



Sustainable and FAIR Data Ecosystem, supporting
new insights in Life Sciences

iRODS UGM 2022

Carl Latham

Lighthouse
customers



ALEXION

REGENERON

AMGEN

medidata



AstraZeneca



Bristol-Myers Squibb

iRODS

ONTOFORCE

Company snapshot

- Founded in 2011
- 56 people – VC Backed
- EY Scale-up of the year 2018
Deloitte Technology “Fast 50”
- Based in EMEA (Ghent, Belgium) and US (Boston)
- Patented software & data ingestion
- Exclusive focus on the biopharmaceutical industry



Which clinical studies have CT lung data and EGFR expression data?

I would like to get sequencing data for all patients onboarded on oncology trials with cardiovascular adverse events.



Which are all the trials that have been tested fasena in phase II?

Do we have data for oncology trials with adults patients reporting cardiovascular adverse events?

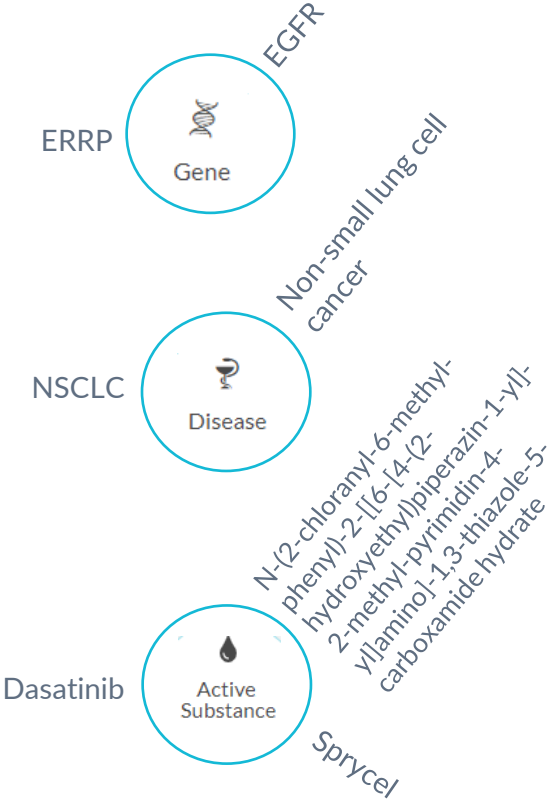
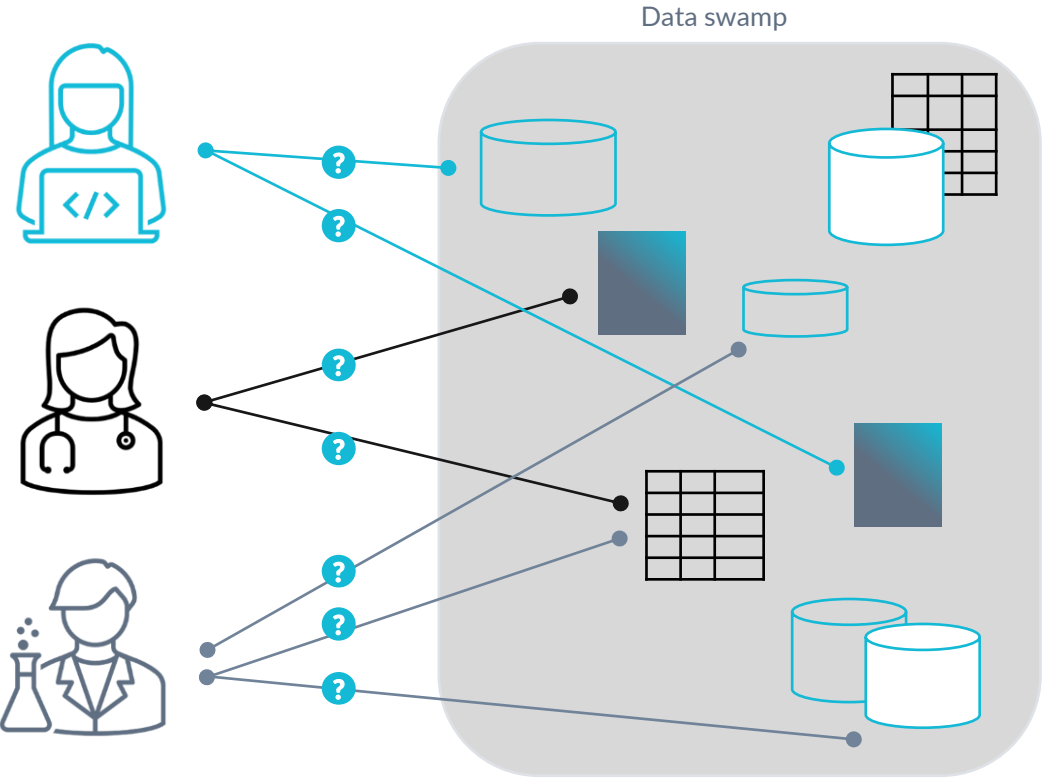
Which clinical sites were involved in studies recruiting patient population with JAK2 V617F mutation and for myeloproliferative diseases?

Life Science data is facing an **ever-growing data volume**, with more complexity and endless difficulties to explore.



As access data platforms become more democratized, the need to **collaborate** seamlessly increases significantly.

Data Is Siloed And Not Harmonized



FAIR Data Principles

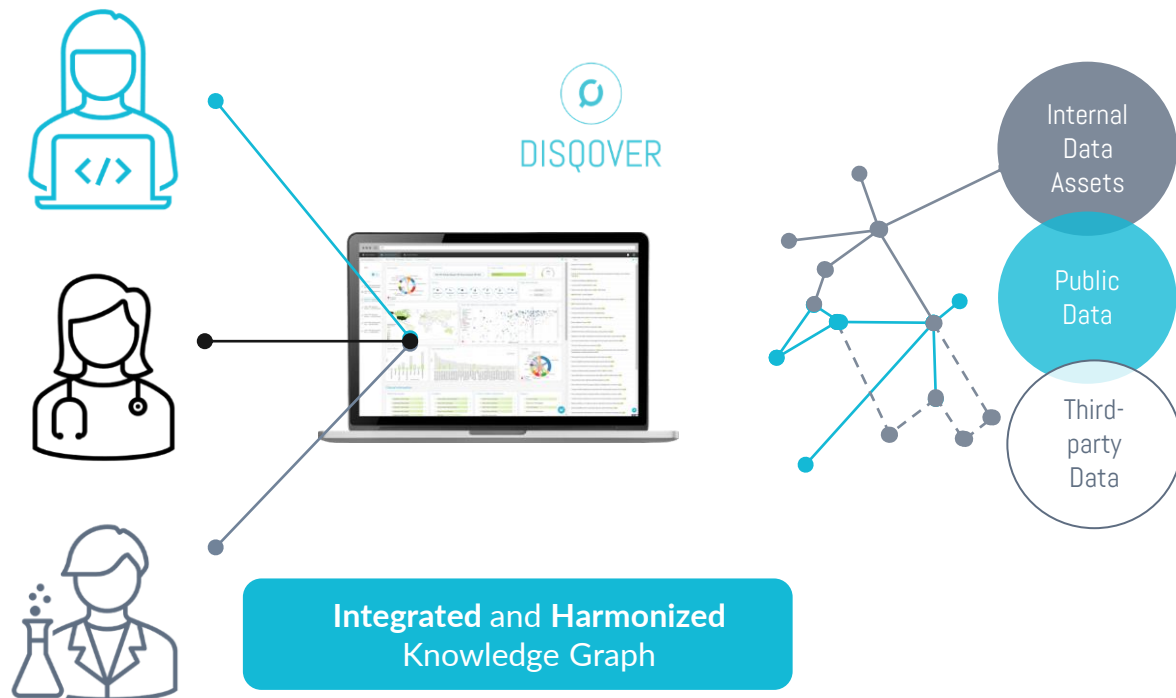
FROM SILOED DATA TO ACTIONABLE DATA USING FAIR DATA PRINCIPLES

ChEMBL
PubMed
ORCID
ICD-10 CM
SNOMED CT
MedDRA
EudraCT
Orphanet
DrugCentral
GRID
Gene Ontology
Clinicaltrials.gov

140+
public
sources



Insights for Life Sciences

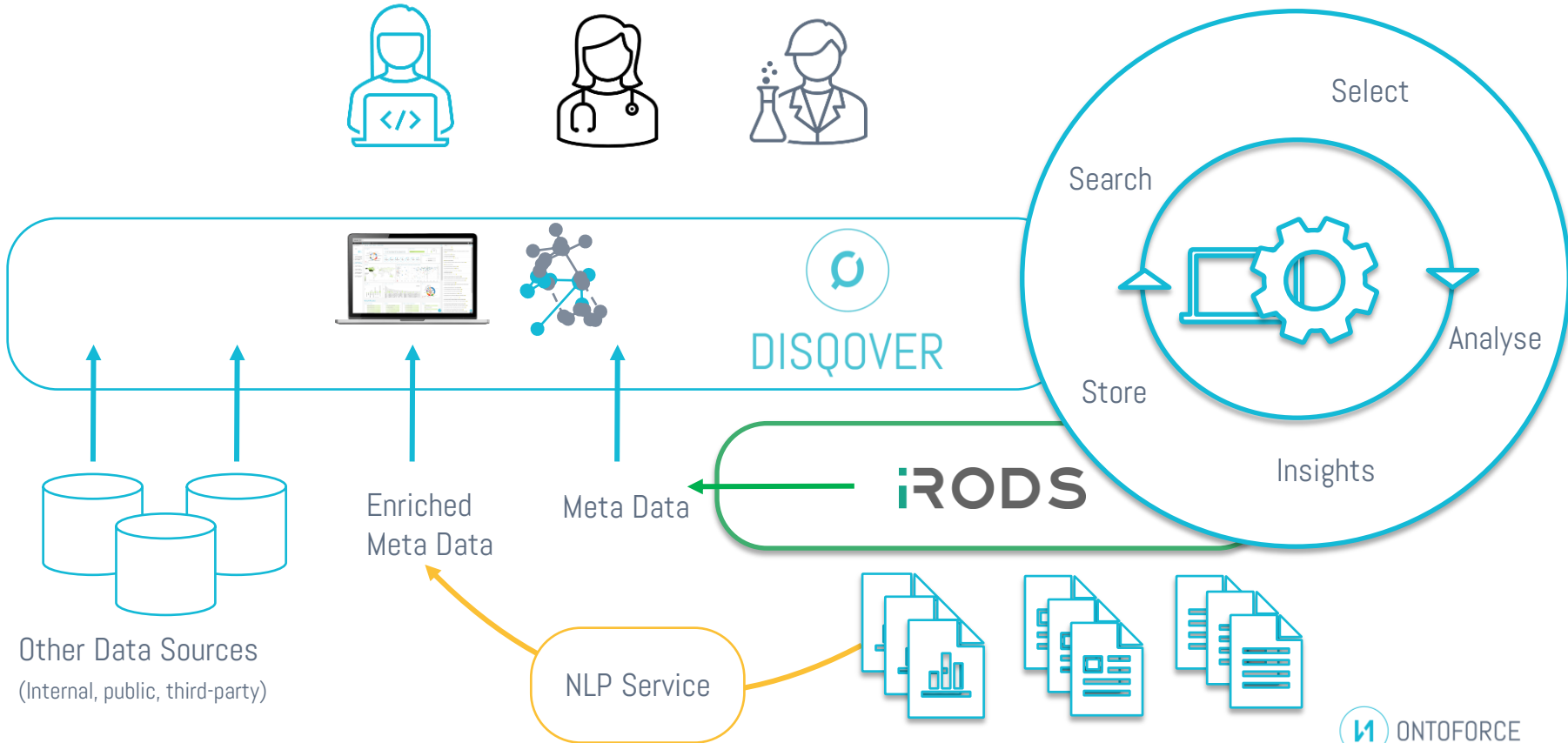


iRODS

Data Management

Consolidated Data Landscape

Sustainable Data Ecosystem



The platform: DISCOVER

THE LIFE SCIENCE KNOWLEDGE PLATFORM THAT LINKS ANY TYPE OF DATA TO DELIVER ACTIONABLE INSIGHTS

INTUITIVE USER INTERFACE

Boost data value extraction and encourage exploration with an engaging interface that caters to the needs of both general and expert users.

INTERACTIVE SEARCH

Avoid information overload and shield users from underlying complexity with interactive data journeys and comprehensive visual dashboards.

POWERFUL CUSTOMIZATION

Adapt views and navigation to specific use cases with powerful self-service customization.



INNOVATIVE DATA INTEGRATION

Merge and link siloed data sources, including third-party and public sources, while retaining traceability, and gain insights based on comprehensive source material.

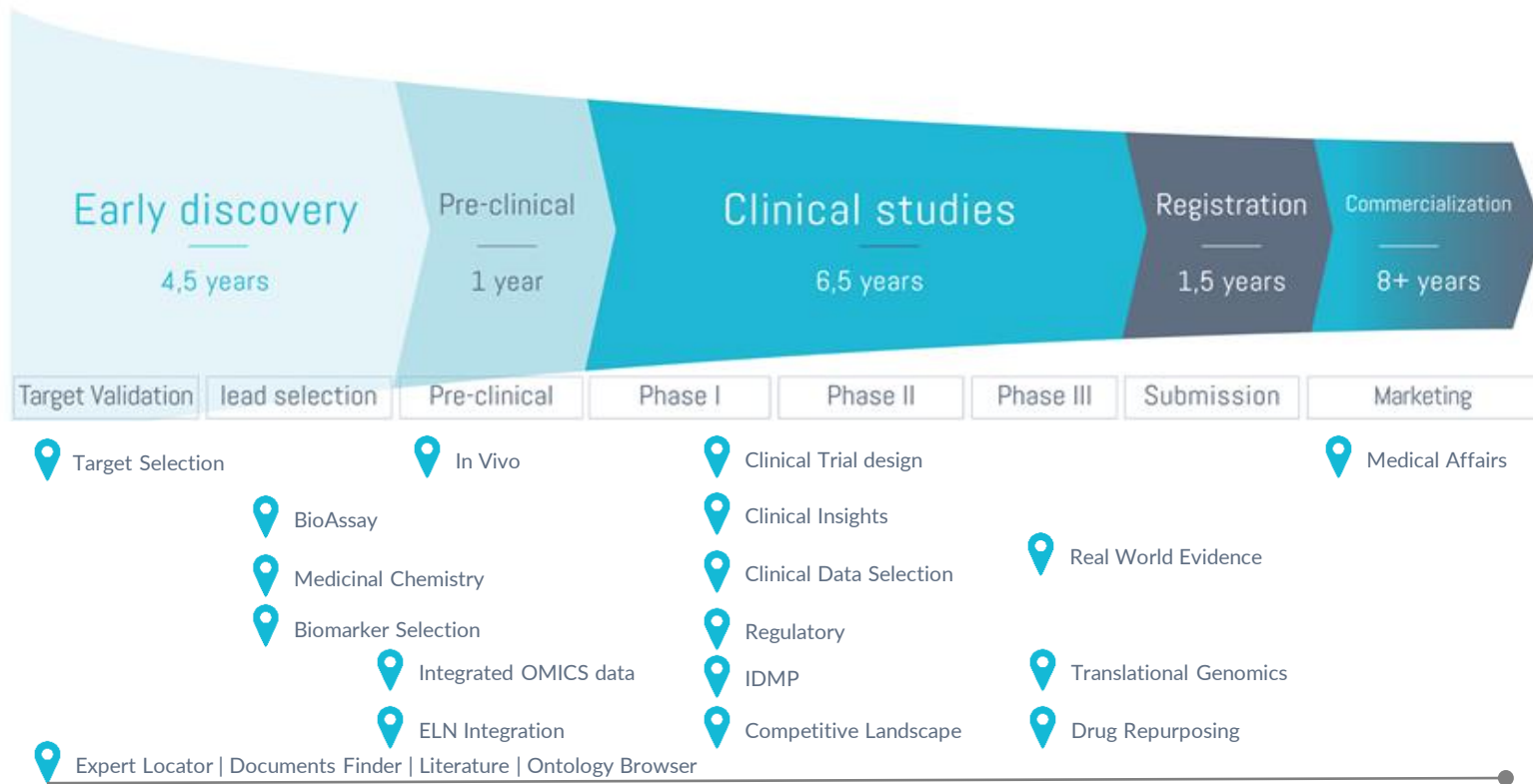
EXPANDABLE PLUGIN ARCHITECTURE

Extend DISCOVER's core functionalities with custom and third-party applications and embed the platform into your enterprise IT ecosystem.

DATA INGESTION PIPELINES

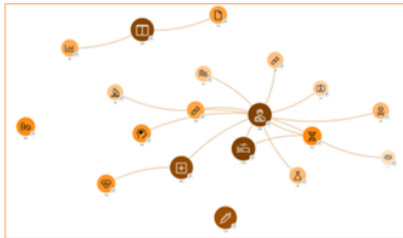
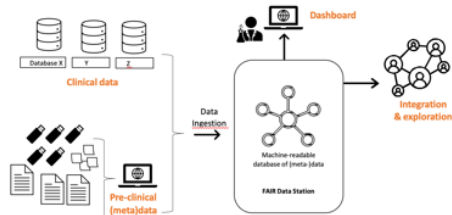
Import, transform and integrate data from different sources using an innovative visual pipeline – no extensive coding required

Supporting Data Access along Drug Development



FAIR Data Station

Integration & Exploration



1 By transforming data as such that it fits to a machine-readable data model, relations between various data types can be automatically inferred and used during data exploration.

Patient
 Filters: CSR: Diagnosis date from 2018-11-05 Sex Female

13 results

Amount of ... 13 / 996

Sex: Female

Data source:

- 13 Central Subject Repository
- 13 SKION
- 4 Organoids
- 4 Toxicity
- 3 BRAINcare
- 3 High-throughput Screening
- 3 The PALGA reports dataset
- 3 cvad

RNA sequencing ava...: Yes/No pie chart (Data missing: 976)

Date of Diagnosis: 2018-11-05 to 2020-10-12

Anatomic location: Pie chart showing various anatomical sites like Brain, N, Ventricle, NOS, L Brain stem, etc.

Links to other categories:

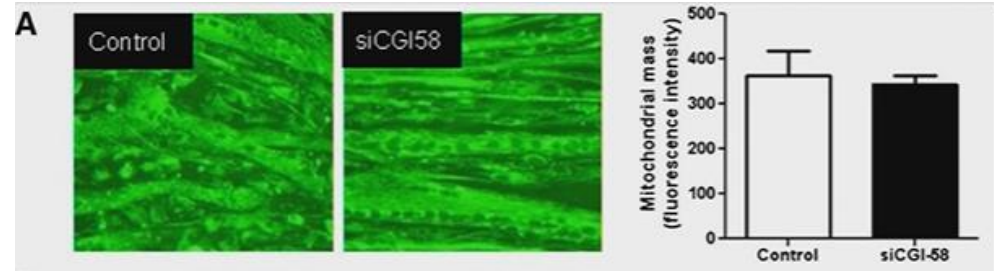
- Biomaterial: 69
- Biosource: 34
- Diagnosis: 26
- BRAINcare: 3
- Organoids: 1
- Lab Result: 1
- Toxicity: 1

13 results list:

- Patient 10322 Date of birth 2005-07-08
- Patient 20173 Date of birth 2010-04-11
- Patient 20465 Date of birth 2008-07-14
- Patient 14025 Date of birth 2011-03-21
- Patient 15429 Date of birth 2004-09-08
- Patient 16025 Date of birth 2009-04-18
- Patient 18059 Date of birth 2010-11-07
- Patient 18114 Date of birth 2005-04-05

Link to iRODS in DISCOVER

The screenshot displays the DISCOVER interface. At the top, a network diagram shows a central node 'cgl-58' connected to 'Gene 41', 'ABHD5 [Ho...]', 'Disease 36', 'Lipid metab...', 'Publication 522', 'Collection 1', and 'Links (15)'. A box labeled 'Is target for' is positioned below the 'Gene' and 'Disease' nodes. Below the diagram, there are three filterable panels: 'Technology' (1 assay), 'Tissue (1)' (Skeletal muscle organ), and 'Related Publication (2)' (Mitochondrial dysf..., Regulation of skeleta...). On the right, a detailed view for 'Collection C000000001' is shown, including 'Project Name P000000002', a 'Link out' to <https://webdav.acc.rit.unimaas.nl/P000000002/C000000001/>, 'Date 5/11/2010', and 'Description CGI-58 is involved in the regulation of energy'.



Conclusion: Why DISCOVER & iRODS

Without DISCOVER & iRODS

Data is not FAIR

Takes days and weeks to get results

Requires 10-40 FTE to maintain

Data duplication and high storage costs

Hard to link public & third-party data

With DISCOVER + iRODS

Data can be reused and reprocessed

Results are instant

Requires 1-2 FTE to maintain

No duplication and reduced storage costs

Out-of-the-box data linkages



DISQOVER

the life science knowledge platform
that links any type of data to deliver
actionable insights

TRY IT NOW ON [DISQOVER.COM](https://disqover.com)

Carl Latham
VP Sales
Carl.Latham@ontoforce.com



Europe: Moutstraat 108 – 9000 Ghent – Belgium
North America: 955 Massachusetts Ave #312 - Cambridge, MA 02139 - USA





ONTOFORCE

ONTOFORCE