



Realizing the Benefits of a Semantically-driven Clinical Smart Data Lake

Anzo Smart Data Lake© Enterprise Solutions

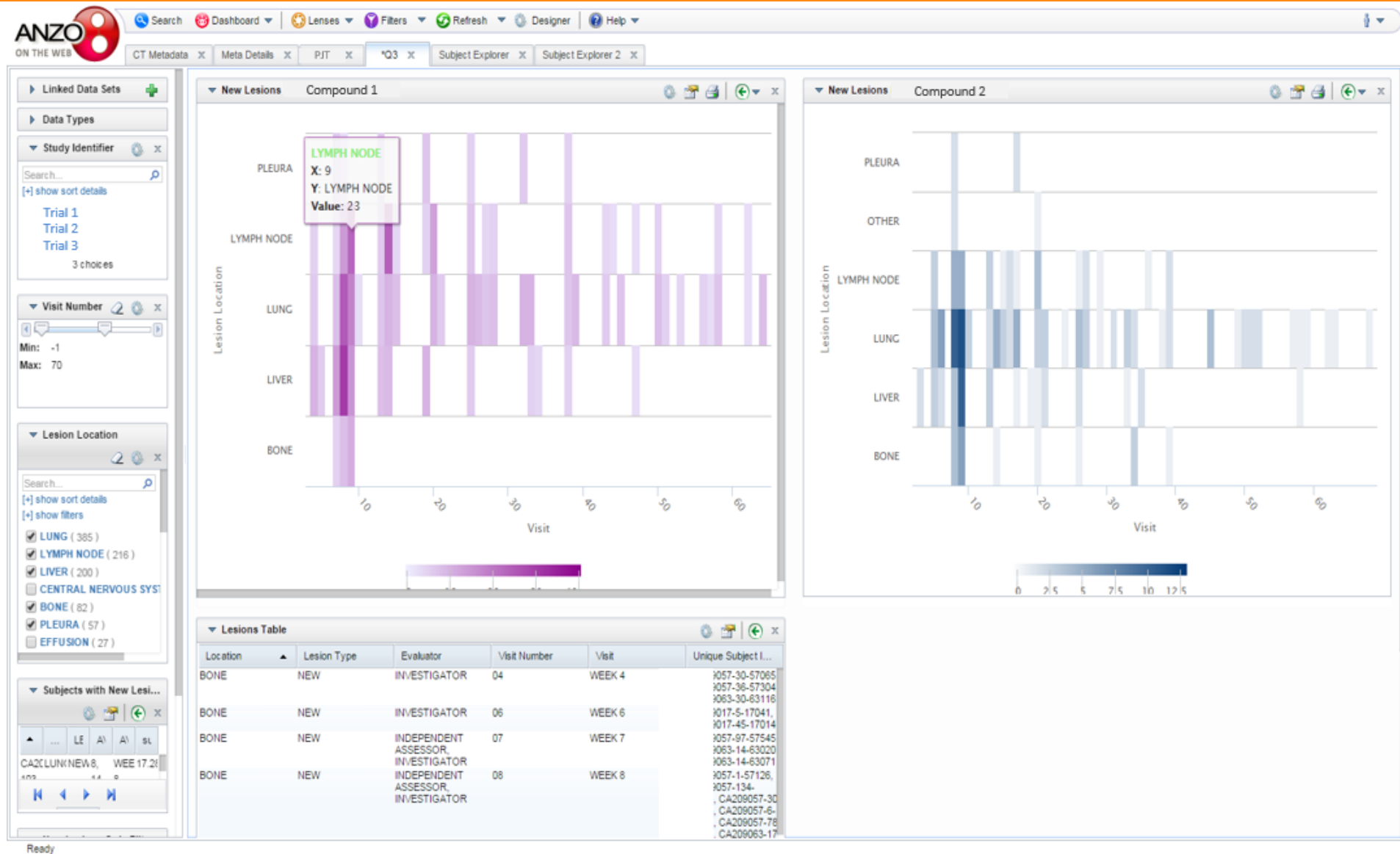
Jim LaPointe

Managing Director, Life Sciences & Healthcare

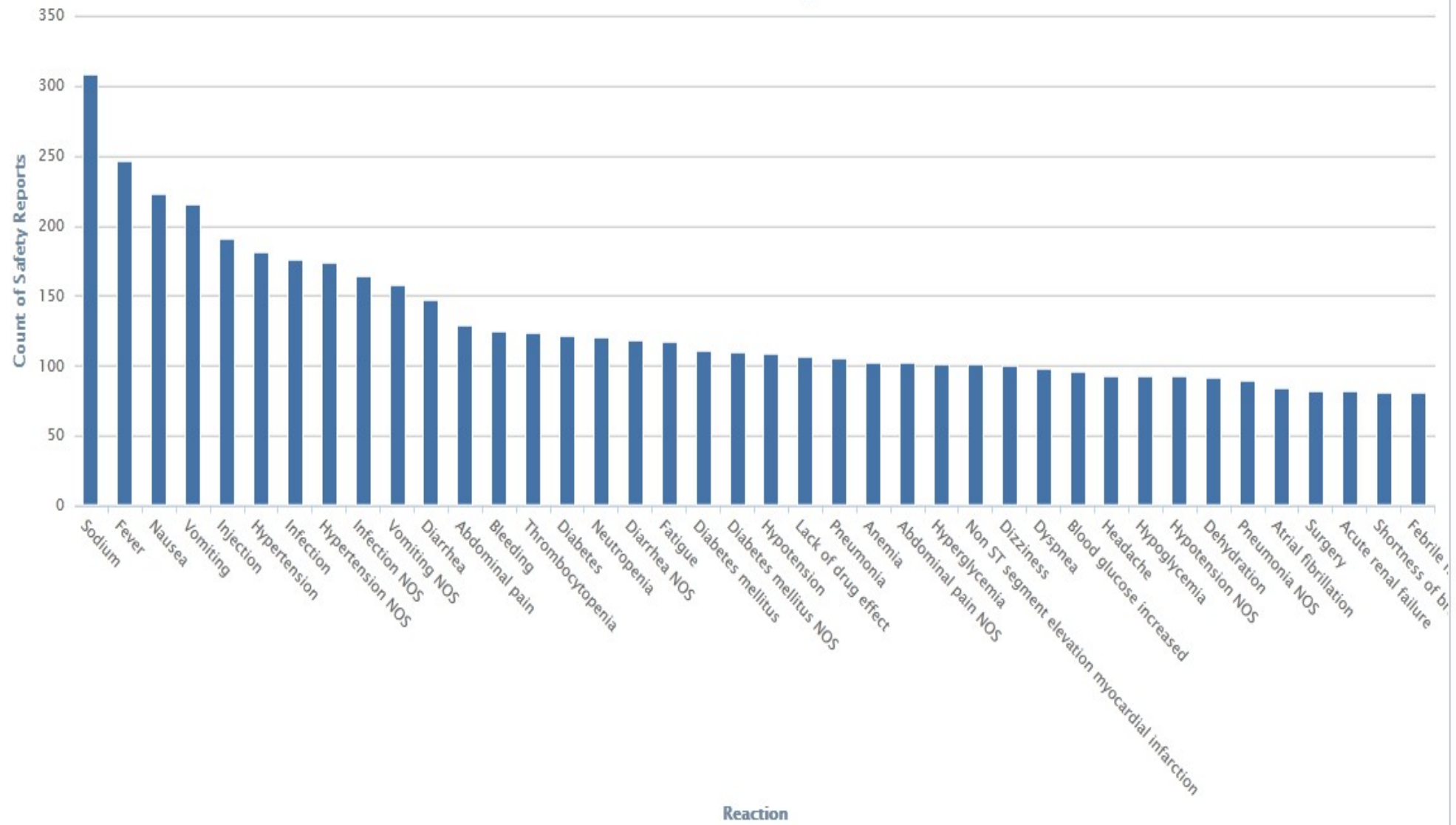
jtlapointe@cambridgesemantics.com

Clinical Smart Data Lake

Clinical Data Integration – Cross Trial Advanced Analytics

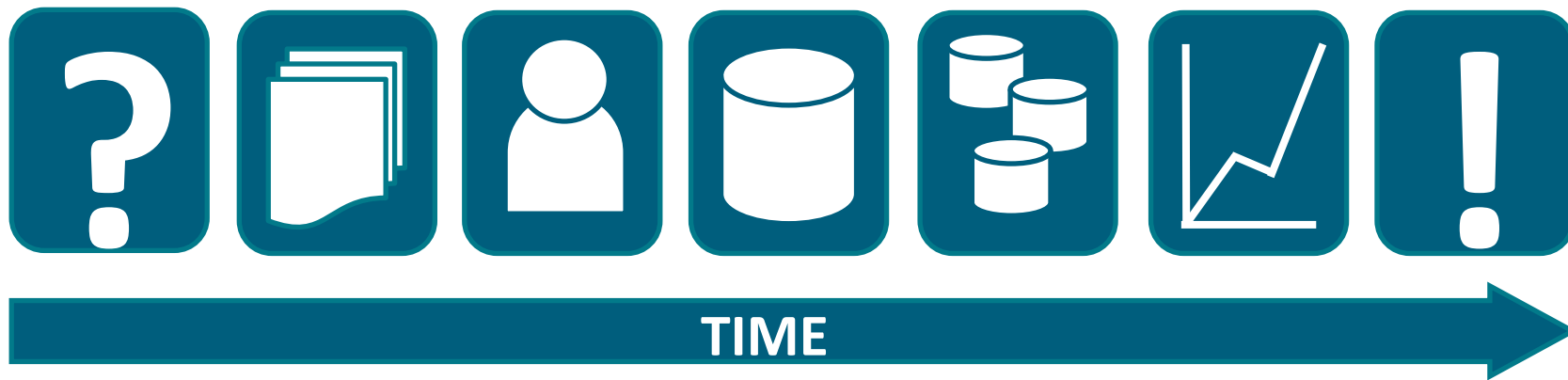


Adverse Reaction Histogram

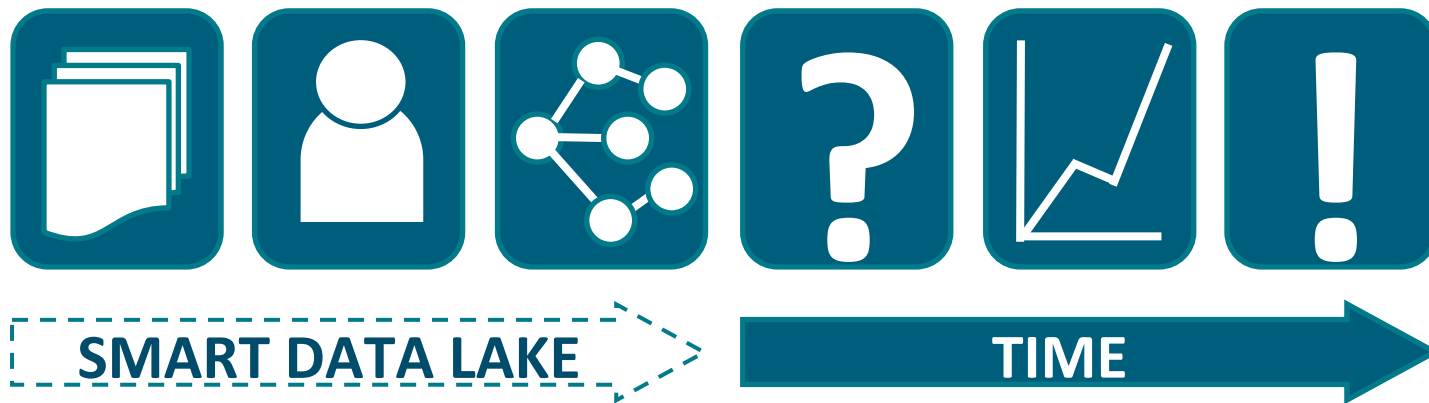


Any Questions?

Any Questions?

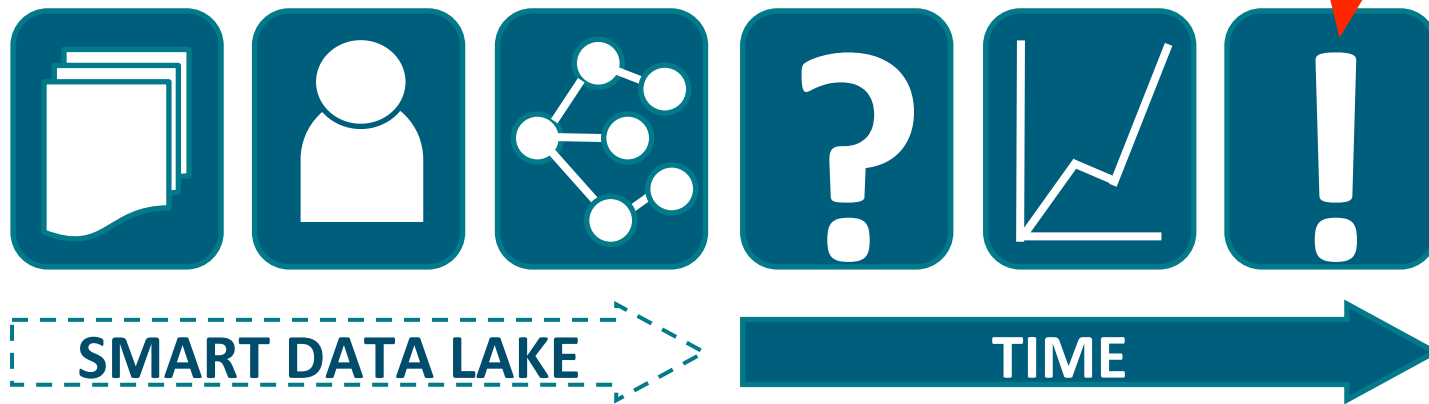


Any Questions?



Any Question?

Smart Data
VALUE:
reducing the time-
to-answer



Cambridge Semantics – Life Sciences Overview

*Cambridge Semantics develops the **Anzo Smart Data Platform®** and the ‘big data’ **Anzo Smart Data Lake®** technologies. The Anzo suite provides an end-to-end graph-based knowledge management platform with sophisticated data integration, data management, and real-time, advanced analytics. Cambridge Semantics delivers Smart Data solutions for the Life Sciences including R&D intelligence, Scientific Affairs, Medical Affairs, Clinical Operations & Data Management, PharmDev and much, much more...*

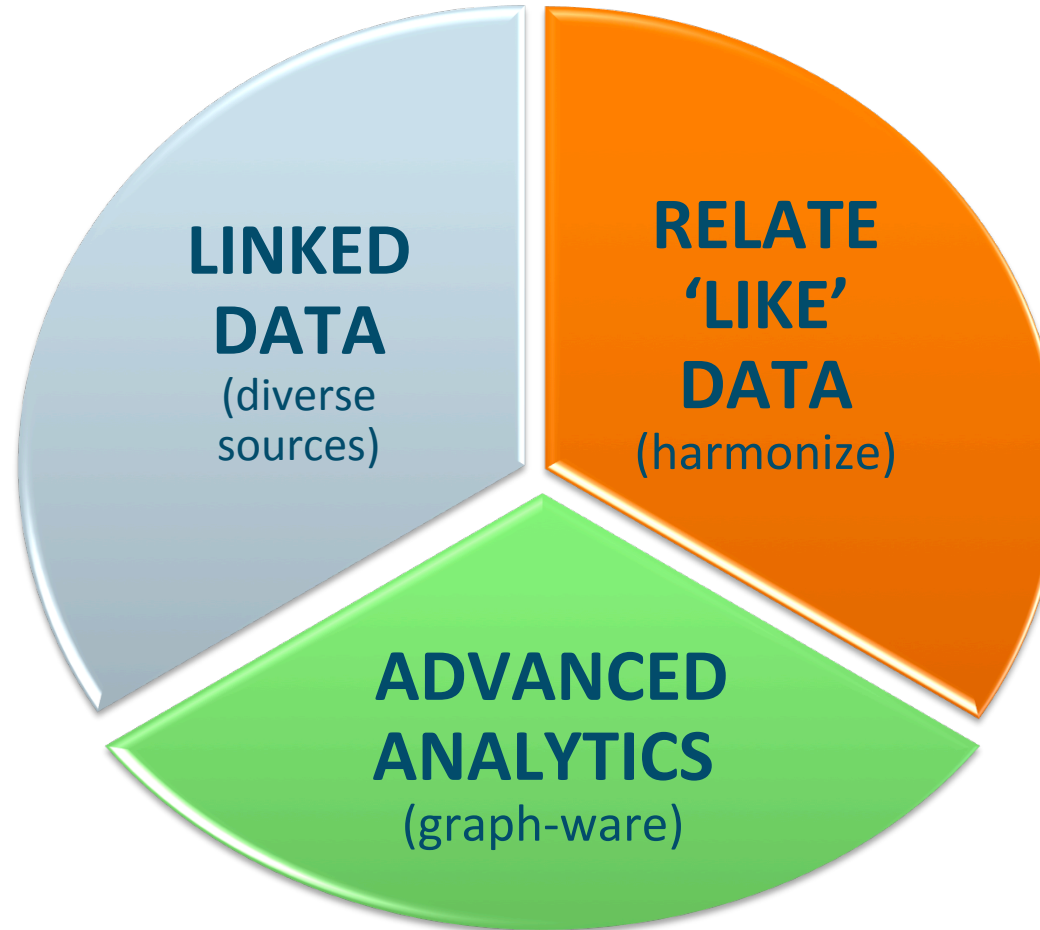
- Award-winning Anzo suite is our 3rd generation implementation of Semantic Web technologies
- Anzo is built on World Wide Web Consortium (W3C) open data standards and 14+ years of R&D
- Founded in 2007 by senior members of IBM’s Advanced Internet Technology Group
- Advanced analytics from the MPP graph DB team of Netezza & Paracel (Amazon Redshift)



What is SMART Data?

LINKED DATA

- Structured (Relational DB, Excel, ...)
- Unstructured (PDF, Word, email, ...)
- Semi-structured (XML, HTML, ...)
- Industry standard graph data structure (RDF* - Resource Description Framework)



ADVANCED ANALYTICS

- Immediately query ALL the data in the graph DB
- Simple self-service analytics with familiar Excel-like functions
- Industry standard query structures (SPARQL*)

RELATE 'LIKE' DATA

- Associate related concepts in a flexible & extensible graph DB
- Understand the data in a human-readable ontology
- Immediately operationalize the data in a machine-readable graph DB
- Industry standard data modeling (OWL* – Ontology Web Language)

* Semantics Web standards –
World Wide Web Consortium (W3C)

Future-proofing Smart Data - Semantic Web Standards

Semantic web technology built with World Wide Web Consortium (W3C) standards

- ✓ *Scale & diversity of data on the web*
- ✓ *Ever-changing environments*

Semantic web enables Smart Data by providing:

- ✓ RDF (Resource Description Framework) - A flexible and logical graph-based data model that can consider data from any source
 - => ***reduced development time and cost***
- ✓ OWL (Ontology Web Language) - An human understandable conceptual model of entities and relationships designed for business analysts and knowledge workers
 - => ***improved application quality, iterative / investigative approach***
- ✓ SPARQL – A complete, high-performance query language standard for accessing RDF data where IT-intensive skills are not needed to access the underlying data structures
 - => ***reduced development time and cost***

W3C Open Standards – A foundation for industry-specific standards

- ✓ Benefit from industry standards e.g. CDISC, MedDRA, SNOMED, OpenPhacts ...
 - => ***improved data quality & interoperability***

Smart Data – Linked & Harmonized Data Sources

Industry Intelligence Database

Company	Website	Mkt Cap
Bio Corp	biocorp.com	\$2.2B
Drugs123	drugs123.com	\$930M
...	...	...

Web news

On Tuesday, **Drugs123** Inc. announced **phase 1** development of their newest **sleep aid** therapeutic, **Narcoleptol**.

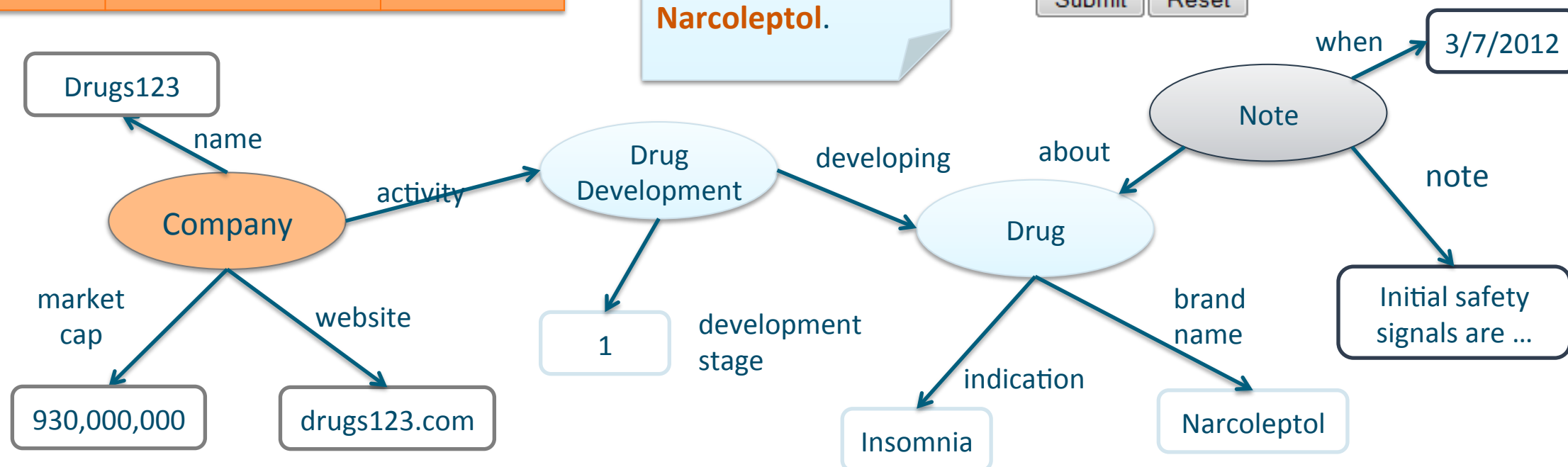
Intelligence Review Notes

Review Notes

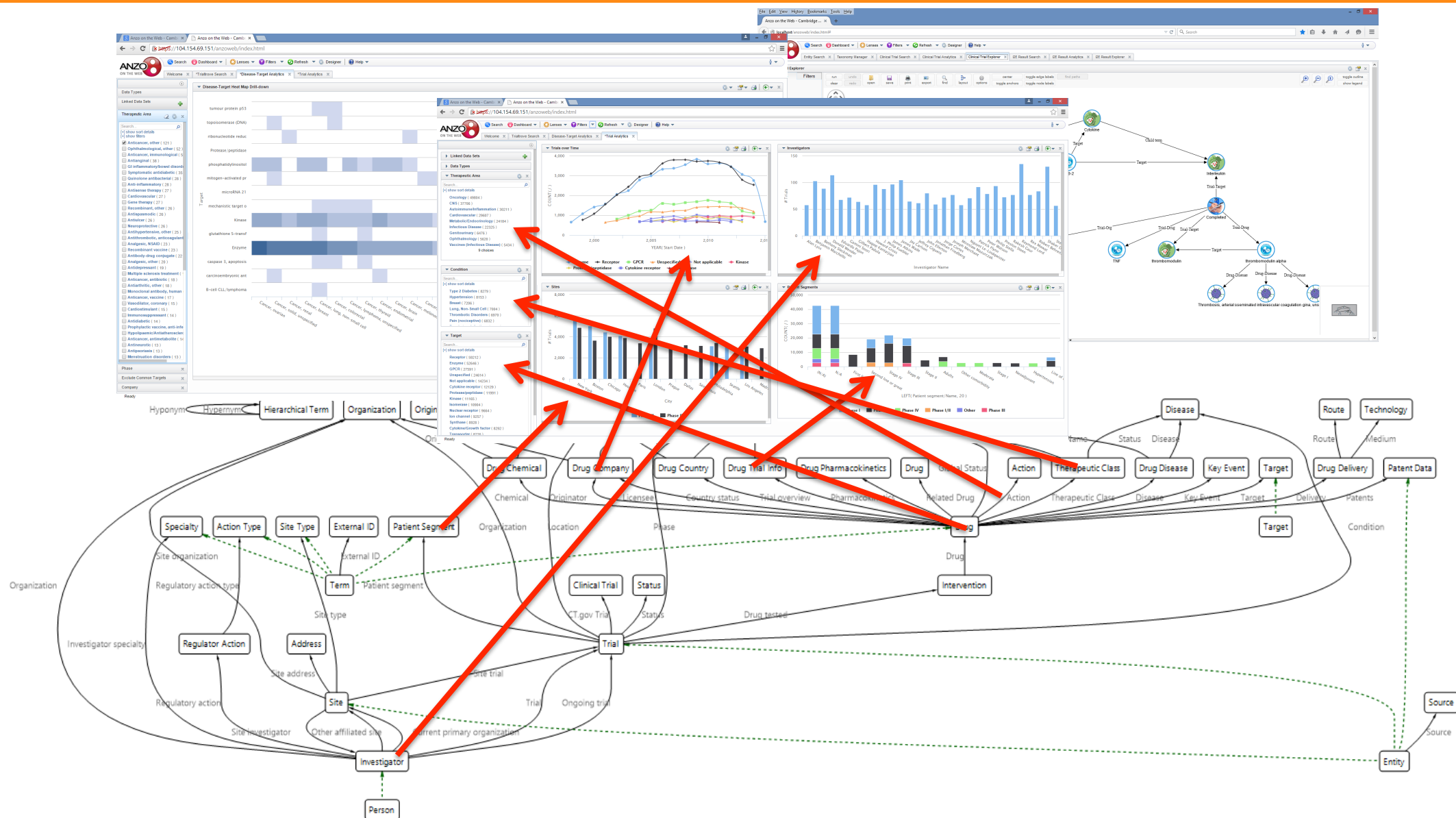
Drug:

Reviewed on:

Notes:



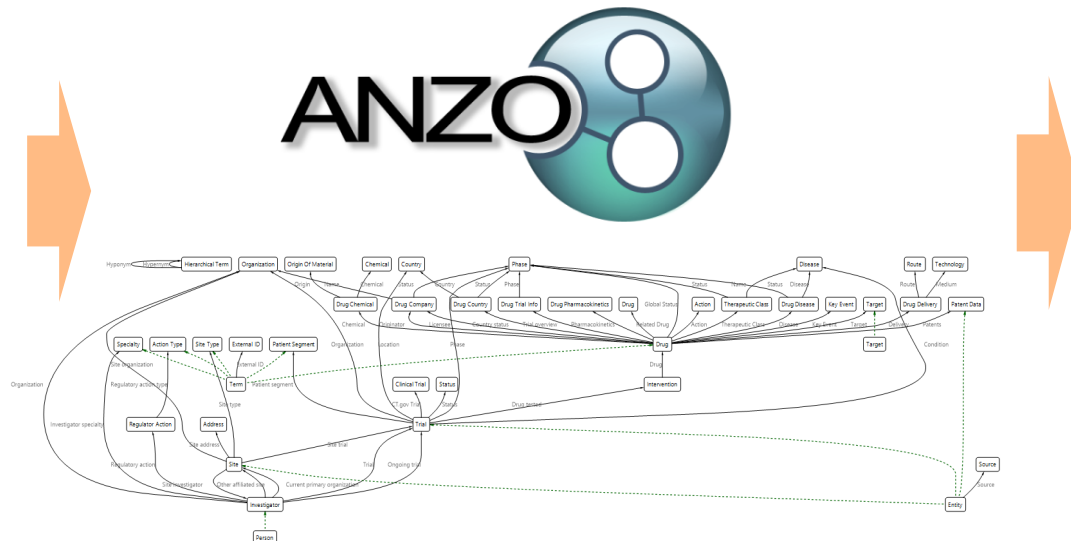
Smart Data – Graph-aware Advanced Analytics



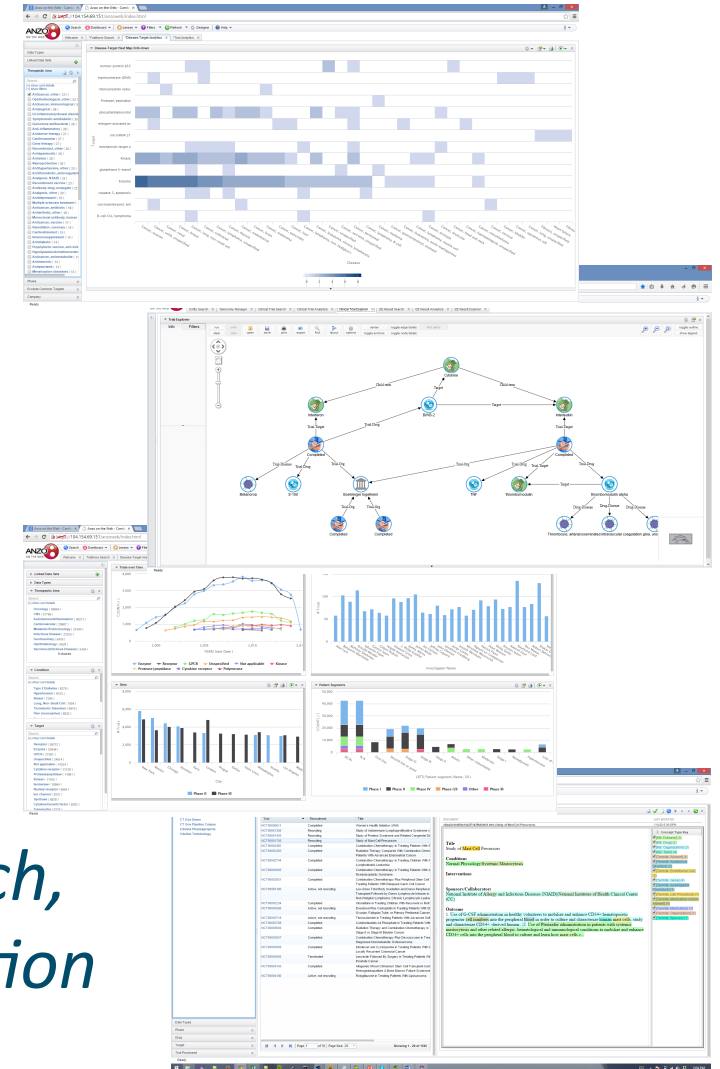
Enterprise Smart Data - Anzo Smart Data Platform©



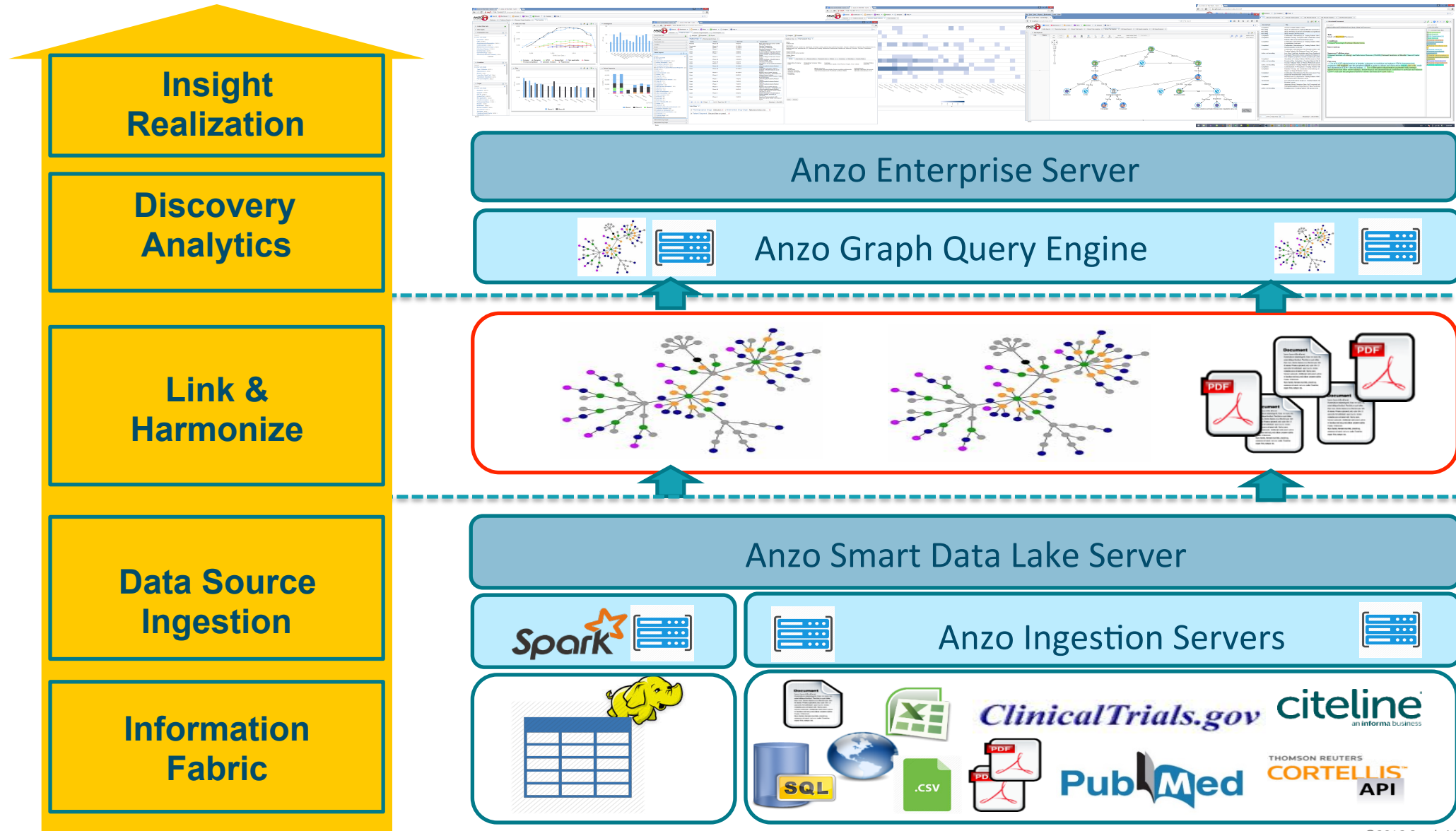
An agile, end-to-end, platform for tackling diverse data challenges



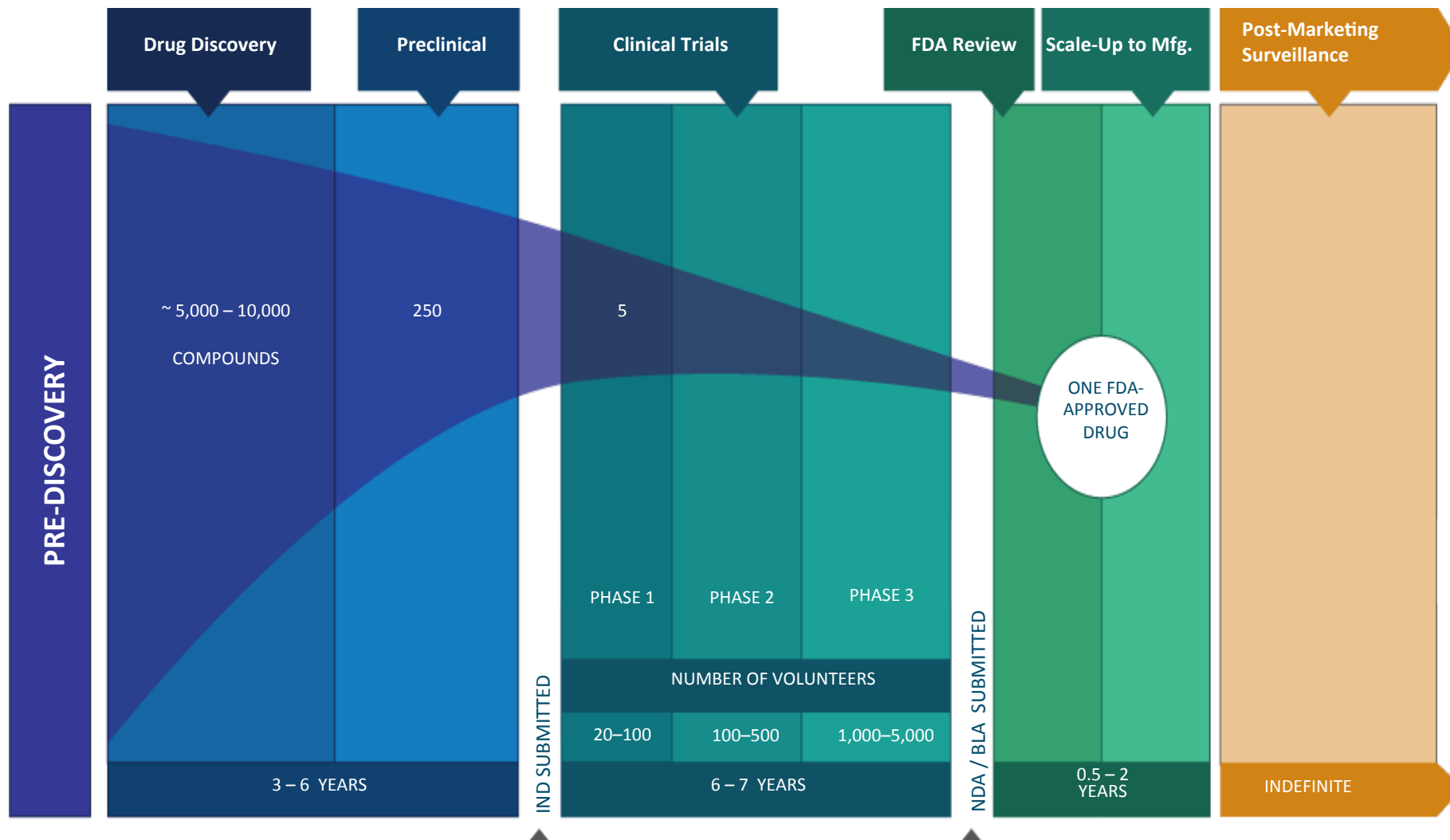
Link and harmonize data for search, advanced analytics, and collaboration



Smart 'Big Data' - Anzo Smart Data Lake©

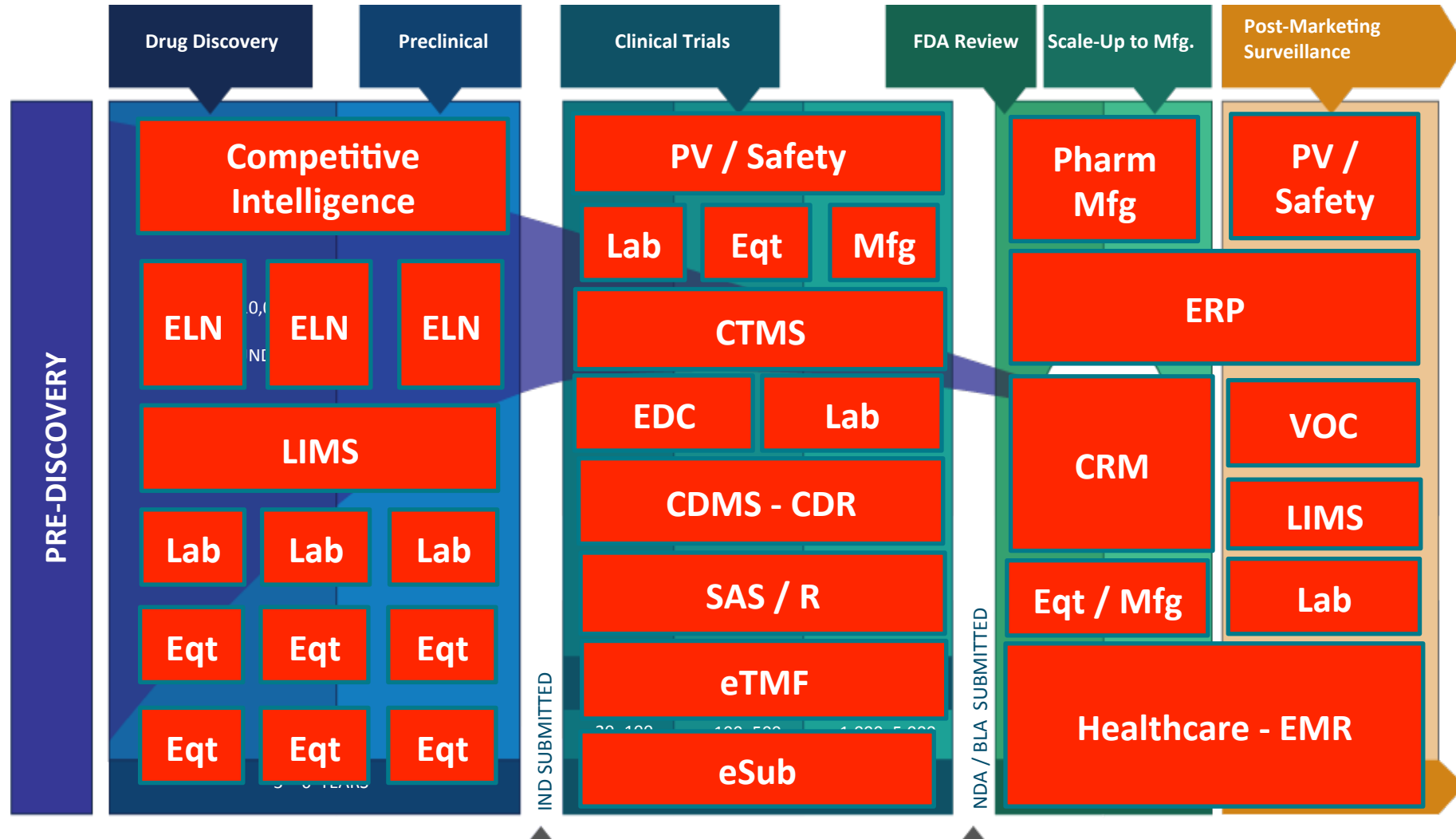


Drug / Biologic / Device R & D Takes 8 to 12 Years!



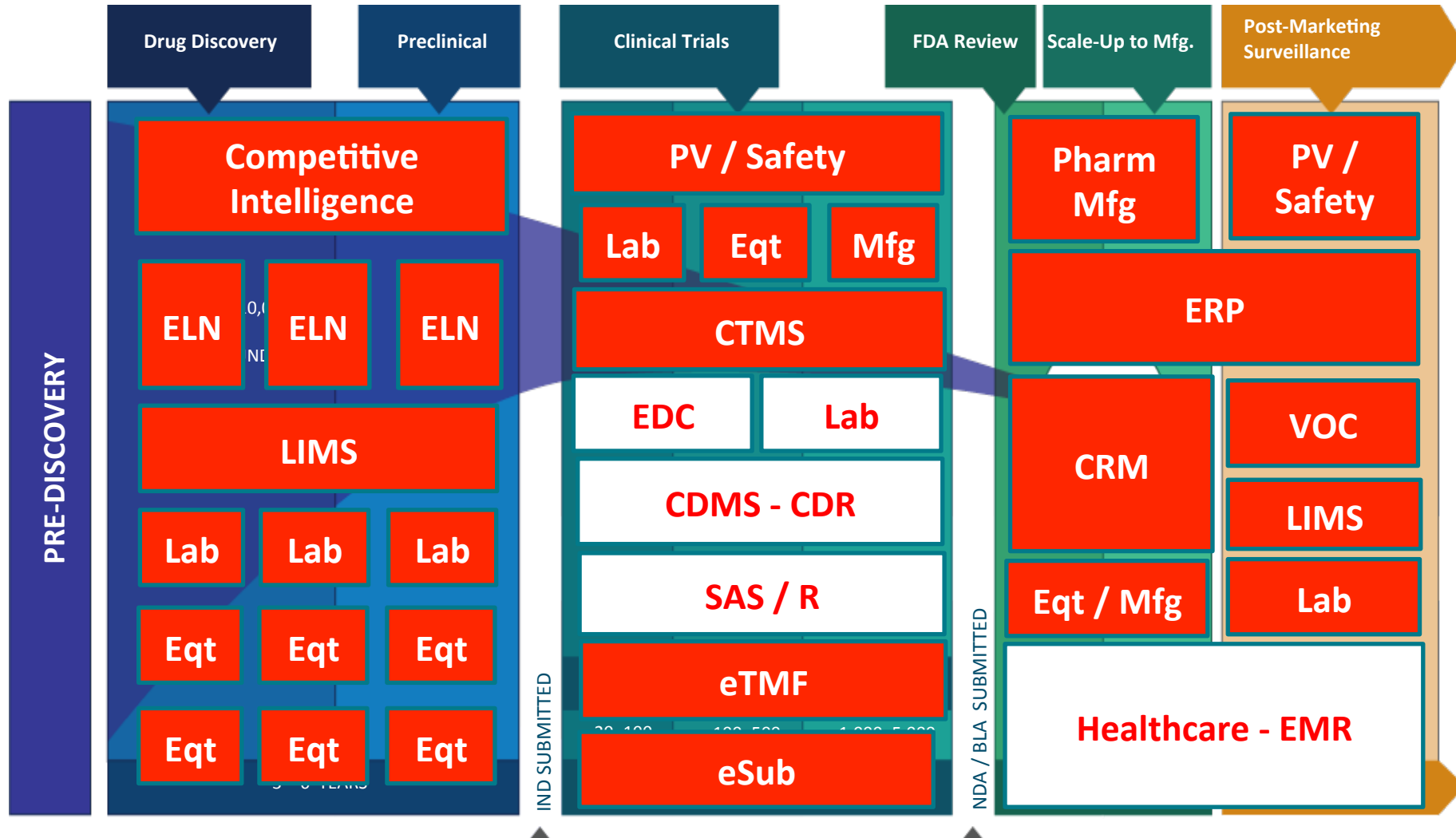
Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

Drug / Biologic / Device R & D Takes 8 to 12 Years!



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

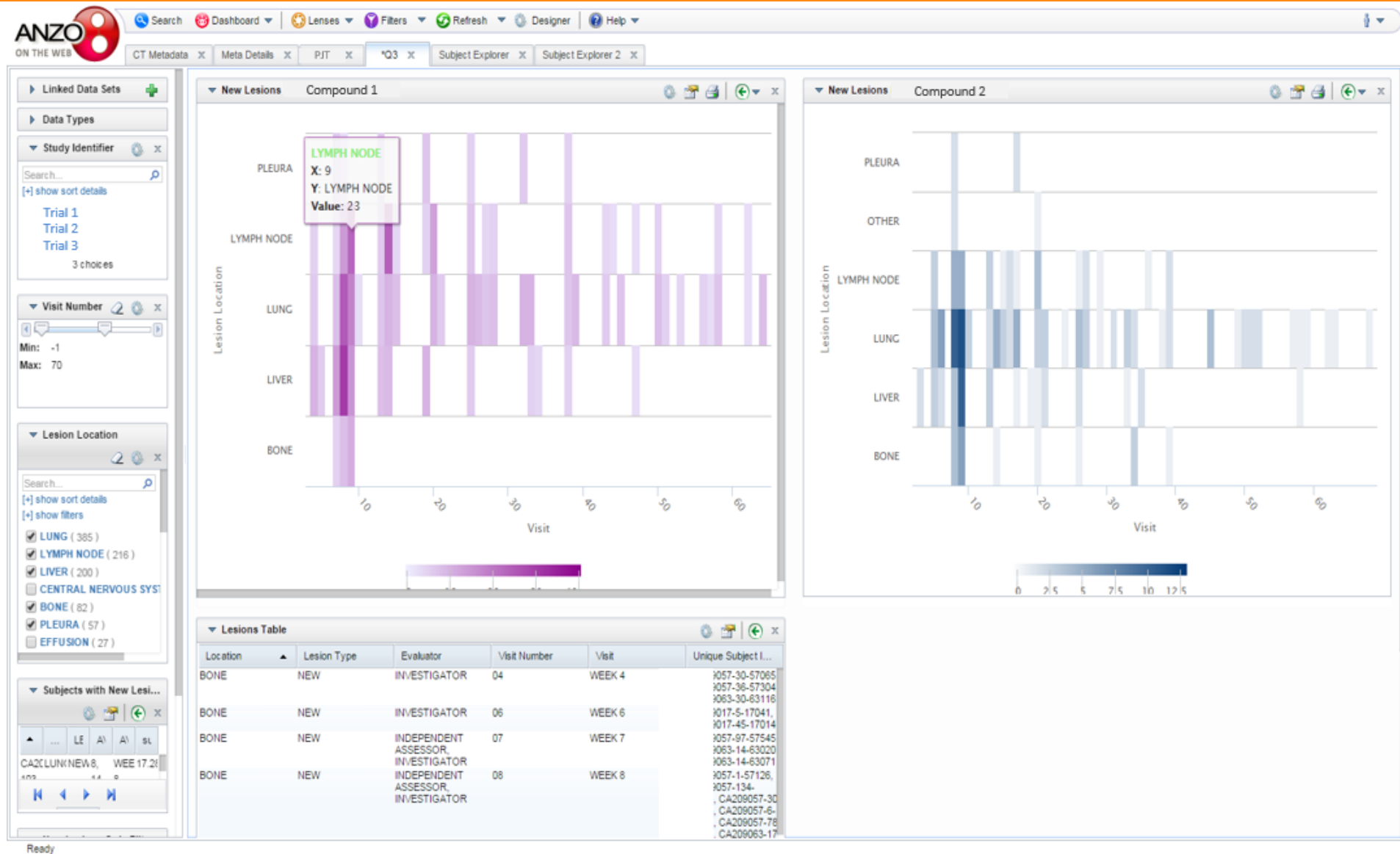
Clinical Smart Data Lake



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

Clinical Smart Data Lake

Clinical Data Integration – Cross Trial Advanced Analytics

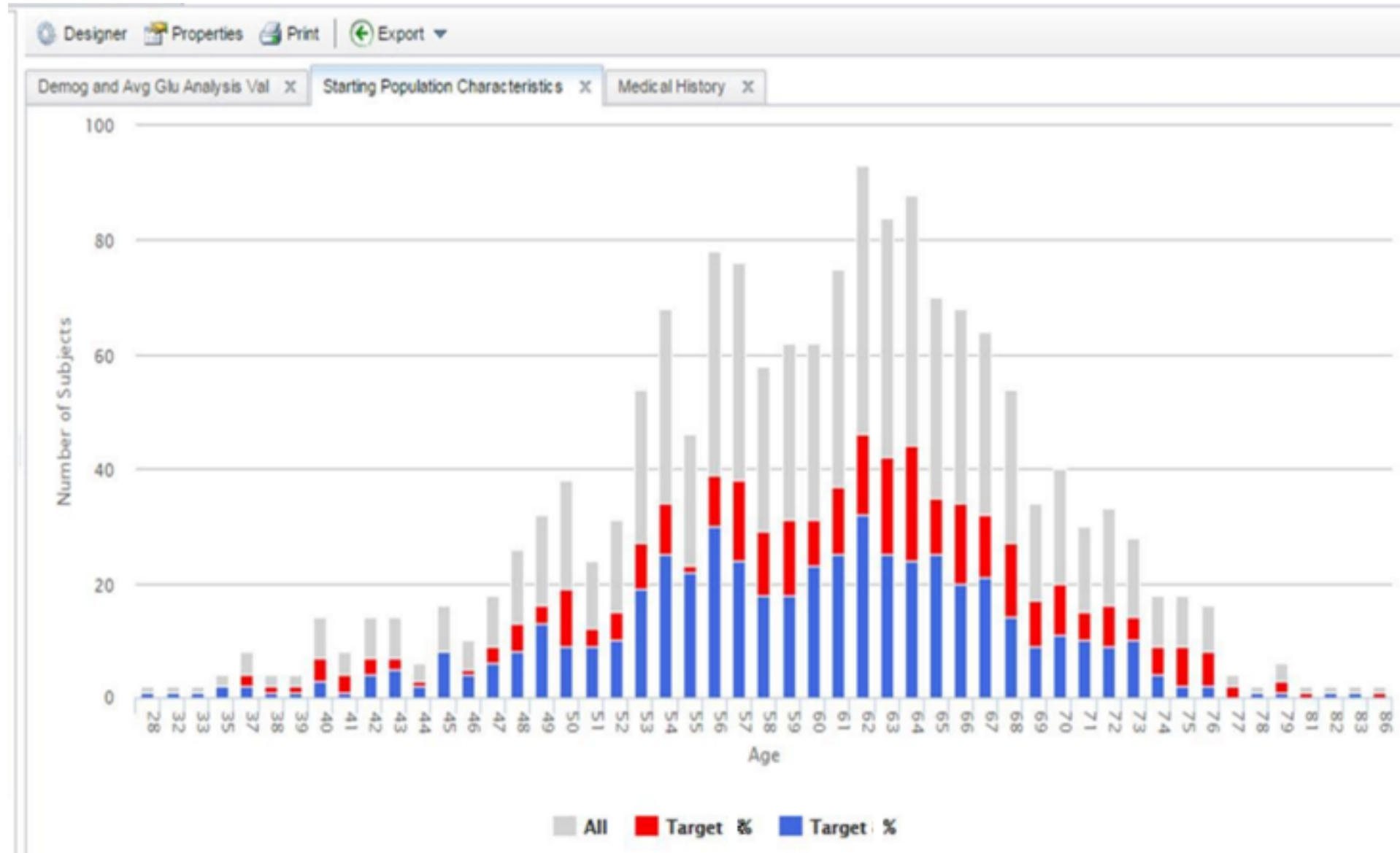


The screenshot displays the ANZO software interface, which is used for analyzing adverse events (AE) and their relationship with overall response. The interface is divided into several sections:

- Top Bar:** Contains the ANZO logo, a search bar, and various toolbars for navigation and analysis.
- Left Sidebar:**
 - Standard Toxicity Grade:** A list of toxicity grades with their respective counts (e.g., 5 (417), 4 (512), 3 (3603), 2 (10165), 1 (25813)).
 - AE Category of the Preferred Term:** A list of AE categories with their respective counts (e.g., SKIN ADVERSE EVENT (746), GASTROINTESTINAL ADVERSE EVENT (553), HEPATIC ADVERSE EVENT (248), ENDOCRINE ADVERSE EVENT (123), RENAL ADVERSE EVENT (105), HYPERSENSITIVITY/INFUSION REACTION (91), PULMONARY ADVERSE EVENT (67)).
 - Dictionary-Derived Term:** A list of dictionary-derived terms with their respective counts (e.g., ABDOMINAL ABSCESS (1), ABDOMINAL DISCOMFORT (34), ABDOMINAL DISTENSION (120), ABDOMINAL HERNIA (1), ABDOMINAL INFECTION (6), ABDOMINAL MASS (1), ABDOMINAL PAIN (518), ABDOMINAL PAIN LOWER (21), ABDOMINAL PAIN UPPER (207), ABDOMINAL RIGIDITY (1), ABDOMINAL SEP SIS (1)).
- Main Panel:**
 - Compound 1:** Shows two heatmaps:
 - AE Class and Best Overall Response:** A heatmap showing the relationship between AE classes (SD, PR, PD, NE, CR) and Best Overall Response (SD, PR, PD, NE, CR) for Compound 1. The color scale ranges from 0 to 200.
 - AE Term and Best Overall Response:** A heatmap showing the relationship between AE terms (e.g., SKIN ADVERSE EVENT, GASTROINTESTINAL ADVERSE EVENT, HEPATIC ADVERSE EVENT, HYPERSENSITIVITY/INFUSION REACTION, PULMONARY ADVERSE EVENT, RENAL ADVERSE EVENT, ENDOCRINE ADVERSE EVENT) and Best Overall Response (SD, PR, PD, NE, CR) for Compound 1. The color scale ranges from 0 to 200.
 - Compound 2:** Shows two heatmaps:
 - AE Class and Best Overall Response:** A heatmap showing the relationship between AE classes (SD, PR, PD, NE, CR) and Best Overall Response (SD, PR, PD, NE, CR) for Compound 2. The color scale ranges from 0 to 200.
 - AE Term and Best Overall Response:** A heatmap showing the relationship between AE terms (e.g., SKIN ADVERSE EVENT, GASTROINTESTINAL ADVERSE EVENT, HEPATIC ADVERSE EVENT, HYPERSENSITIVITY/INFUSION REACTION, PULMONARY ADVERSE EVENT, RENAL ADVERSE EVENT, ENDOCRINE ADVERSE EVENT) and Best Overall Response (SD, PR, PD, NE, CR) for Compound 2. The color scale ranges from 0 to 200.

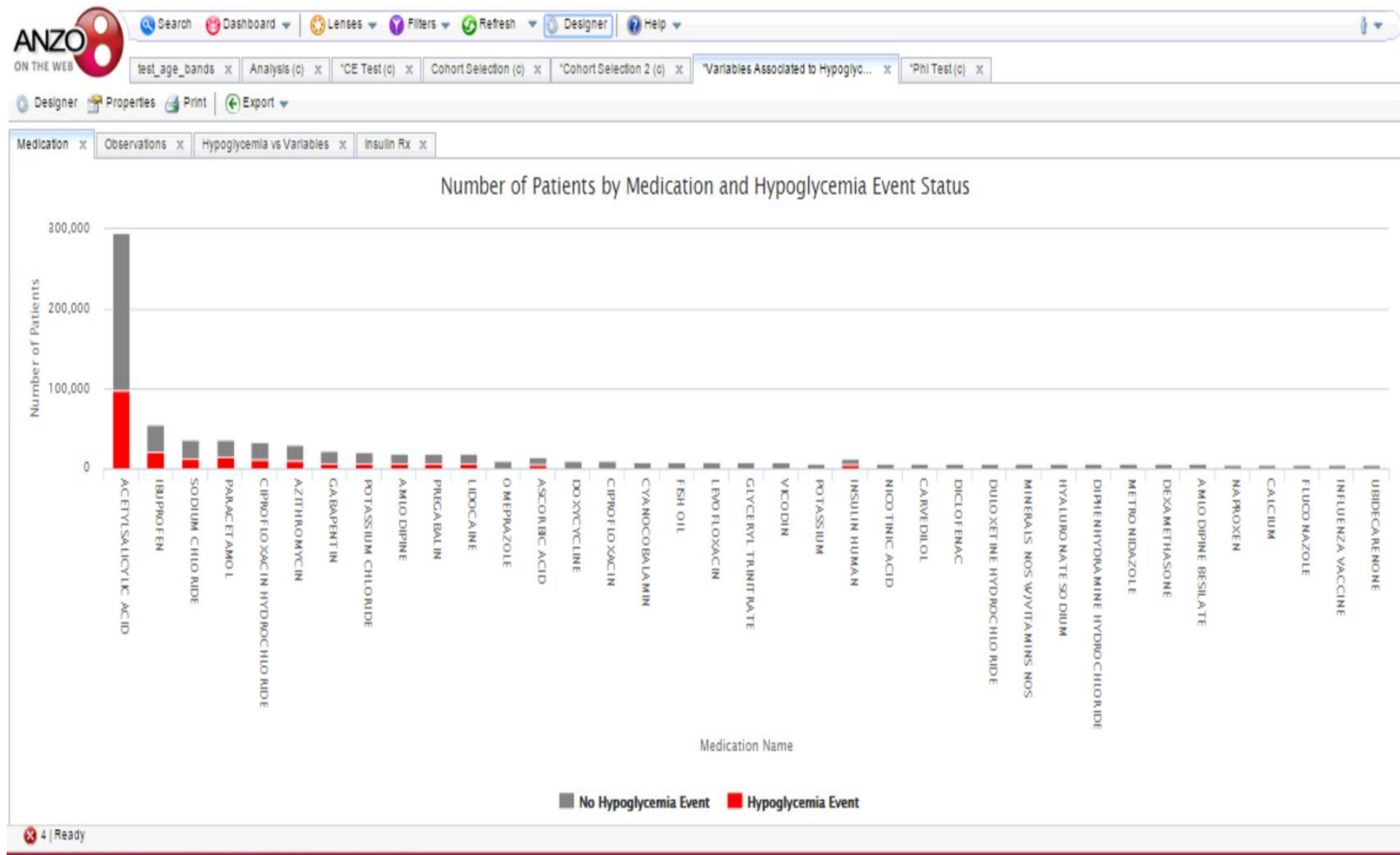
Clinical Smart Data Lake

Biomarker-driven Clinical Trial Analytics



Clinical Smart Data Lake

Real World Data (RWD)-driven Clinical Trial Analytics



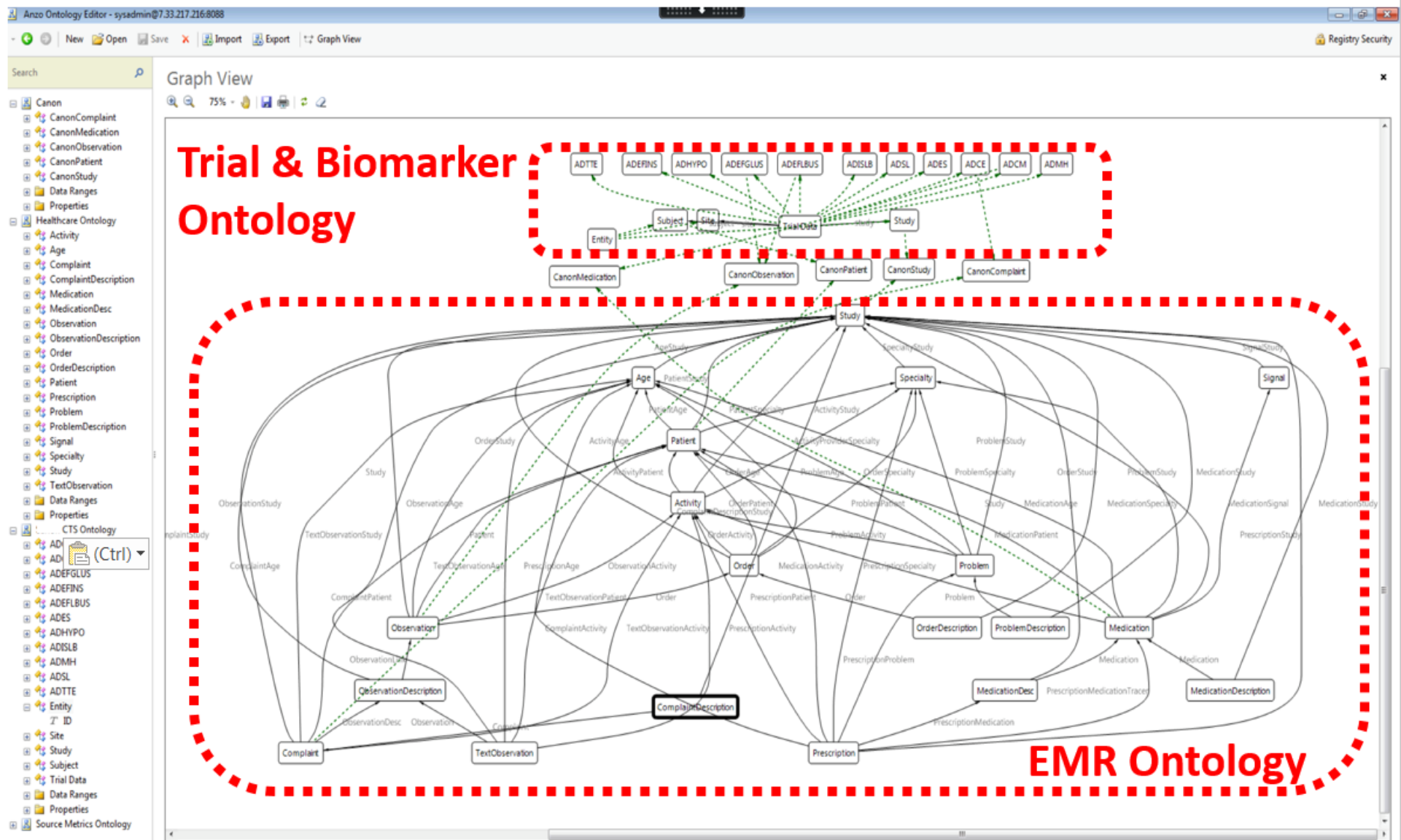
Clinical Smart Data Lake

Real World Data (RWD)-driven Clinical Trial Analytics

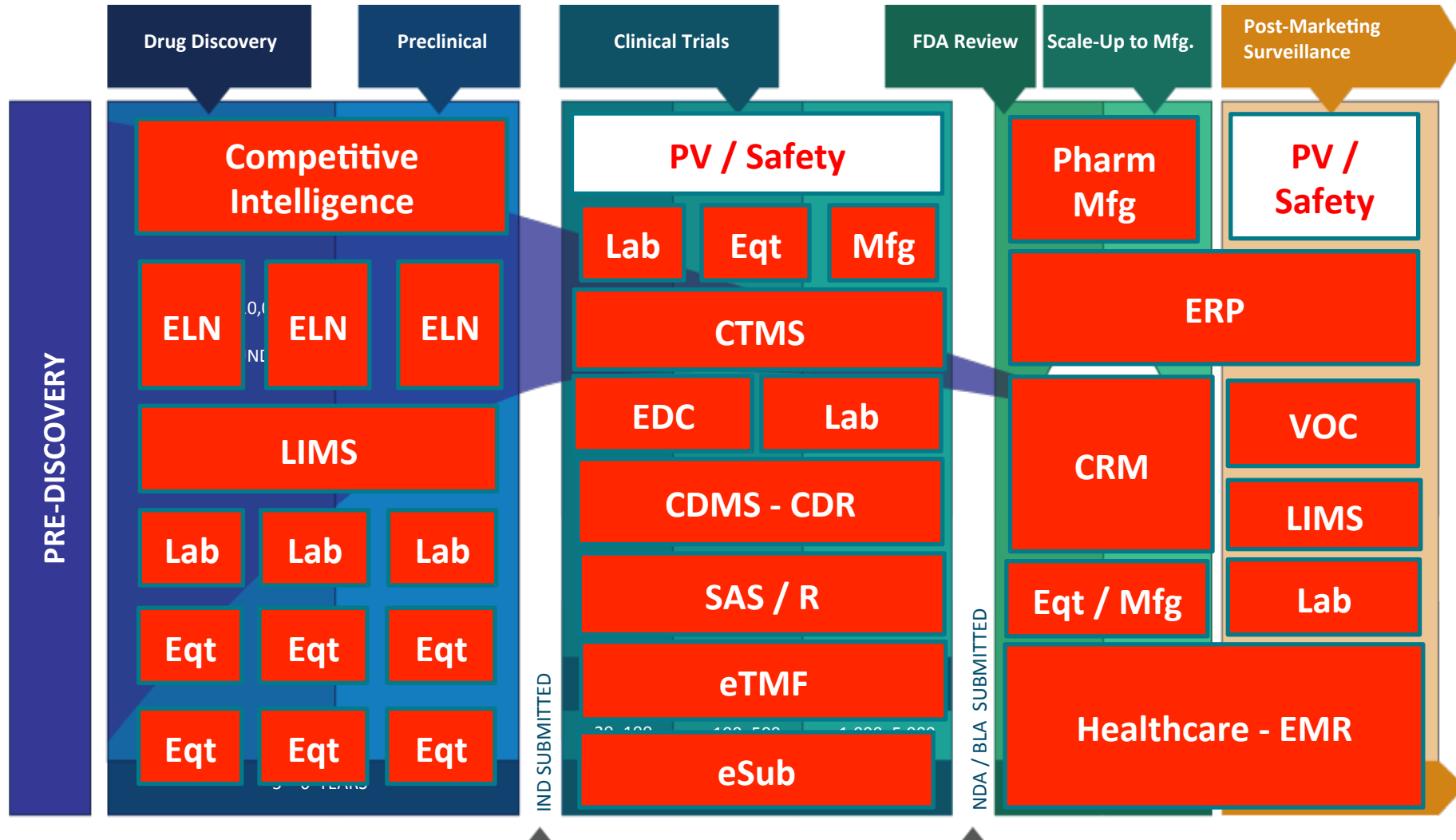


Clinical Smart Data Lake

Trial - Biomarker - EMR Data Integration



Clinical Smart Data Lake



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

A 45-year-old male patient, treated with tylenol + codine, experienced hypersensitivity to tylenol manifested by edema of uvula and dysphagia .

Patient's medical history included atopic allergic rhinoconjunctivitis to dust. Family history of atopic conjunctivitis in mother.

The patient is treated with tylenol for stiffness.

The patient was previously started on therapy with oral tylenol + codine. A few hours after the first intake of tylenol + codine, the patient experienced edema of uvula and dysphagia requiring medical intervention (nos) at emergency department. Corrective treatment was not specified.

The event resolved on an unspecified date.

Anzo Smart Data Lake© - Unstructured Machine Language Translation & Annotation

The screenshot displays the Anzo Smart Data Lake web application. The browser address bar shows the URL: `localhost/anzoweb/index.html?username=sysadmin&password=123`. The interface includes a navigation menu with options like Search, Dashboard, Lenses, Filters, Layout, Refresh, Toolbar, and Help. The main content area is titled "Chinese Document Consumption" and features a "Table of Documents" with columns for Document, Source, Author, Last Modified, Summary, and Intake Time. The table lists several documents, including one about a landslide in Maizhokunggar County, Tibet. Below the table, there is an "Annotated Document" section showing a Chinese news article. The article text is highlighted in yellow, and a sidebar on the right lists various annotations such as [Rosette: Date] (2), [Rosette: Location] (6), [Rosette: Organization] (13), [Rosette: Person] (10), and [Rosette: Title] (2). A callout box at the bottom provides an English translation of the highlighted text.

Document Table:

Document	Source	Author	Last Modified	Summary	Intake Time	
习近平李克强对西藏墨竹工卡县山体滑坡作出重要指示	0 original		3/29/2013 4:33 PM	—	4/1/2013 11:08 AM	Delete
习近平李克强对西藏墨竹工卡县山体滑坡作出重要指示	1 rosette		3/29/2013 4:33 PM	—	4/1/2013 11:04 AM	Delete
习近平李克强对西藏墨竹工卡县山体滑坡作出重要指示	2 translation		3/29/2013 4:33 PM	—	4/1/2013 11:06 AM	Delete
Xi Jinping, Li Keqiang made important instructions Maizhokunggar County, Tibet landslides	3 translated annotations		3/29/2013 4:33 PM	—	4/1/2013 11:56 AM	Delete
Xi Jinping, Li Keqiang made important instructions Maizhokunggar County, Tibet landslides	4 full translation		3/29/2013 4:33 PM	Xi Jinping, Li Keqiang made important instructions Maizhokunggar County, Tibet landslides Requirement launched by the re...	4/1/2013 11:14 AM	Delete

Annotated Document:

习近平李克强对西藏墨竹工卡县山体滑坡作出重要指示

要求全力展开救援 尽最大努力抢救被困人员

新华社北京3月29日电 3月29日晨，西藏自治区拉萨市墨竹工卡县境内，中国黄金集团公司下属的华泰龙矿业开发有限公司甲玛矿区发生大面积山体滑坡，据初步统计，83名工人被埋。党中央、国务院对此高度重视，正在国外访问的中共中央总书记、国家主席、中央军委主席习近平，分别作出重要指示，要求全力展开救援，尽最大努力抢救被困人员，防止次生灾害。李克强、郭声琨、王勇等中央领导同志也作出批示。

自前，西藏自治区党委、政府和有关企业正在全力组织抢险救援，安全转移伤员，国家安监总局、国土资源部、国家质检总局等部门已派工作组赴现场指导救援工作。

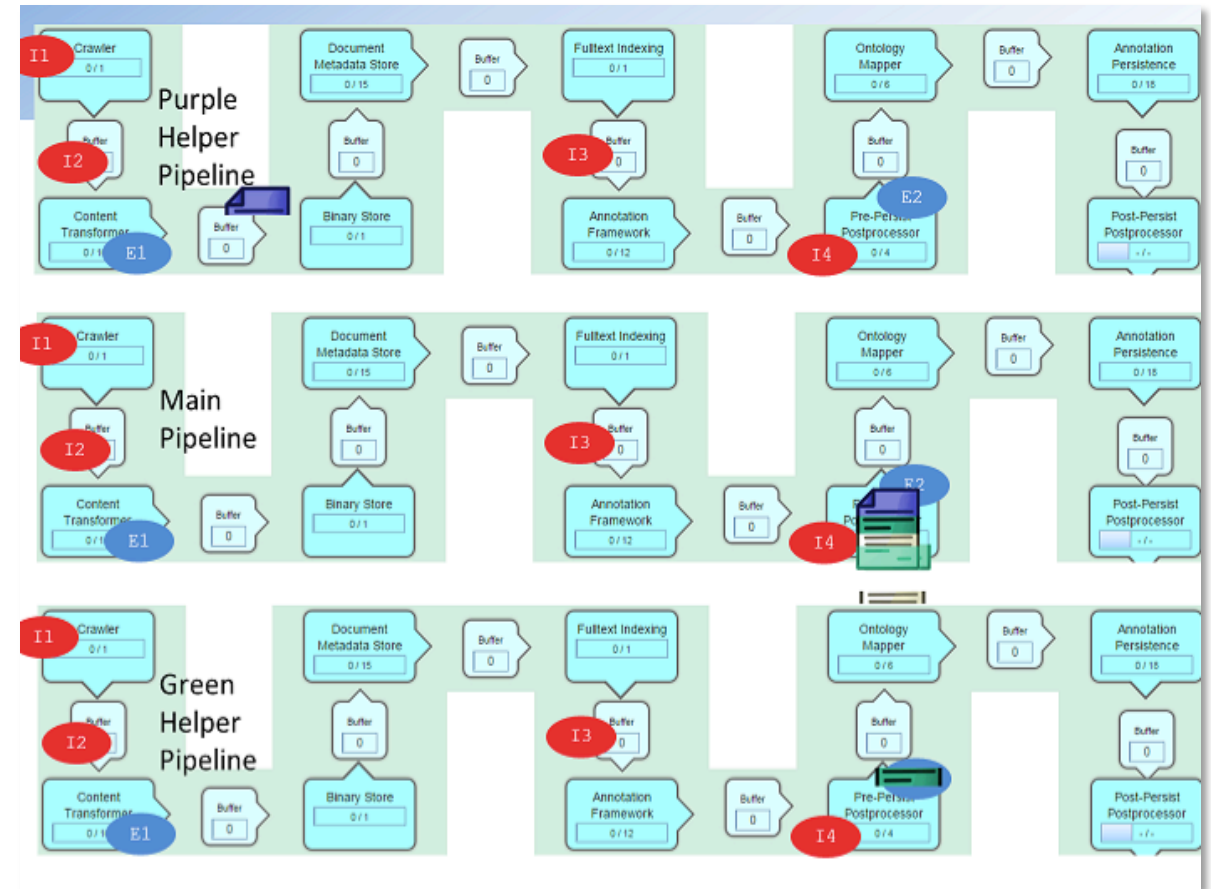
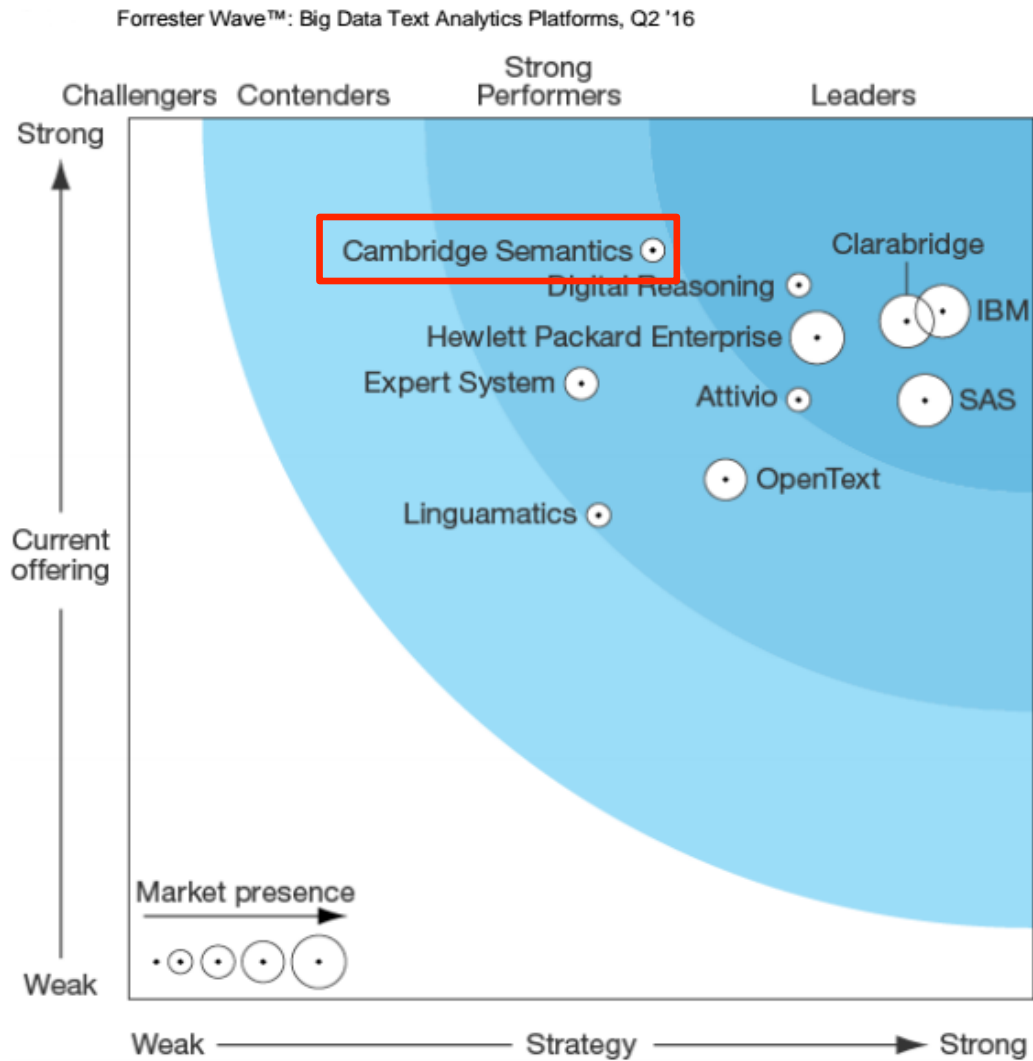
Callout Box:

Xinhua Beijing March 29, the morning of March 29, Maizhokunggar County, Lhasa, the subsidiary of China National Gold Group Corporation Huatalong Mining Development Co., Ltd. Jiama Mine is a large area of landslides, according to preliminary [Annie: Organization] China National Gold Group Corporation Central Committee and the State Council attach great importance to the CPC Central Committee General S Corporation Chairman Xi Jinping, member of the Standing Committee of the Political Bureau of the CPC Central Committee, State Council Premier Li Keqiang visit abroad, made important instructions, called for efforts to start the rescue, do its utmost to rescue trapped, to prevent secondary disasters, and pay close attention to the verification of trapped number identify landslide reason, to carry out all remedial work. Liu Yunshan, Zhang Gaoli, Liu Qibao, Guo Shengkun, Wang Yong and other central leaders also given instructions.

Anzo Smart Data Lake© - Unstructured

Industry-leading NLP Technology

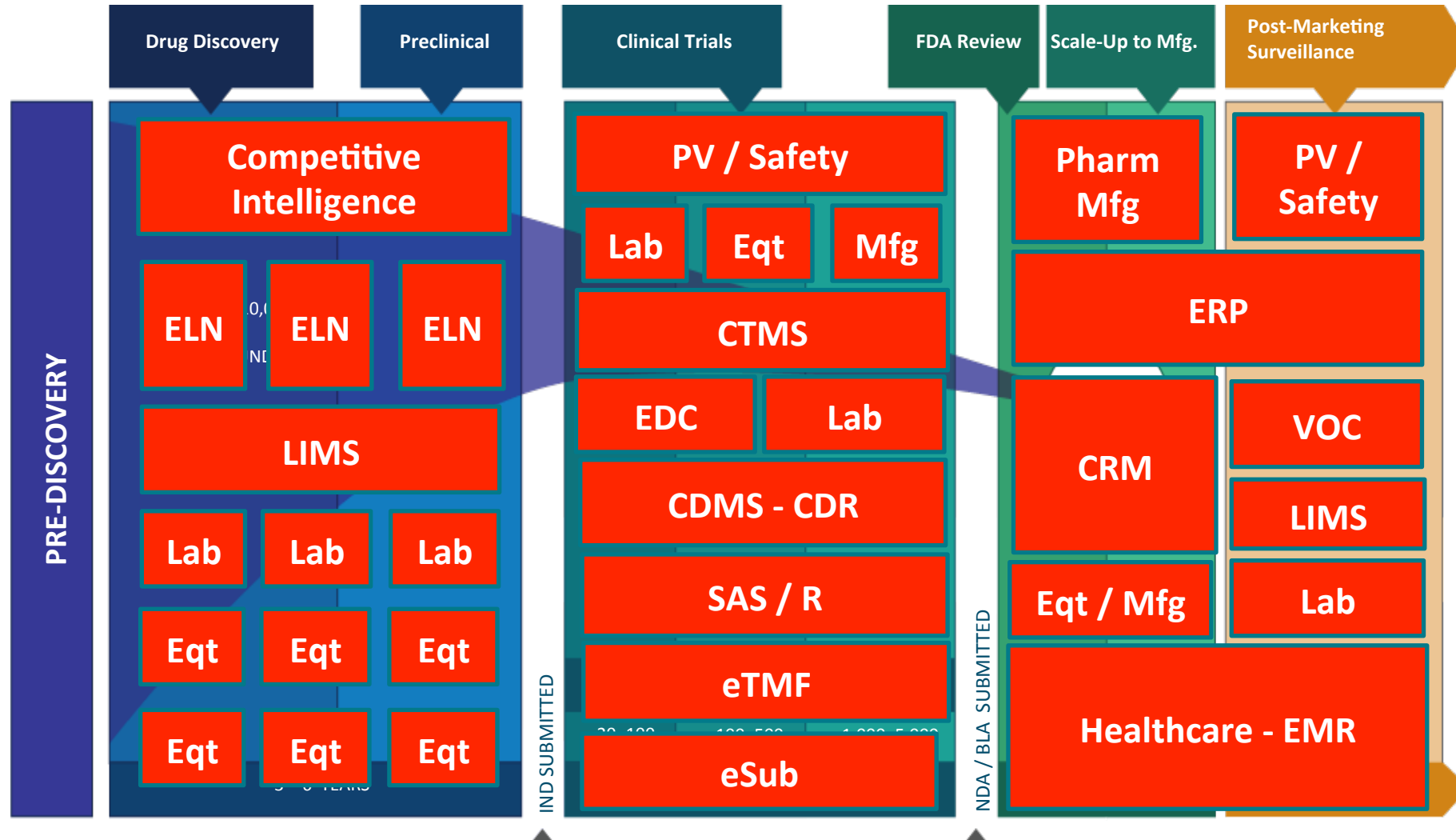
Forrester Wave, 2016



Pipeline Orchestration: multi-threaded & multi-annotated (40+ annotators)

Smart Data Lake – reducing the time-to-insights

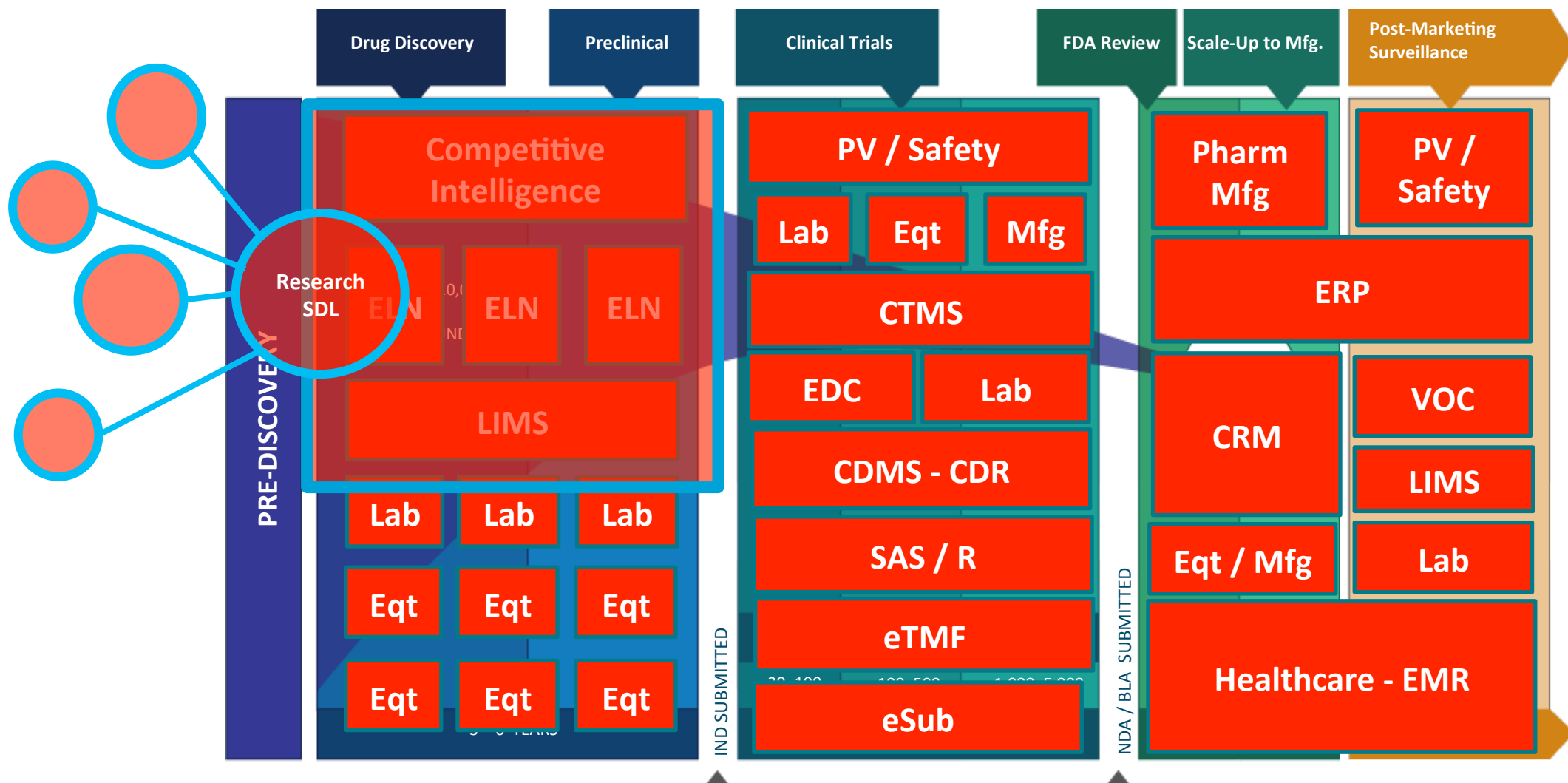
Anzo Smart Data Lake© Enterprise Solutions



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

Smart Data Lake – reducing the time-to-insights

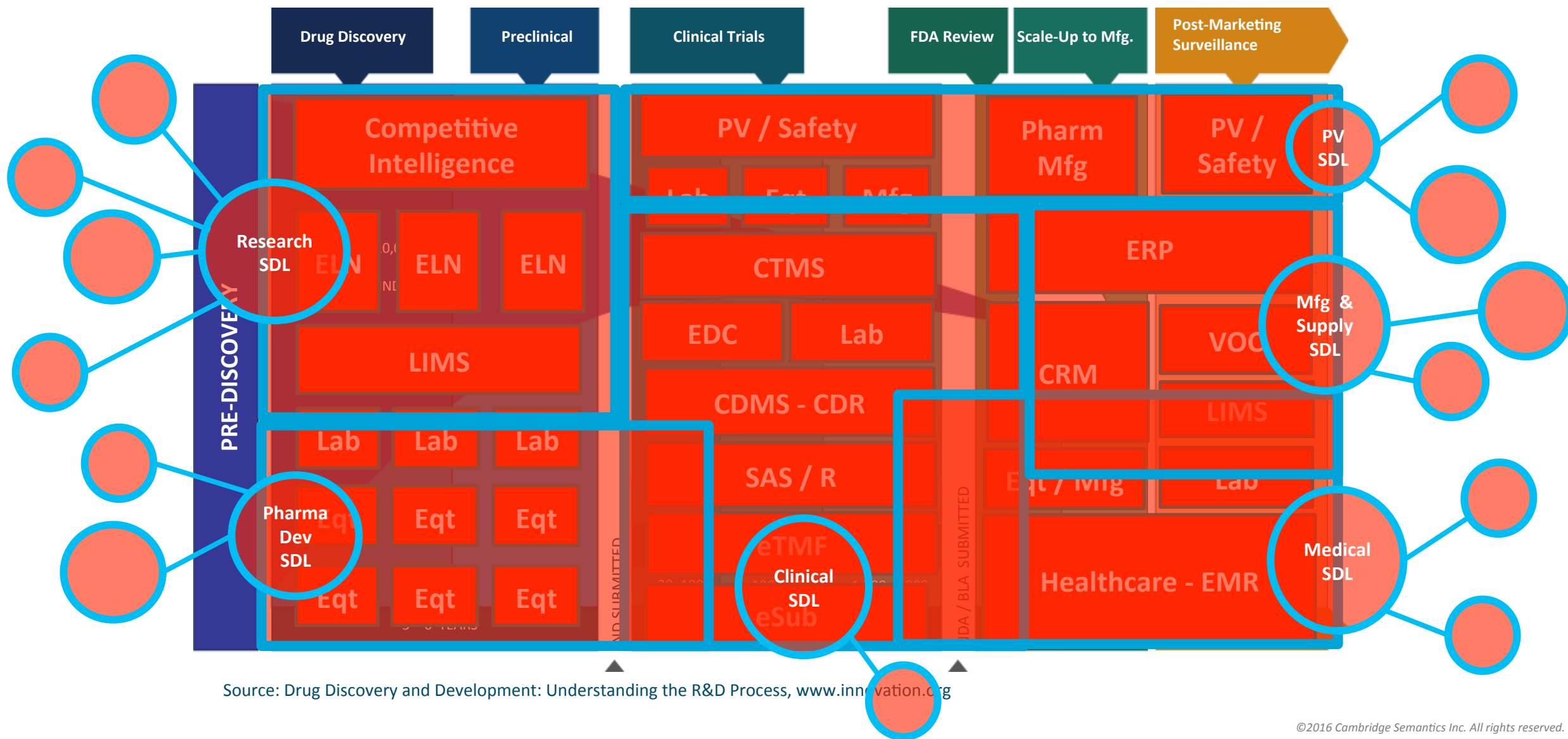
Anzo Smart Data Lake© Enterprise Solutions



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

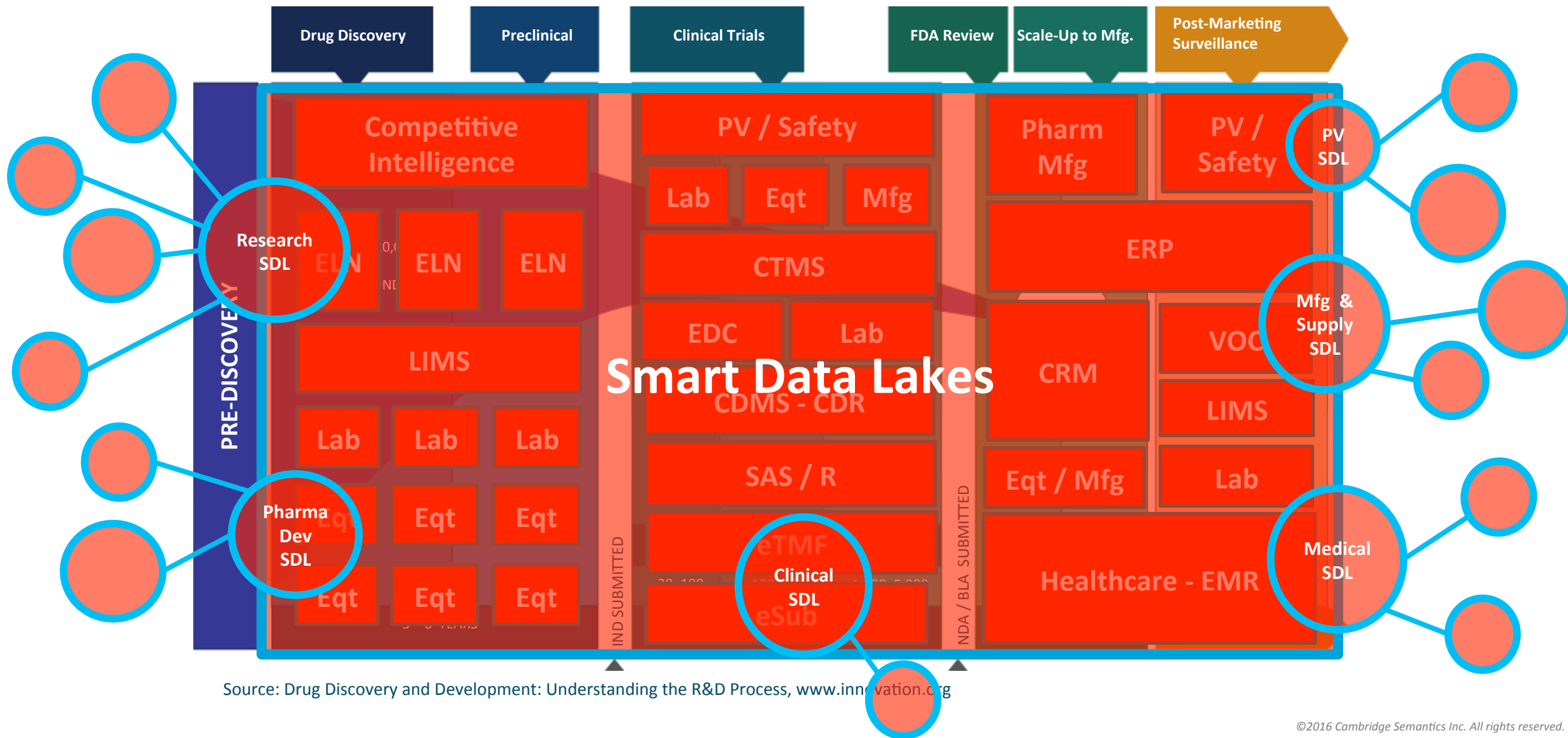
Smart Data Lake – reducing the time-to-insights

Anzo Smart Data Lake© Enterprise Solutions



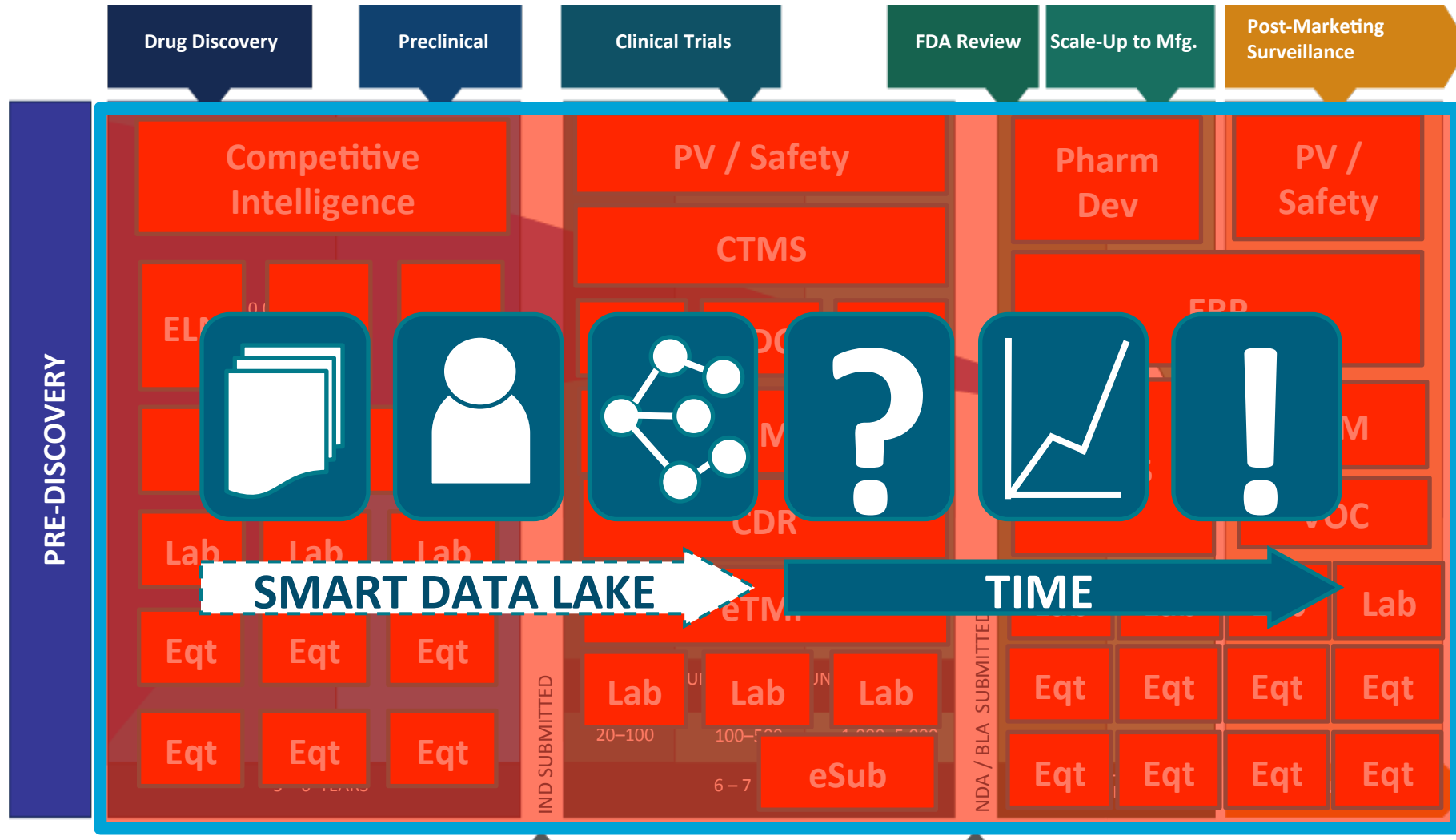
Smart Data Lake – reducing the time-to-insights

Anzo Smart Data Lake© Enterprise Solutions



Smart Data Lake – reducing the time-to-insights

Anzo Smart Data Lake© Enterprise Solutions



Source: Drug Discovery and Development: Understanding the R&D Process, www.innovation.org

Any Questions?