

## Big data and AI in Pharma: comprehensive Insights come from one comprehensive Platform

Making the case for the integration of all data

*The digital revolution is changing healthcare practice at a rapid pace.*

The digital revolution is changing healthcare practice at a rapid pace and with it the demands to all the players in the field. All major pharma companies have established ambitious programs to leverage the new technology to

enhance the efficacy and efficiency of all business processes in the quest for better patient outcomes and enhanced financial results. Pharma investments in the big data/Artificial Intelligence (AI) field are estimated to have tripled since 2014 to an estimated \$10 billion per year in 2017. And there is a broad variation of practical use cases:

- With the understanding of patho-mechanisms on genetic and cellular receptor levels, research is transcending classical medical categories. Molecules impacting i.e. vessel growth in tumors also help cure eye diseases.
- The treatment of metabolic disorders like i.e. diabetes become a well – orchestrated, data driven cooperation of wearables and apps, medical, nutritional and behavioral specialists that monitor key parameters in real time to allow individualized therapies with significantly enhanced outcomes.
- The new field of translational medicine is generating data to provide information and evidence across classical medical, biological, behavioral and epidemiological disciplines to generate new insights into pathological patterns and potential new cures.
- New targeted therapies generated from pharmaceutical and clinical research are available with specific companion diagnostics and change treatment paradigms and in some cases turn fatal illnesses into chronic diseases, or even provide cure.

Data were always the driver of research, but what has changed is the exponential rate in which the volume of data and the number of new relevant scientific disciplines increases.

*Data are not the new gold – but the ability to put them together in a relevant and analyzable way is.*

“Data are not the new gold – but the ability to put them together in a relevant and analyzable way is” is a regularly heard statement of high-ranking pharma data specialists. What’s most needed is a comprehensive look at the data available for decision making in research and in business contexts: from scientific communication and medical information to commercial functions.

Innoplexus has accompanied pharmaceutical companies in the establishment of big data and AI solutions. The key data issues that we were faced with were

1. We know that we should have the data - but in which silo?
2. We search for relevance – but we receive information. Is there a way to get to relevance?
3. If we get relevance - how can we turn it into insights?

## The 3 steps to put comprehensive, relevant and analyzable data at your fingertips

### 1. Put it all in one place

The well-known issue of data siloes has been discussed for 20 years and addressed in the pharma industry. In the area of production and supply chain management it has often been tackled by change management initiatives that led to comprehensive business processes: these new ways to do business have always been supported and enabled by one comprehensive data platform. In these enterprise platforms, all relevant data are stored and can be extracted at all times by all users across all functions and departments. The resulting efficacy gains are significant: reports and insights can now be derived at all times and the coordinated business process across departments and functions have led to new levels of productivity and transparency.

There is no evidence against the assumption that this analogy – applied to the scientific data used in the business contexts from R&D to scientific communication, from medical information to marketing communication - would not yield similar productivity gains.

*Put all internal and external data in one place and use a smart technology to store them such that relevance can be extracted.*

The simple concept: put all internal and external data in one place, store them such that all formats can be recognized and use a smart technology to store them such that relevance can be extracted as needed by the respective user.

Such a system integrates published (publications, clinical trials, congress articles etc.), unpublished (e.g. from universities and research institutes), restricted, third party (e.g. CRM and IMS data) and enterprise data in real time. It collects such data through various acquisition techniques (API connectors, crawlers, etc). Fetched data are extracted and cleansed using various AI and data processing methods. Data are then indexed using a comprehensive ontology of Life Science to build the information graph and various analytics in such a way that a user can get a 360-degree view of any relevant Life Science concept.

Through this process Innoplexus' *Ontosight*<sup>TM</sup> platform delivers:

1. Always up to date information to the user in real time allowing insights based on the most recent data.
2. Completely secured enterprise data that always remain inside the firewall of the company (even unpublished data are never sent to the intermediary and the meta data remain encrypted across the network.)
3. Complete, easily searchable, sharable and accessible data sets for further analytics in real time.

What you get is one comprehensive depository for all internal and external information, “freed of its untraceable data silo, made available to generate insights”:

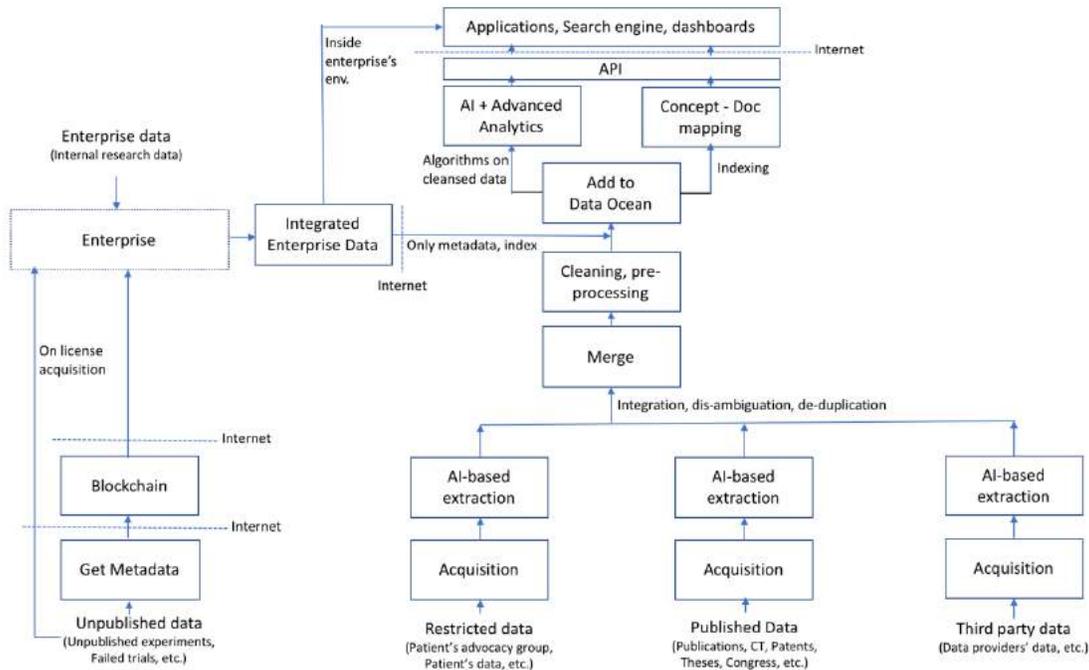


Figure 1: Depiction of Ontosight's integration process of various data types

## 2. Organize your data “contextually” – so that database searches retrieve relevance – not data

Librarians have – amongst other qualities – the ability to search information not only according to search terms like title or author, but also according to the context of any given search. They are effective because they understand the scientific language of biomedicine and healthcare. They are valuable because they understand the context of their search and where to find the information

*Equip your systems with an ontology that understands the language of healthcare and medical science and sorts content according to context.*

But how to store the data so that relevant information can be found? The simple answer to a complex process is: teach the search engine the language of healthcare and medical sciences. Translated into IT terms: equip your systems with an ontology that understands the language of healthcare and medical science and sorts content according to context. This way the machine does not look for similar

words but for similar biomedical and healthcare concepts. What the machine extracts are not similar sequences of letters but adjacent concepts of content.

Ontosight™ supports all the Life Science entities including drugs, diseases, indications, targets (proteins), pathways, genes, biomarkers, etc.

Once all the Life Science data are assembled in one comprehensive database, *Ontosight™* performs ontology-based searches to extract relevant information. *Ontosight™* encompasses 25 million semantic concepts, representing a unique biomedical language algorithm. The analogy often used to describe the difference between keyword-based search and an ontology driven approach is the difference between fly-fishing and fishing with a fishnet: in an ontology driven search, the “Data Ocean” of relevant data is skimmed with a custom-tailored fishing net, where the nodes are the concepts and the strings are the semantic associations.

**What you get** is an ontology-based data storage with an extraction machine that understands the context of your business needs and picks out *relevant* information:

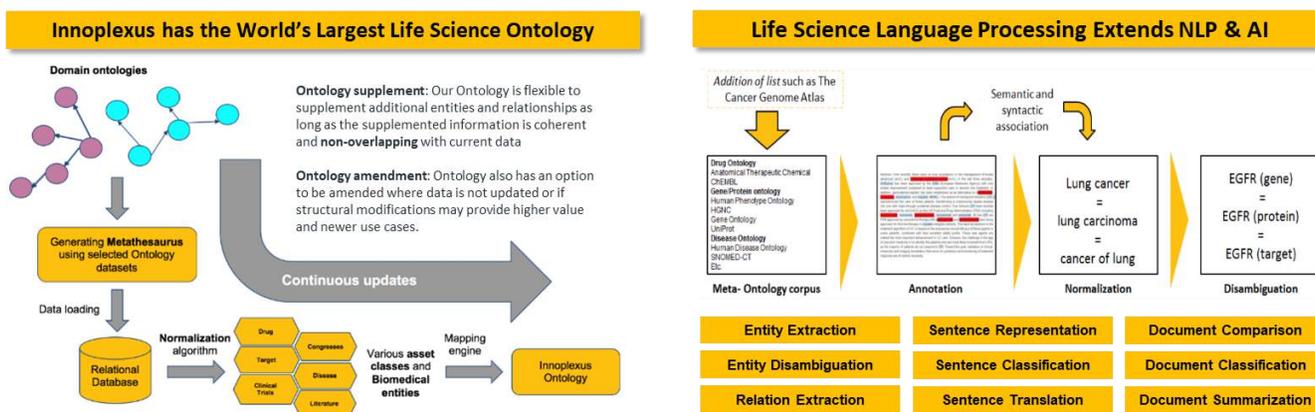


Figure 2: Processes of a self-learning Life Science Ontology

### 3. Visualize relevance so that it becomes analyzable to generate insights

Anyone who has ordered a book online knows that recommendations like “if you like this – you may also like...” next to the chosen title come automatically. The algorithms suggest titles according to the frequency of similar orders. What works for novels is not specific enough in the scientific context: no researcher would be happy with a recommendation “if your read this edition of *Jama*, you may also like that edition of the *New England Journal*.”

But imagine the scientific literature search could be chosen according to different contexts; i.e. mechanisms of action, or similar poster topics, or congress presentations with the same therapeutic concept ... and a pie chart would appear in which you could see the distribution of the mechanism of action of your chosen MoA across major congresses and when you click on it, the actual publications appear.

Or, if in the context of a scientific communication initiative you would try to understand the way different molecules are published across different channels and a machine would present to you the trends over time in the various scientific publication channels, and when you click on any data point, the original paper would become readable no matter in which format it was stored. Then the machine would in milliseconds provide you

with *relevant* information. The time you would have spent in the past counting publications and channels - you would now invest into generating hypotheses for the reasons for a certain pattern and you could at any time deep-dive into the data to further explore.

*Data do not provide information,  
but data in context do.*

How to turn information into insight? Data do not provide information, but data in context do. *Ontosight™* allows you to generate contextualized search results. The question remains: if that information is available, how do I get to

insights and actionable results for your decisions? The machine can visualize information so that patterns become recognizable and that is the starting point for the generation of new insights.

This has only recently become possible: the technological advances that are often referred to as the big data/AI revolution enable to shorten to milliseconds the process of data gathering and cleaning, which has before taken many man hours. This big data/AI revolution frees time to generate insights through permanently updated information in time frames never seen.

*Ontosight™* extracts (structured and unstructured) information through various Artificial Intelligence based techniques like Natural Language Processing and Computer Vision. The data received through this process are then cleansed and merged with already processed data.

The resulting “Data Ocean” hosts millions of documents (publications, clinical trials, patents, etc). To enable real time search through such a vast amount of data, each document is indexed through an *Ontosight™* specific process for ease of search. This allows searches to quickly identify the associated document through an ElasticSearch process and as a result provides comprehensive information: i.e. the relevant abstracts as well as the link to open the original document.

Just like in the analogy of Google Maps, insights become possible by overseeing the full picture of a given research to then “zoom in” on patterns and details of the information provided.

*Ontosight™* results come in form of dashboards and intuitive visualizations. Applications like i.e. sentiment analytics, guideline analysis, competitive landscaping, comparative studies, triangulate the information and visualize data in intuitive and comprehensive “information maps”. These visualizations can be customized according to the use case. By reducing the complexity of information retrieved from the “Data Ocean” to an easy to comprehend overview, insight generation can move from the “big picture” to the detail of the provided information in real time.

**What you get** are intuitive visualizations of all relevant information in real time, providing clear and structured overviews that allow insight generation

## The path to comprehensive insights starts with one comprehensive data platform

Innovative technologies can help pharma companies to overcome data challenges. Relevant data questions are solvable if you choose the right technology support:

### 1. How to overcome the data silo issue to make all data available for analysis?

*Create one comprehensive platform with all internal and external data.*

### 2. How do I extract relevant information from a comprehensive data base?

*Use an ontology-based search engine that “understands” the language of healthcare and all biomedical and scientific terms and contexts*

### 3. How to generate insights from information?

*Structure data to gain insights from purpose built-automated and customized visualizations for pattern recognition.*