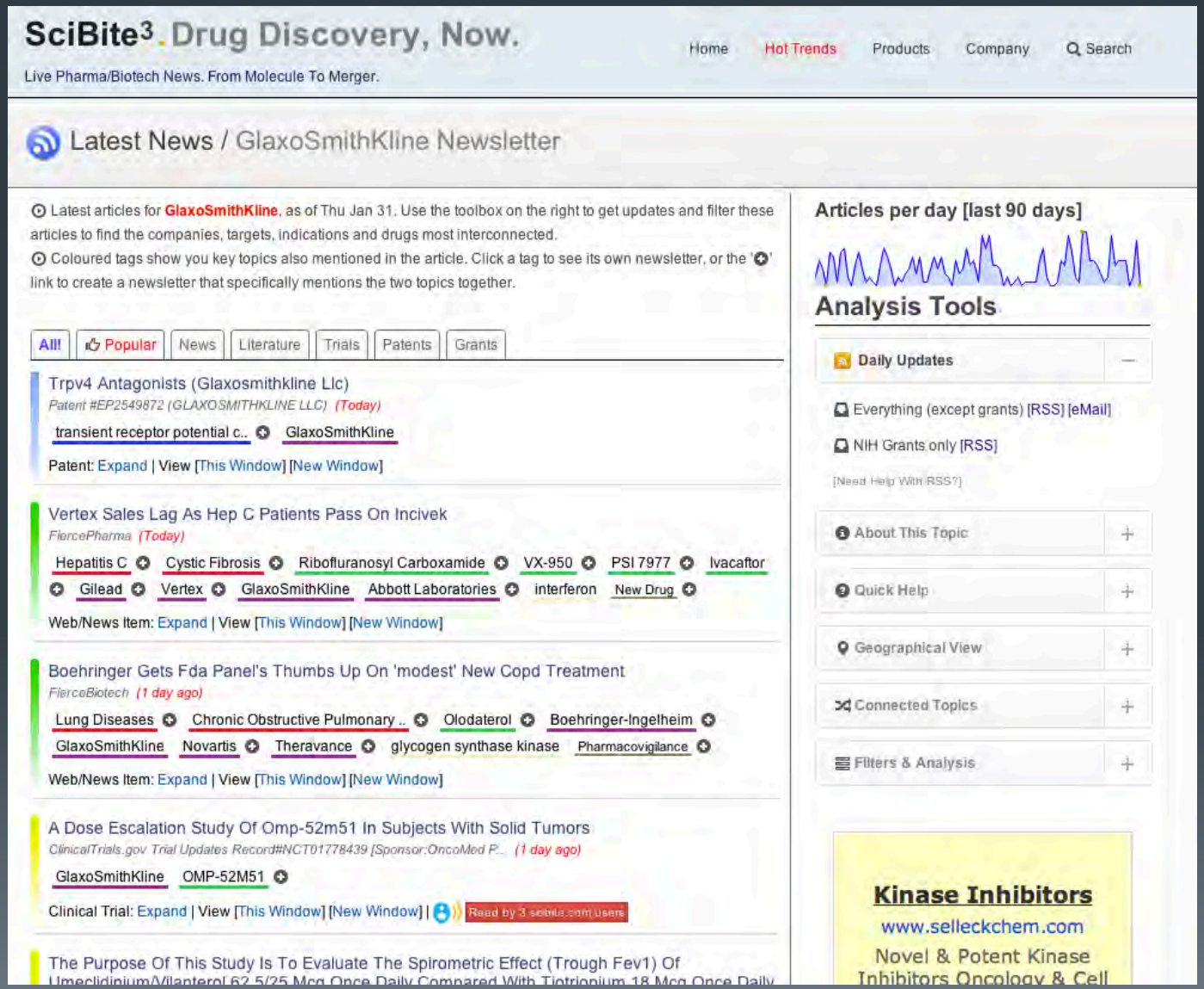
Therapeutic Intelligence

I care about asthma.

What does the current landscape of (Drugs, Targets, Companies, Trials, Collaborations…) look like?

Where does TI come from?

Twitter-like Newsfeed

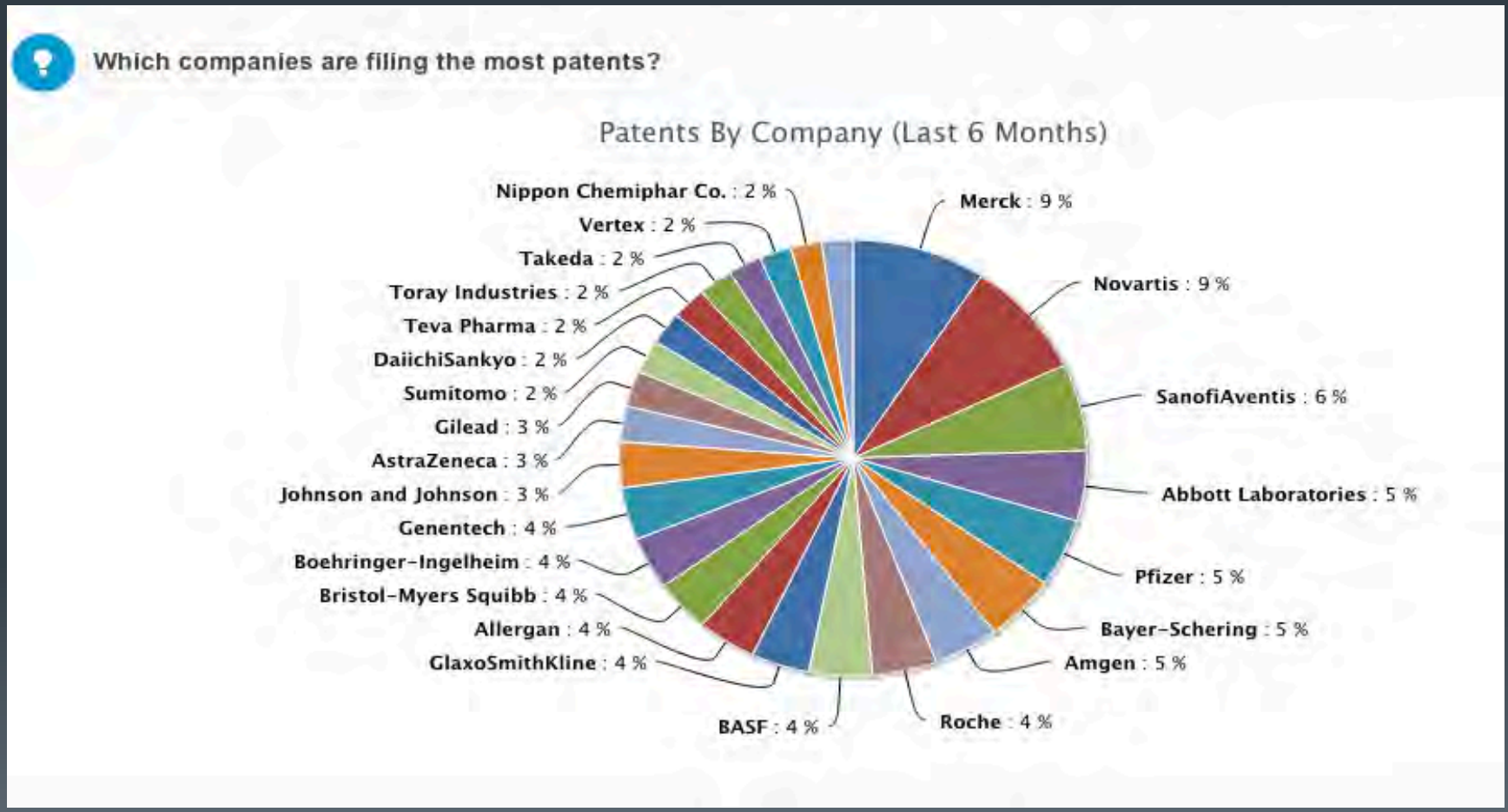


Operations

SciBite Main API Methods (v3.1)

- **EntityBooleanQueryService: Get News For One Or More Entities** — `/EntityBooleanQueryService` GET
- **EntityNewsService: Get recent news articles for an input entity** — `/EntityNewsService` GET
- **EntitySearchService: Find ID for an entity** — `/EntitySearchService` GET
- **EntityWebLinksService: Get HTML Links for interesting sites** — `/EntityWebLinksService` GET
- **EntityTextToNewsService: Entity news by text query** — `/EntityTextToNewsService` GET
- **DocumentToNewsService: Get news for entities in a document** — `/DocumentToNewsService` GET
- **DocumentEntitiesService: Get entities in a document** — `/DocumentEntitiesService` GET
- **ListEntitiesService: List all entities** — `/ListEntitiesService` GET
- **TermiteService: Run termite NER on a piece of text** — `/TermiteService` GET
- **InsightLinksService: Get entities most correlated with an input entity** — `/InsightLinksService` GET
- **InsightTopTypesService: Top entities in the news** — `/InsightTopTypesService` GET
- **InsightJointInSourceService: Co-occuring entities in a source** — `/InsightJointInSourceService` GET
- **InsightLitSpaceService: Co-occuring entities in historical medline abstracts** — `/InsightLitSpaceService` GET
- **InsightLitSpacePairService: Co-occuring entity-entity data in historical medline abstracts** — `/InsightLitSpacePairService` GET

Slice & Dice With The API



Hot Target Barcodes

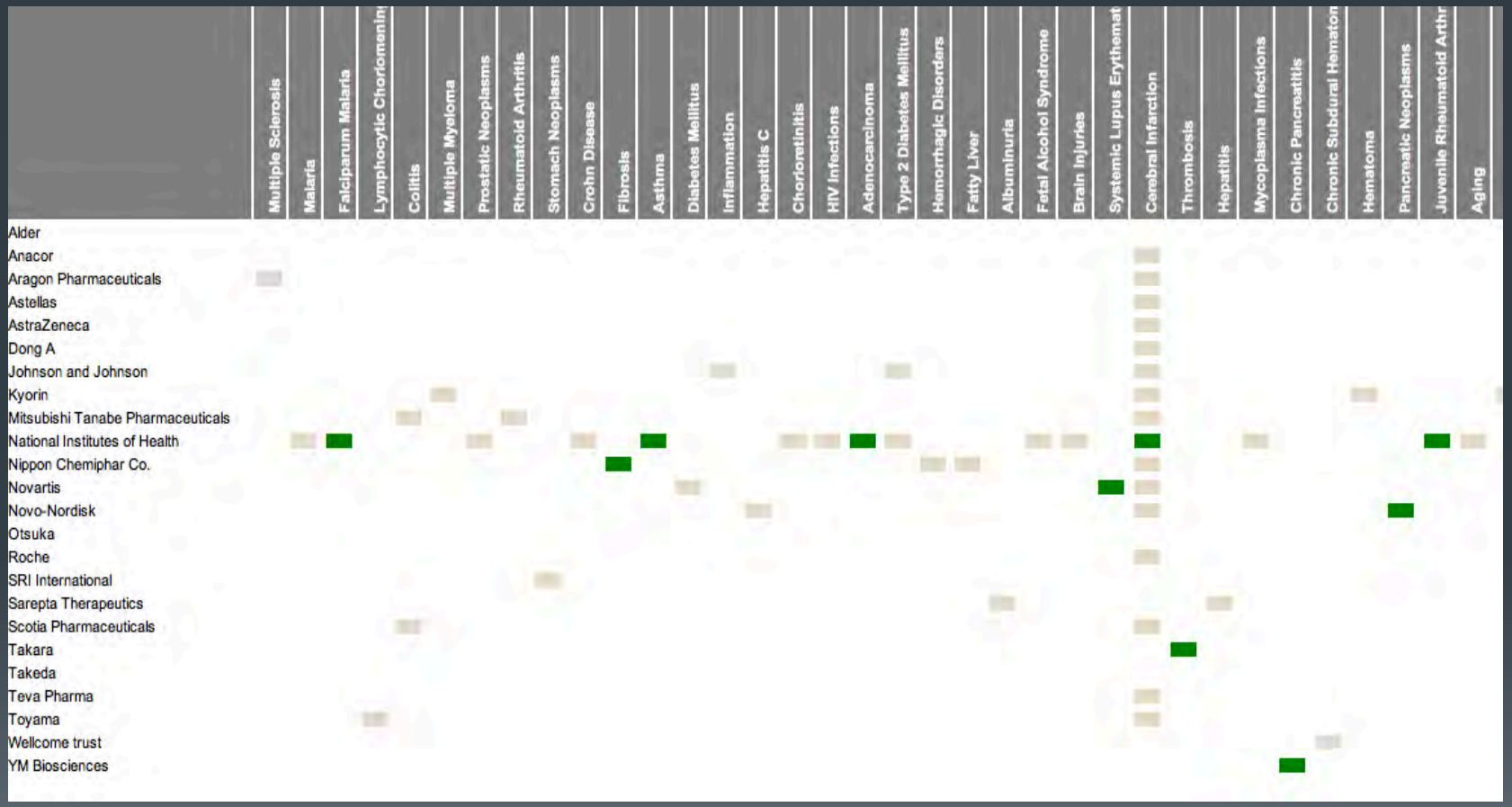
ID	Name	2002-2013 Literature Significance Plot
CD4	CD4 molecule	
CD8A	CD8a molecule	
HLA-DRB1	major histocompatibility complex, class II, DR beta 1	
IFNB1	interferon, beta 1, fibroblast	
IFNG	interferon, gamma	
MMP9	matrix metallopeptidase 9 (gelatinase B, 92kDa gelatinase, 92kDa type IV collagenase)	
MOG	myelin oligodendrocyte glycoprotein	
IL10	interleukin 10	
IL4	interleukin 4	
APOE	apolipoprotein E	
IL6	interleukin 6 (interferon, beta 2)	
BDNF	brain-derived neurotrophic factor	
CCL2	chemokine (C-C motif) ligand 2	
CTLA4	cytotoxic T-lymphocyte-associated protein 4	
IL12A	interleukin 12A (natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35)	
IL1B	interleukin 1, beta	
CRYAB	crystallin, alpha B	
IL2	interleukin 2	
IL2RA	interleukin 2 receptor, alpha	
CCR5	chemokine (C-C motif) receptor 5	
ISG20	interferon stimulated exonuclease gene 20kDa	

CCR2	chemokine (C-C motif) receptor 2	
HLA-DRB5	major histocompatibility complex, class II, DR beta 5	
PON1	paraoxonase 1	
FASLG	Fas ligand (TNF superfamily, member 6)	
HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase	
ICAM1	intercellular adhesion molecule 1	
IL7	interleukin 7	

Indication Opportunity Map

Indications →

Organisations →



3. Open PHACTS

Navigating Pharmacological Space

Open PHACTS Project Partners

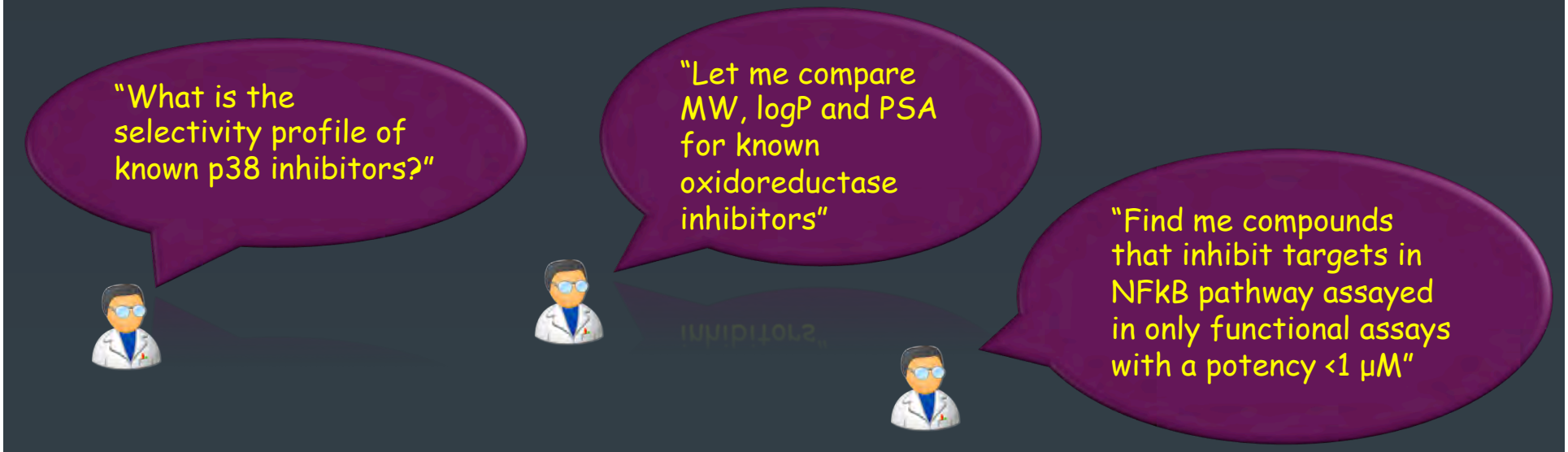
Pfizer Limited – Coordinator
Universität Wien – Managing entity
Technical University of Denmark
University of Hamburg, Center for Bioinformatics
BioSolveIT GmbH
Consorci Mar Parc de Salut de Barcelona
Leiden University Medical Centre
Royal Society of Chemistry
Vrije Universiteit Amsterdam

Spanish National Cancer Research Centre
University of Manchester
Maastricht University
Aqnowledge
University of Santiago de Compostela
Rheinische Friedrich-Wilhelms-Universität Bonn
AstraZeneca
GlaxoSmithKline
Esteve

Novartis
Merck Serono
H. Lundbeck A/S
Eli Lilly
Netherlands Bioinformatics Centre
Swiss Institute of Bioinformatics
ConnectedDiscovery
EMBL-European Bioinformatics Institute
Janssen
OpenLink

pmu@openphacts.org

It *Should* be easy to ask…



ChEMBL | DrugBank | Gene Ontology | Wikipathways | GeneGo

ChEBI | UniProt | UMLS | GVKBio

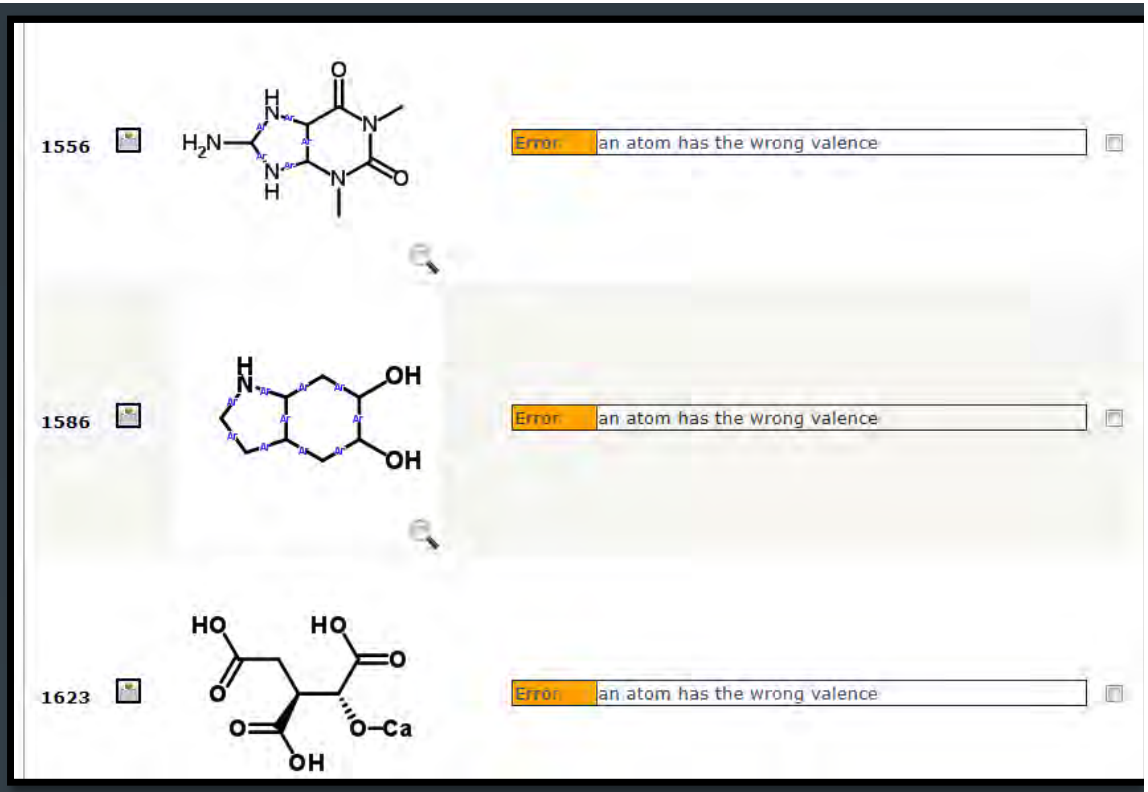
ConceptWiki | ChemSpider | TrialTrove | TR Integrity

Business Question Driven Approach

Number	sum	Nr of 1	Question
15	12	9	All oxidoreductase inhibitors active <100nM in both human and mouse
18	14	8	Given compound X, what is its predicted secondary pharmacology? What are the on and off,target safety concerns for a compound? What is the evidence and how reliable is that evidence (journal impact factor, KOL) for findings associated with a compound?
24	13	8	Given a target find me all actives against that target. Find/predict polypharmacology of actives. Determine ADMET profile of actives.
32	13	8	For a given interaction profile, give me compounds similar to it.
37	13	8	The current Factor Xa lead series is characterised by substructure X. Retrieve all bioactivity data in serine protease assays for molecules that contain substructure X.
38	13	8	Retrieve all experimental and clinical data for a given list of compounds defined by their chemical structure (with options to match stereochemistry or not).
41	13	8	A project is considering Protein Kinase C Alpha (PRKCA) as a target. What are all the compounds known to modulate the target directly? What are the compounds that may modulate the target directly? i.e. return all cmpds active in assays where the resolution is at least at the level of the target family (i.e. PKC) both from structured assay databases and the literature.
44	13	8	Give me all active compounds on a given target with the relevant assay data
46	13	8	Give me the compound(s) which hit most specifically the multiple targets in a given pathway (disease)
59	14	8	Identify all known protein-protein interaction inhibitors

… paper coming very soon in DDT

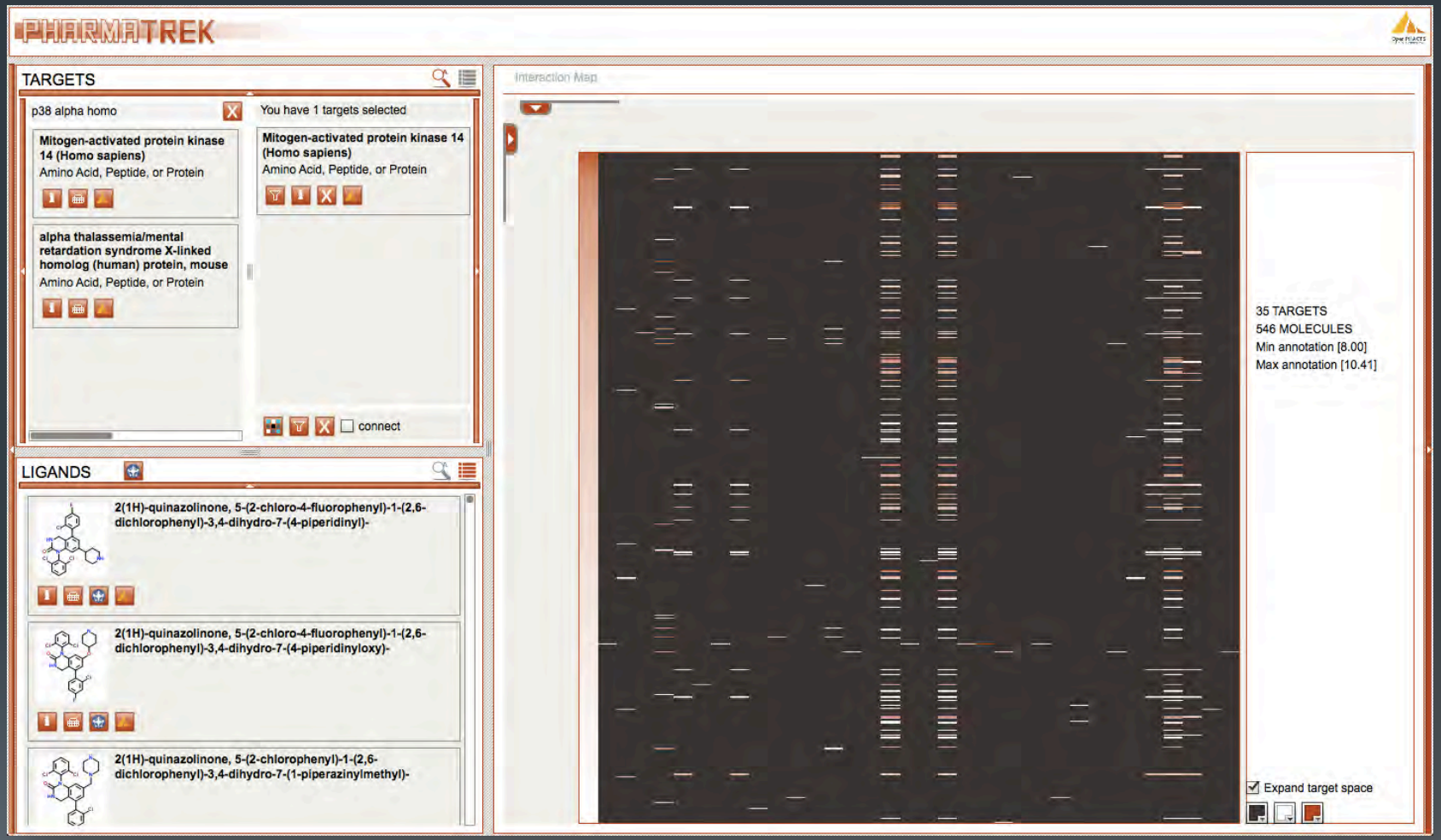
ChemSpider
The free chemical database



- Chemical Registration System
  - Utilizes ChemSpider Validation and Standardization platform including collapsing tautomers
  - Utilizes FDA rule set as basis for standardization (GSK lead)
  - All molecules get a CVSP report – helps us biologists!

dev.openphacts.org

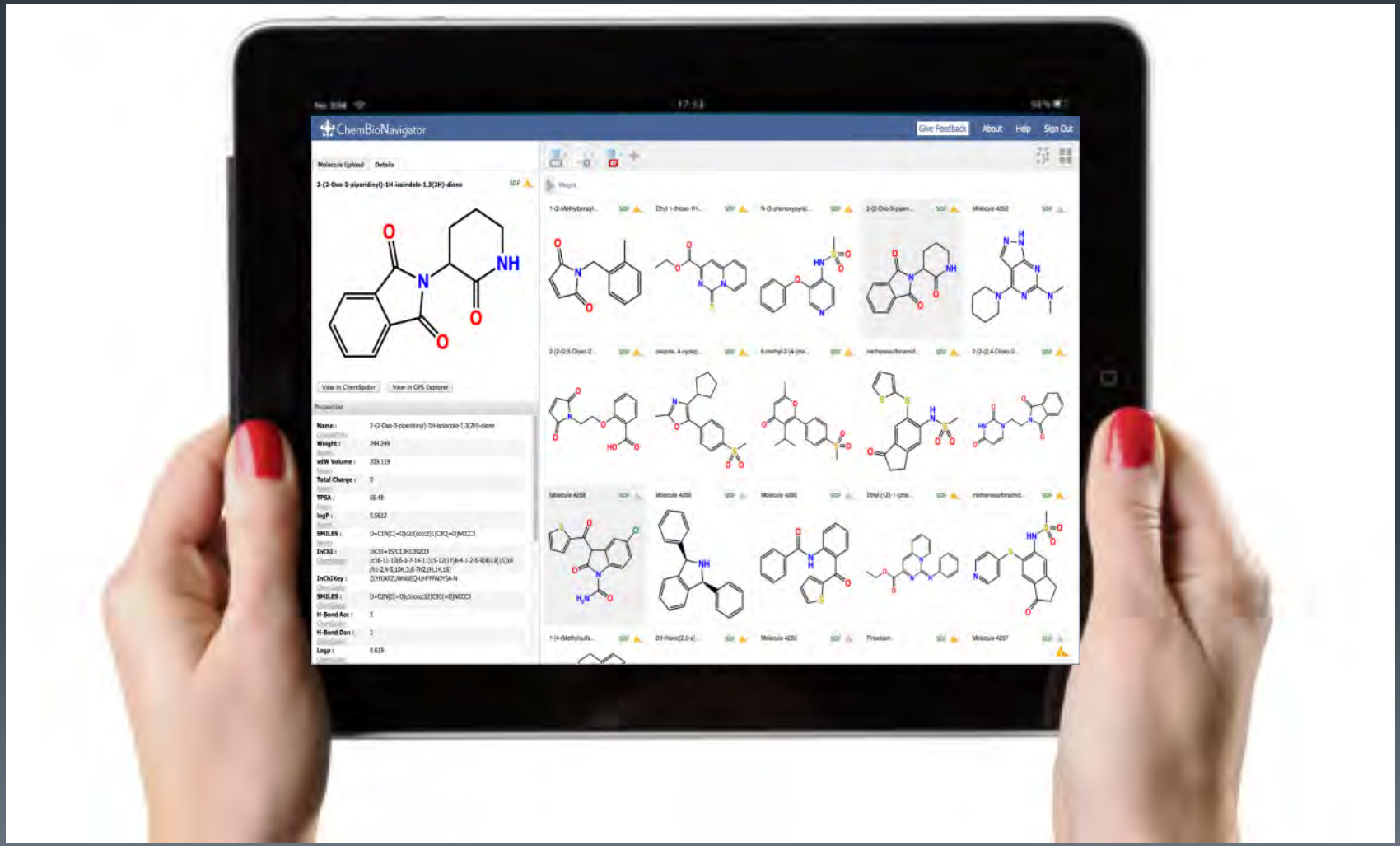
PharmaTrek



http://pharmatrek.org

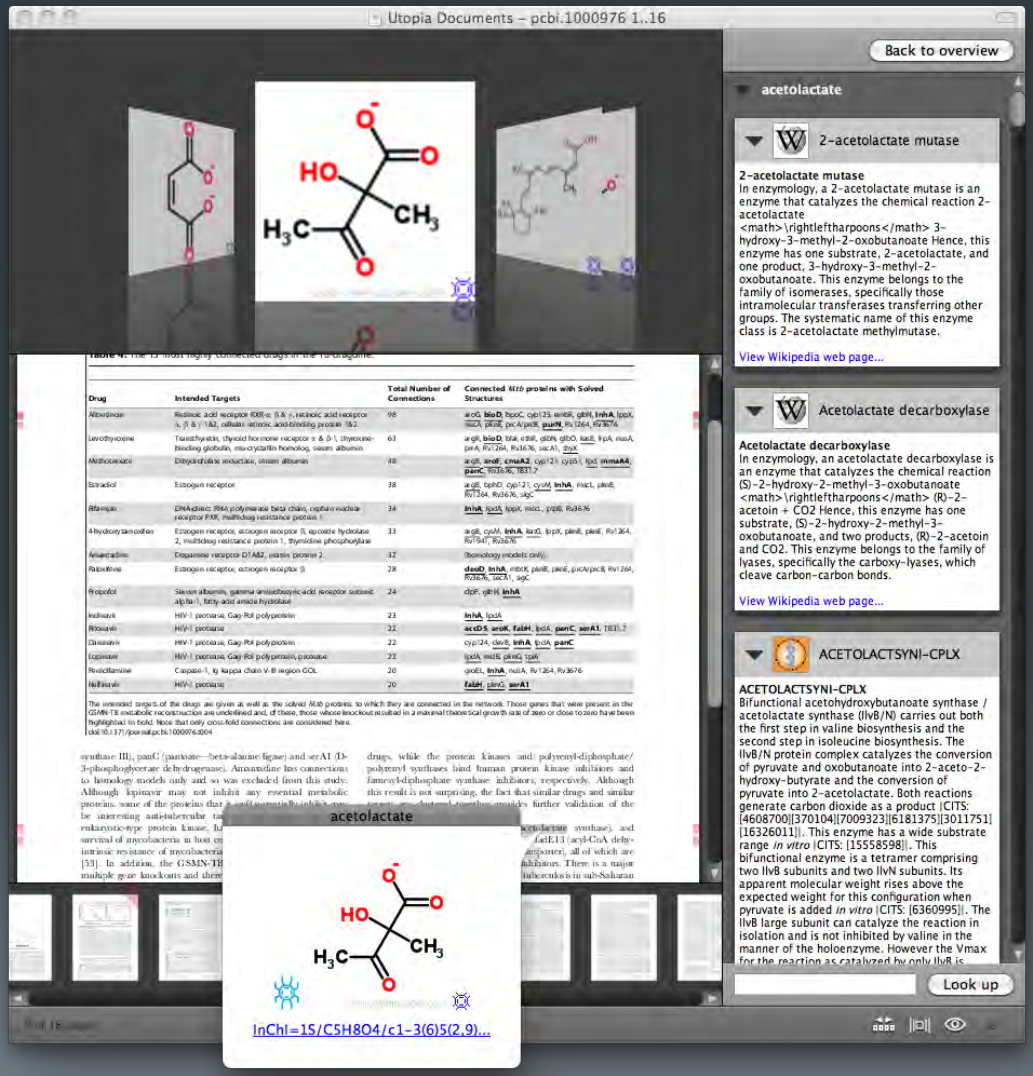
The ChemBio Navigator

http://chembionavigator.org/



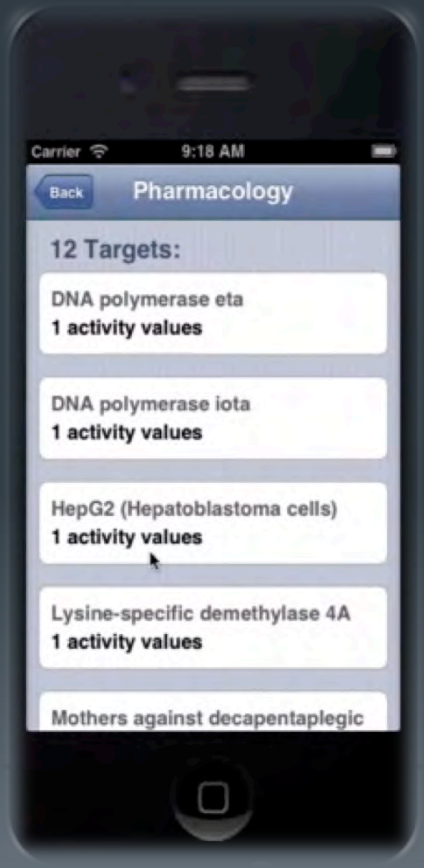
Utopia

http://getutopia.org

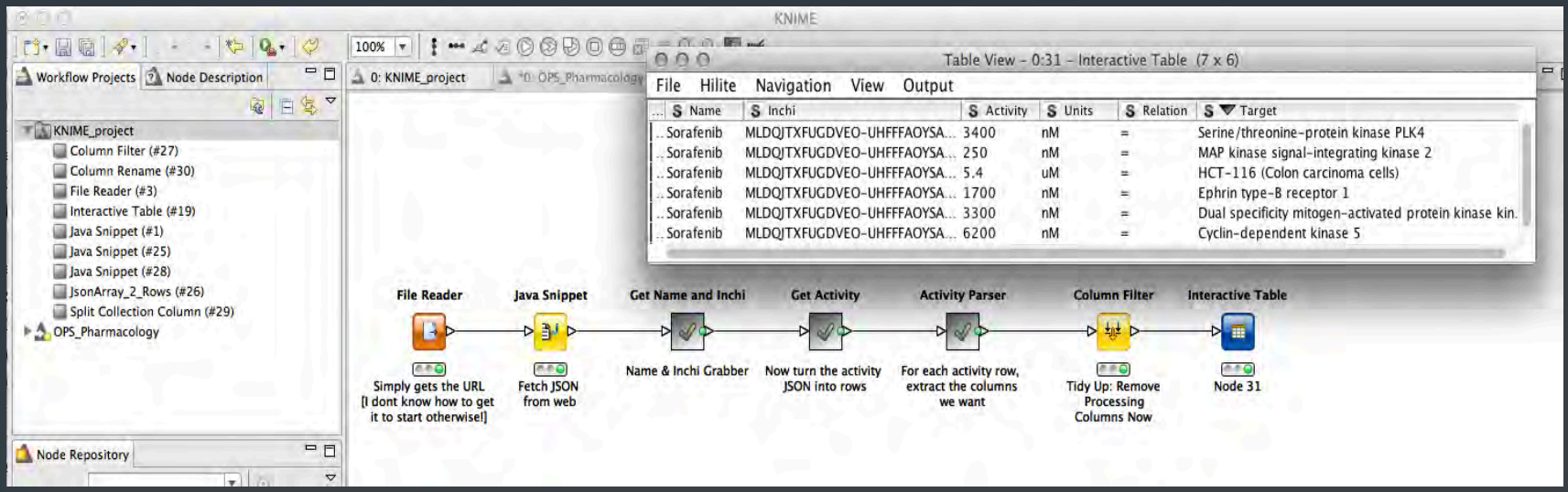


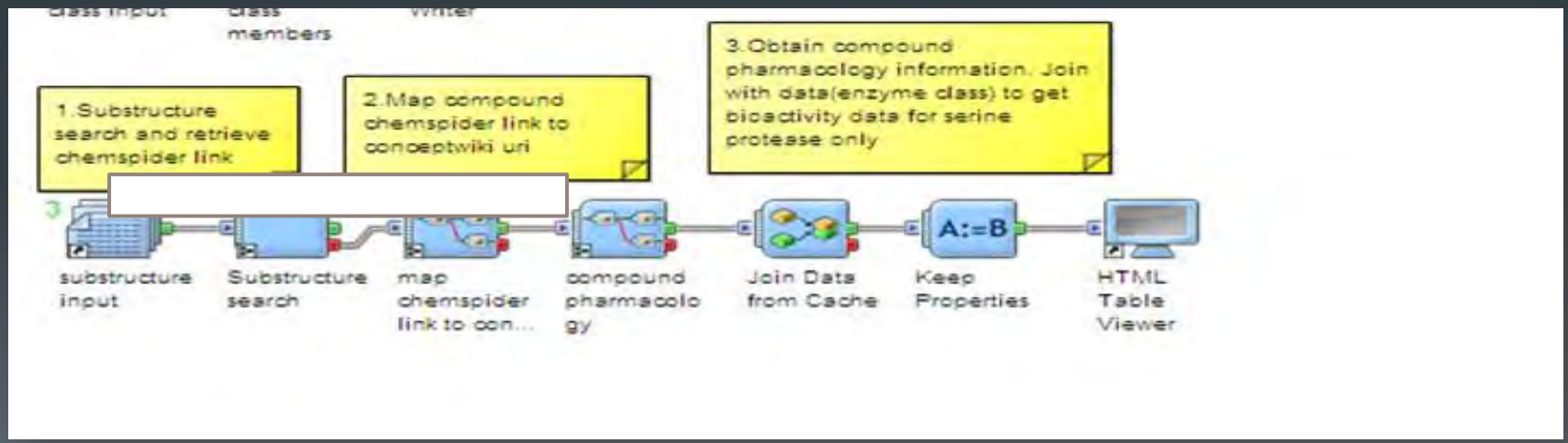
Hacking The iPhone

http://www.youtube.com/watch?v=0aGB6YqtuQ0



Workflow Tools



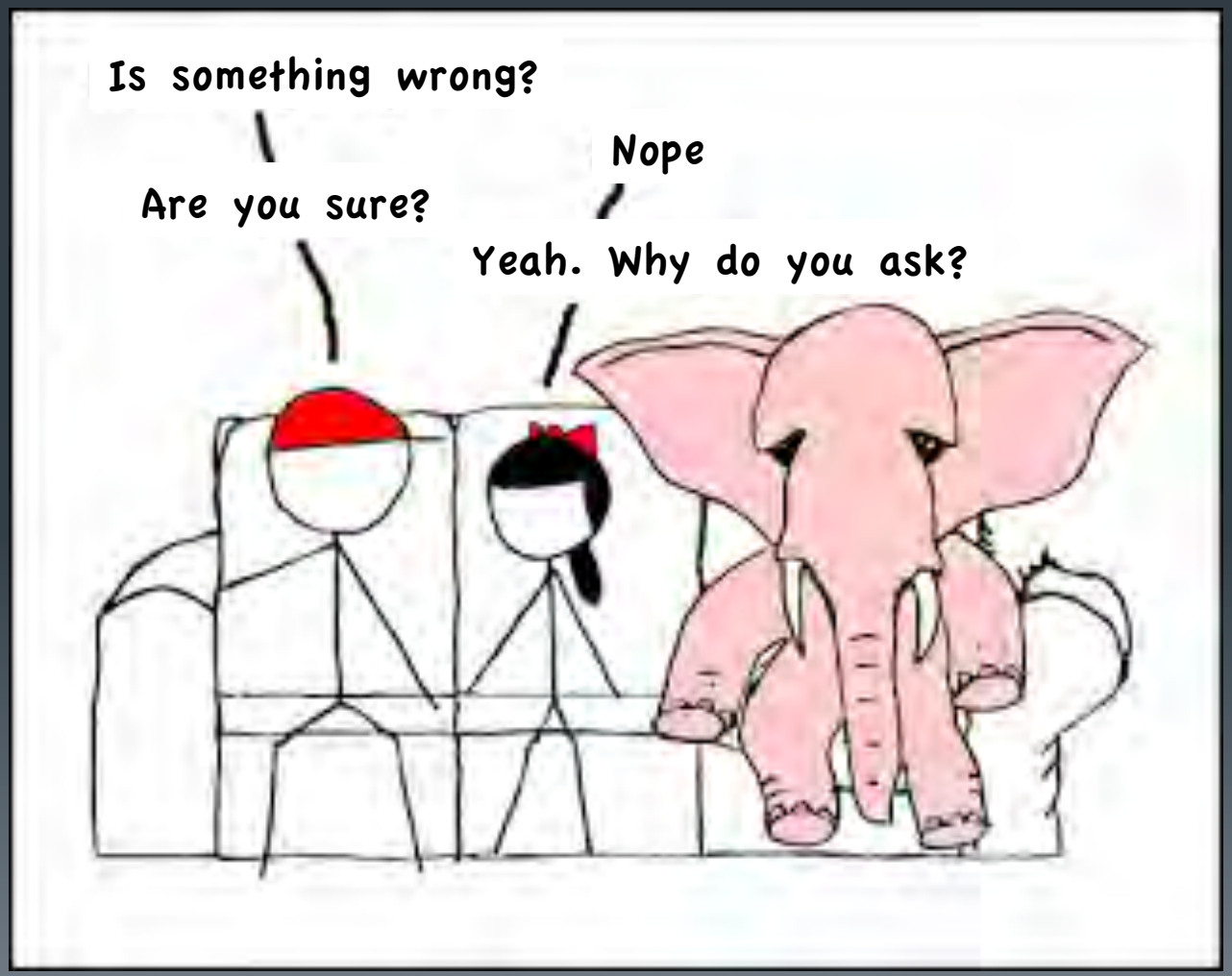


+ Google Refine, Helium For Excel, Taverna

Anyone can now ask…

- Who's working on what targets for what indications?
- What chemical matter has been tested against those targets?
- What does that chemical space look like?
- Whats been patented in that area?
- What other targets do those compounds hit?
- Who's working on those targets for what indications…?

Security?



Security: Still A Concern For Some

- But more companies moving to cloud externalisation
- Starting to see emergence of security principles (Pistoia, Open PHACTS)
- VM’s (local, remote)
- Clever workflows
- Will opportunity overtake fear?

Conclusions

- Accessible API's increasing
- Let people explore, they'll find new use cases
- Can't build a big interface on big data
- Usability, Discovery & Security key challenges

Acknowledgements

- SureChem Limited: http://surechem.com
- SciBite Limited: http://scibite.com
- Open PHACTS Consortium: http://openphacts.org | http://dev.openphacts.org | http://explorer.openphacts.org