

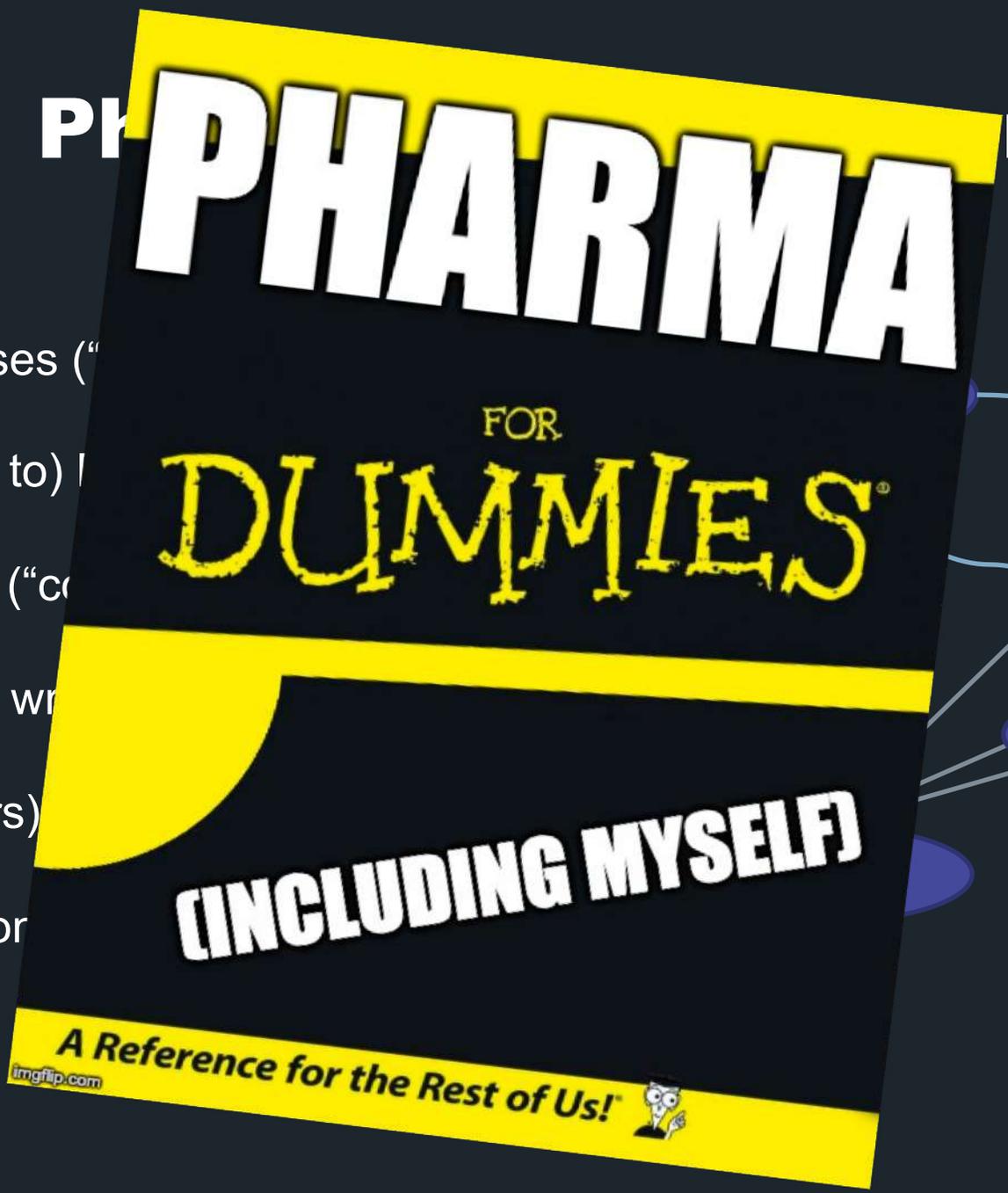
Making data FAIR with the Pharma Research Hub

Emil Zegers, MarkLogic

emil.zegers@marklogic.com

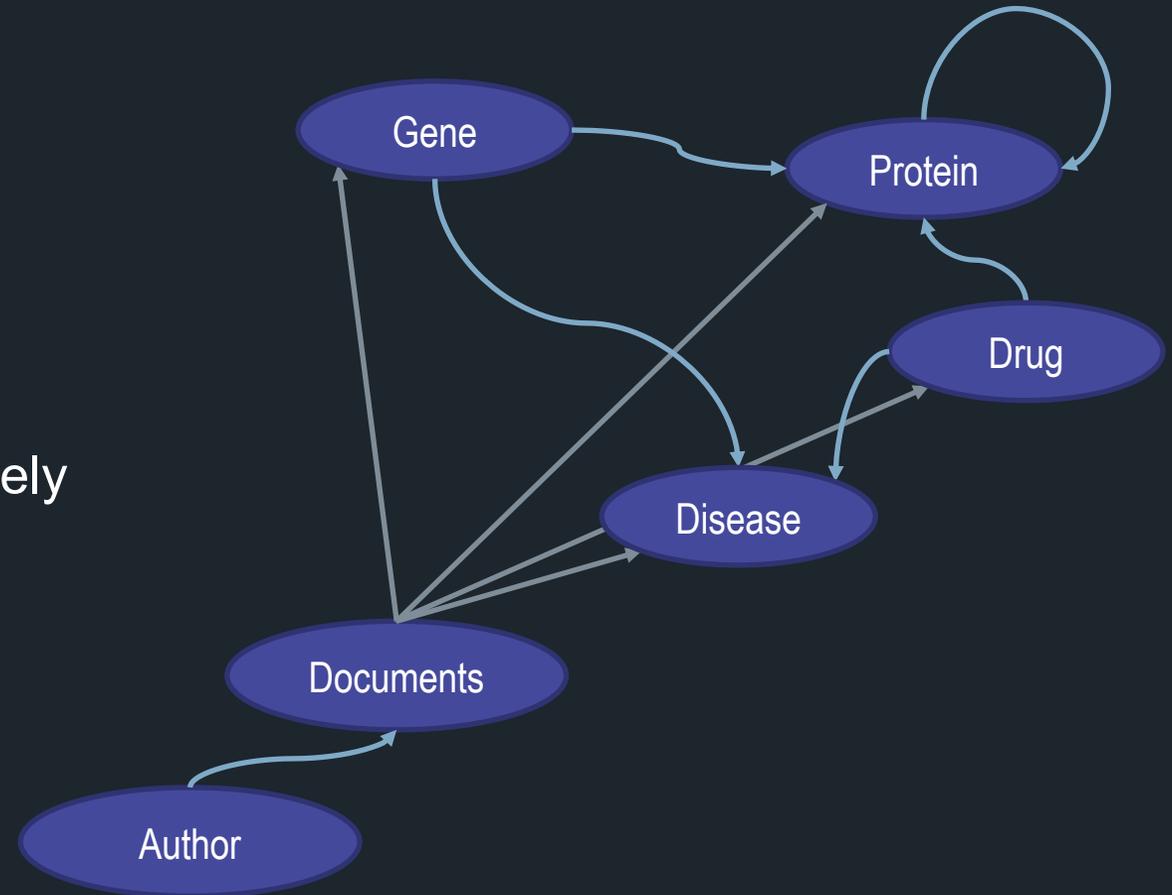
PHARMACOLOGY

- Drugs treat Diseases (“cure”)
- Drugs target (bind to) Proteins
- Genes synthesize (“code for”) Proteins
- And all this stuff is written in Books
- By People (Authors)
- Proteins mediate or regulate

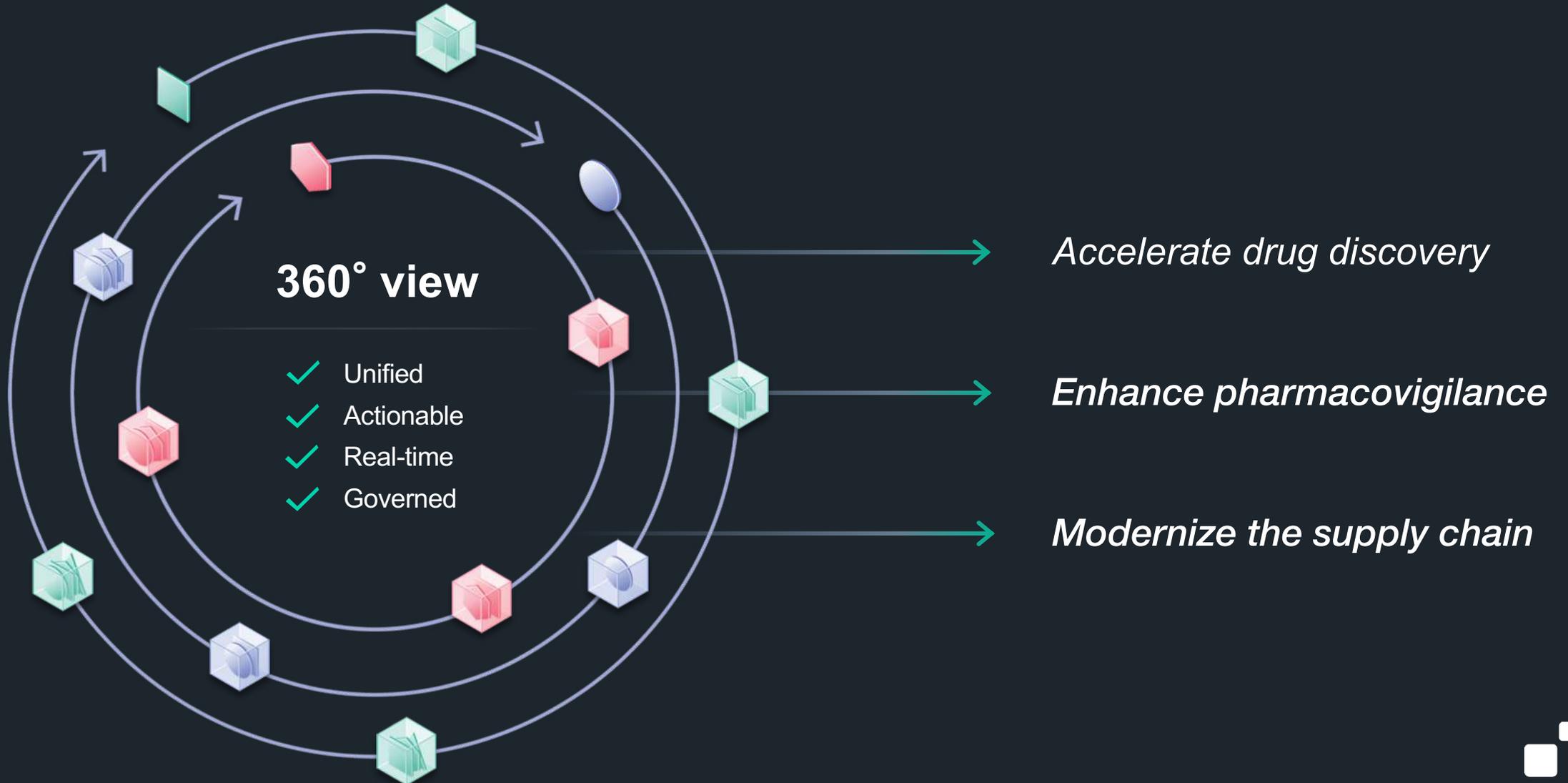


Pharma 101: A Primer

- Drugs treat Diseases (“Indications”)
- Drugs target (bind to) Proteins
- Genes synthesize (“code for”) Proteins
- And all this stuff is written about, extensively
- By People (Authors)
- Proteins mediate one another



Organizations need a 360° view



A 360° view of Research Assets

With that 360° view (new!) scientists, sales and management can answer questions such as:

- What are the trials we are running on IL23?
- What are the indications is this for?
- What other companies / partners are doing this?
- Are there any assay that we have? How recent are the assays?
- What biorepositories are available?
- What ELNs are available? Can I see them?
- What valuable information packages can we deliver through our sales channel?

W en.wikipedia.org › wiki › Interleukin_23 ▾

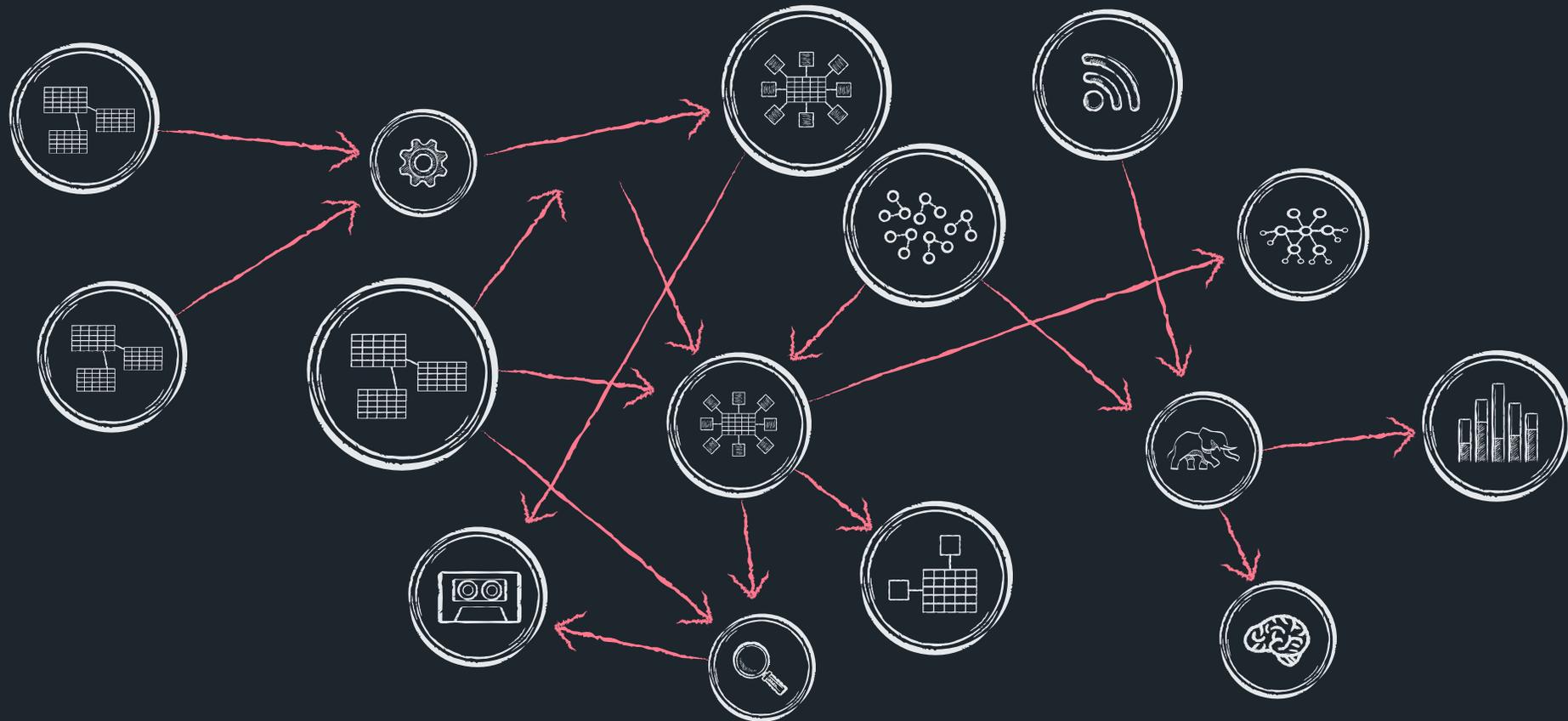
[Interleukin 23 - Wikipedia](#)

Interleukin-23 (IL-23) is a heterodimeric cytokine composed of an IL12B (IL-12p40) subunit (that is shared with IL12) and the IL23A (IL-23p19) subunit. A functional receptor for IL-23 (the IL-23 receptor) has been identified and is composed of IL-12R β1 and IL-23R.

An Electronic Laboratory Notebook (**ELN**), also referred to as an electronic lab notebook, is a software program designed to replace paper laboratory notebooks. Lab notebooks in general are used by scientists and technicians to document research, experiments and procedures performed in a laboratory.



The reality is data is (still) in silos



Pharma Research Hub Platform



Data sets in the Pharma Research Hub

- NIH's NCBI on genes
- Protein data from UniProt's curated data set, along with UniProt gene relationships
- PubMed Central's publication and author content
- US FDA's information on Drugs
 - Plus the RxNorm drug ontology, listing all drugs
- the SNOMED ontology, with disease and other information
- MedGen gene-to-disease mapping files

* Pharma Data Hub is built to load any data set in your company, from biological, to study or trial data, to person or marketing data, and link it together. We have a demonstration with "open" data sets, because we obviously cannot show our customers proprietary data, but the intent is to include both public and company private data and people

Demo Content

- A visual user interface on all data and domain knowledge
- Data loaded *as-is* -> JSON, XML, Relational, Documents, RDF Ontologies
- Both structured and unstructured data sources
- Multi-model approach:
 - Search
 - Semantic graph overarching data silos
- Workspaces



Demo



The MarkLogic Pharma Research Hub

A single pane of glass that provides easy access to the widest possible array of R&D data available

Healthcare & Life Sciences

SECTORS

- Pharmaceutical
- Medical device
- Health insurance
- Federal and state payers
- Hospital systems/ACOs
- Provider tech
- Medical associations

CUSTOMERS

Global Pharmas Trust MarkLogic

By working with *five of the largest global pharmaceutical companies and other leaders in the healthcare ecosystem*, MarkLogic has long used its Data Hub technology to help solve pharmaceutical companies' business and data challenges

USE CASES

- Pharmaceutical supply chain
- Real world evidence platform
- Research & Development Portal
- Pharmacovigilance
- Clinical data hub
- MMIS modernization
- Healthcare.gov **ObamaCare**





Thank you! (are we FAIR enough?)

Making data FAIR with the Pharma Research Hub

Check our blog series "FAIR for Pharma" at <https://www.marklogic.com/blog>

Emil Zegers, MarkLogic
emil.zegers@marklogic.com

Blog series FAIR for Pharma

1. Transform Your Enterprise Data with FAIR Principles
<https://www.marklogic.com/blog/transform-your-enterprise-data-with-fair-principles/>
2. Embracing Open Standards for Accessible Data in the Pharma Industry
<https://www.marklogic.com/blog/embracing-open-standards-for-accessible-data-in-the-pharma-industry/>
3. Interoperable Data in the Pharma Industry: Achieving a 'Single Source of Truth'
<https://www.marklogic.com/blog/interoperable-data-in-the-pharma-industry-achieving-a-single-source-of-truth/>
4. Faster R&D in the Pharma Industry through Re-Use of Critical Data
<https://www.marklogic.com/blog/faster-rd-in-the-pharma-industry-through-re-use-of-critical-data/>



Pharma Research Hub background info



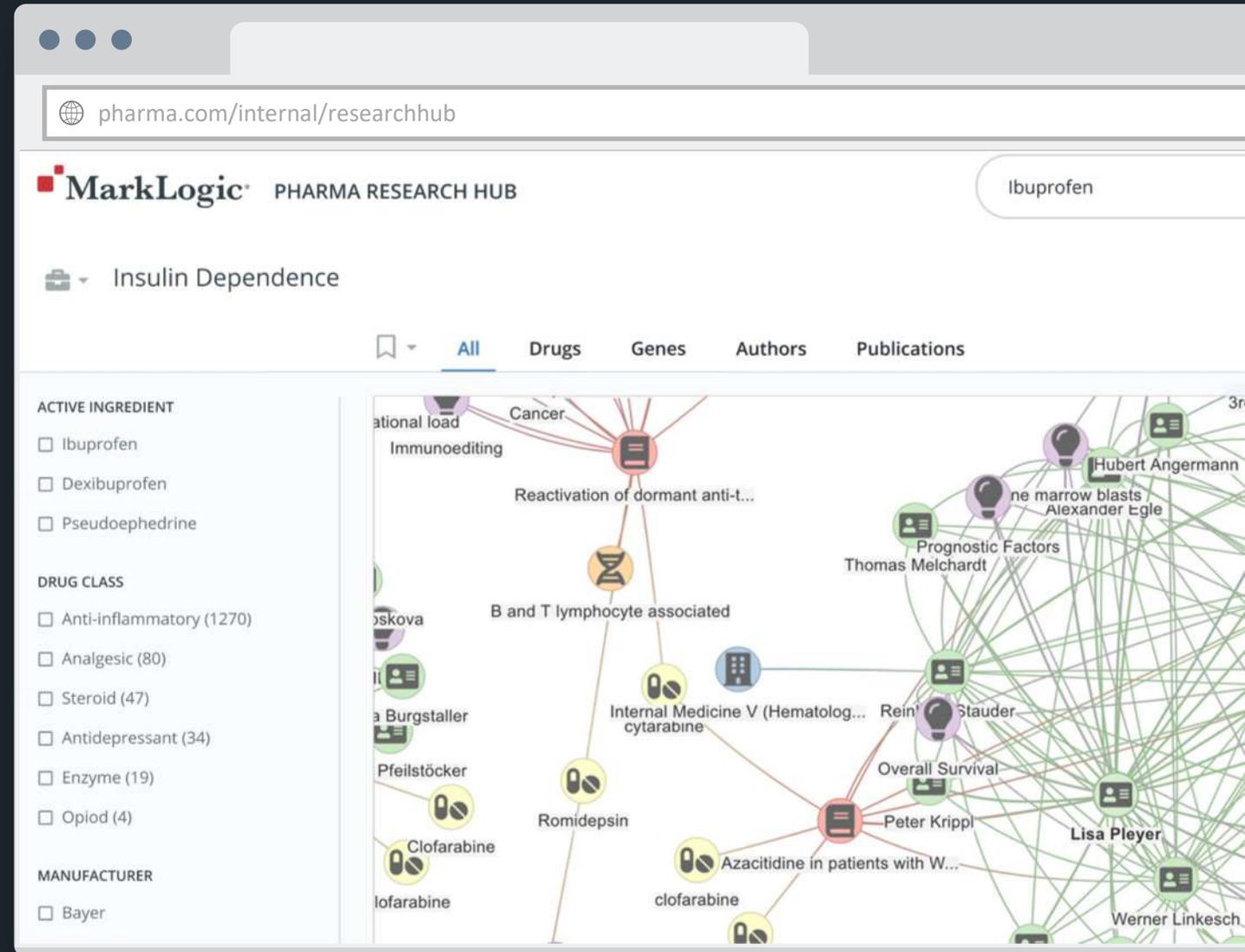
Filtered results

Search across drugs, genes, authors, publications and use facets to filter the results.

The screenshot shows a web browser window with the URL `pharma.com/internal/researchhub`. The page header includes the MarkLogic logo and 'PHARMA RESEARCH HUB'. A search bar on the right contains the text 'Ibuprofen'. Below the header, a breadcrumb trail shows 'Insulin Dependence'. A navigation menu at the top right has tabs for 'All', 'Drugs', 'Genes', 'Authors', and 'Publications', with 'Drugs' selected. A dropdown menu is open over the 'Drugs' tab, showing options: 'Select all', 'Deselect all', and 'Add selected to projects'. The main content area displays two search results. The first result is for 'Ibuprofen', with a 'DRUG' tag, a description: 'commonly used nonsteroidal antiinflammatory (NSAID) drug which is available both by prescription and over-the-counter...', and 'Active ingredients: Ibuprofen'. The second result is for 'Seractil', with a 'DRUG' tag, a description: 'Seractil is a type of medicine called a non-steroidal anti-inflammatory drug (NSAID). NSAIDs are used to relieve pain and inflammation...', and 'Active ingredients: Dexibuprofen'. On the left side, there are three filter sections: 'ACTIVE INGREDIENT' with checkboxes for 'Ibuprofen', 'Dexibuprofen', and 'Pseudoephedrine'; 'DRUG CLASS' with checkboxes for 'Anti-inflammatory (1270)', 'Analgesic (80)', 'Steroid (47)', 'Antidepressant (34)', 'Enzyme (19)', and 'Opioid (4)'; and 'MANUFACTURER' with a checkbox for 'Bayer'.

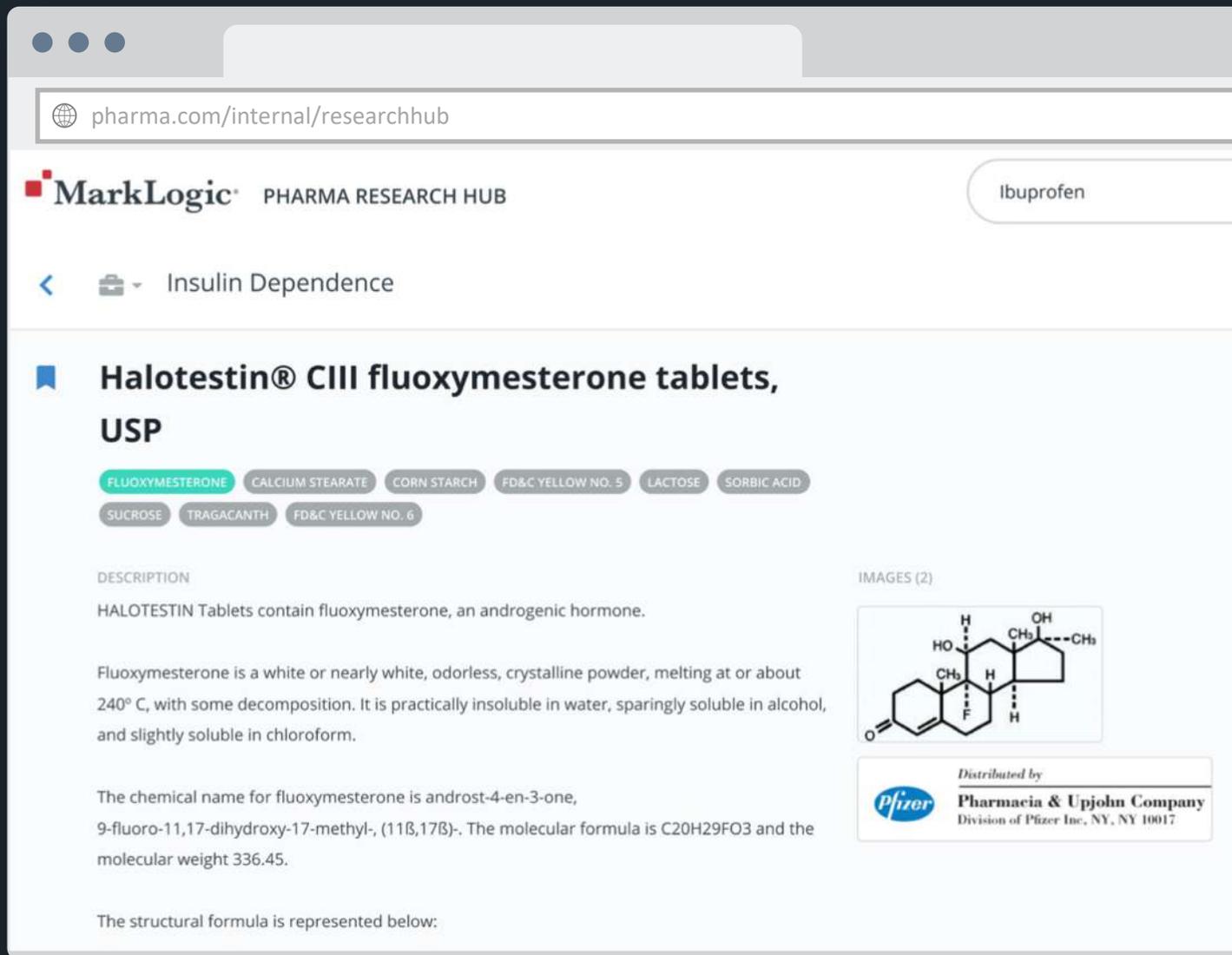
Graph of connections

Toggle the graph view to see connections between entities



Rich information

See aggregate info for each entity,
including related topics



The screenshot shows a web browser window with the URL `pharma.com/internal/researchhub`. The page header includes the MarkLogic logo and "PHARMA RESEARCH HUB". A search bar on the right contains the text "Ibuprofen". Below the header, a breadcrumb trail shows a back arrow, a briefcase icon, and the text "Insulin Dependence".

Halotestin® CIII fluoxymesterone tablets, USP

FLUOXYMESTERONE CALCIUM STEARATE CORN STARCH FD&C YELLOW NO. 5 LACTOSE SORBIC ACID
SUCROSE TRAGACANTH FD&C YELLOW NO. 6

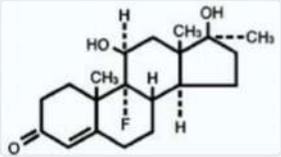
DESCRIPTION
HALOTESTIN Tablets contain fluoxymesterone, an androgenic hormone.

Fluoxymesterone is a white or nearly white, odorless, crystalline powder, melting at or about 240° C, with some decomposition. It is practically insoluble in water, sparingly soluble in alcohol, and slightly soluble in chloroform.

The chemical name for fluoxymesterone is androst-4-en-3-one, 9-fluoro-11,17-dihydroxy-17-methyl-, (11 β ,17 β)-. The molecular formula is C₂₀H₂₉FO₃ and the molecular weight 336.45.

The structural formula is represented below:

IMAGES (2)



The chemical structure of fluoxymesterone is shown as a steroid nucleus with a ketone group at C-3, a double bond between C-4 and C-5, a fluorine atom at C-9, a methyl group at C-10, a hydroxyl group at C-11, and a methyl group at C-17. Stereochemistry is indicated with wedges and dashes.

Distributed by
Pfizer Pharmacia & Upjohn Company
Division of Pfizer Inc, NY, NY 10017

Data Integration

- Pharma Research Hub shows how six data sources (silos) were integrated
 - Drug data
 - Drug ontology with ingredients
 - Gene database
 - Protein database
 - Publications
 - Authors extracted from Publications



Multi-model

- Most data is mapped into JSON records
- The “content” is in XML to support text markup
- Inputs include:
 - CSV/Relational – NCBI Genes; Gene/Disease relationships
 - XML – PubMed, UniProt proteins,
 - RDF/Ontology – RxNorm and SNOMED data for drugs, drug ingredients, diseases



Key features



Load any pharma data

Rapidly load any pharmaceutical data set, internal or external



Leverage AI and ML

Better search results on higher quality data with our Smart Mastering feature



Visualize relationships

Enhance search and define data relationships with Semantics



Create, share workspaces

Save data to workspaces, then share with colleagues



Notifications and alerts

Get notifications when new research data is available



Custom search

Customize search results to receive only relevant information

