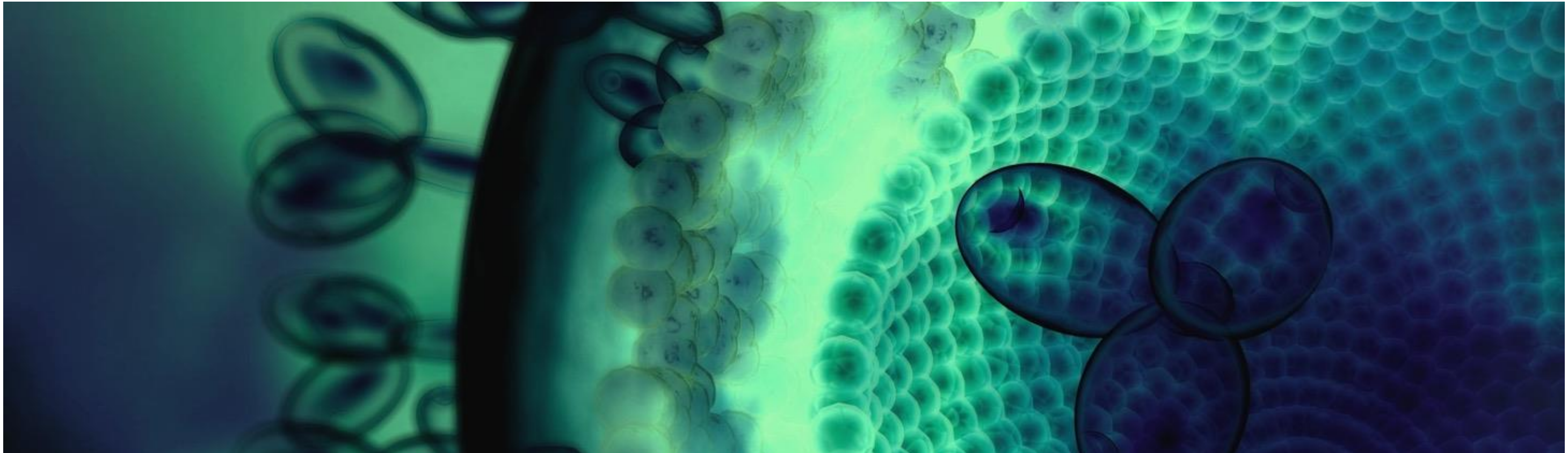




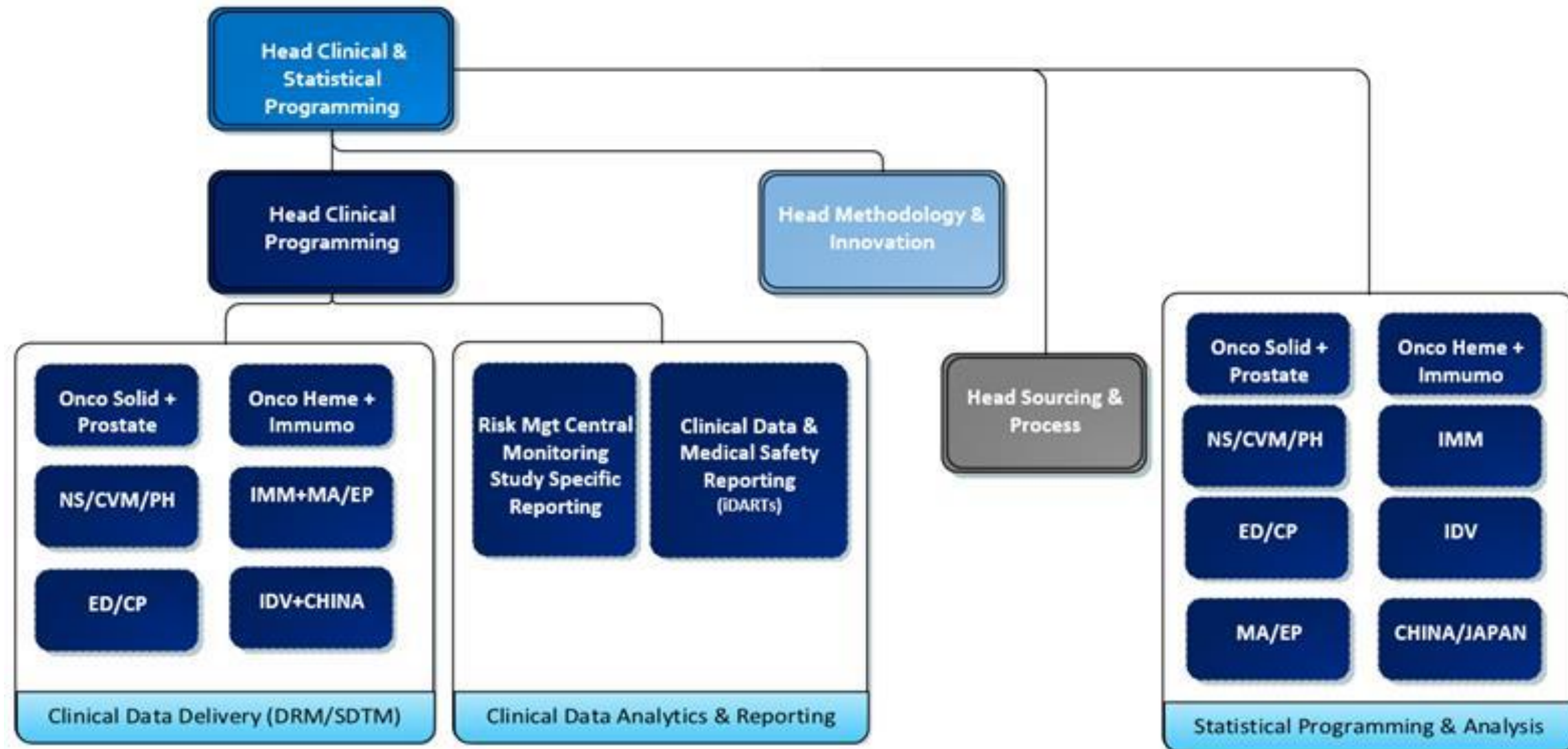
Data Engineering & Analytics at Janssen

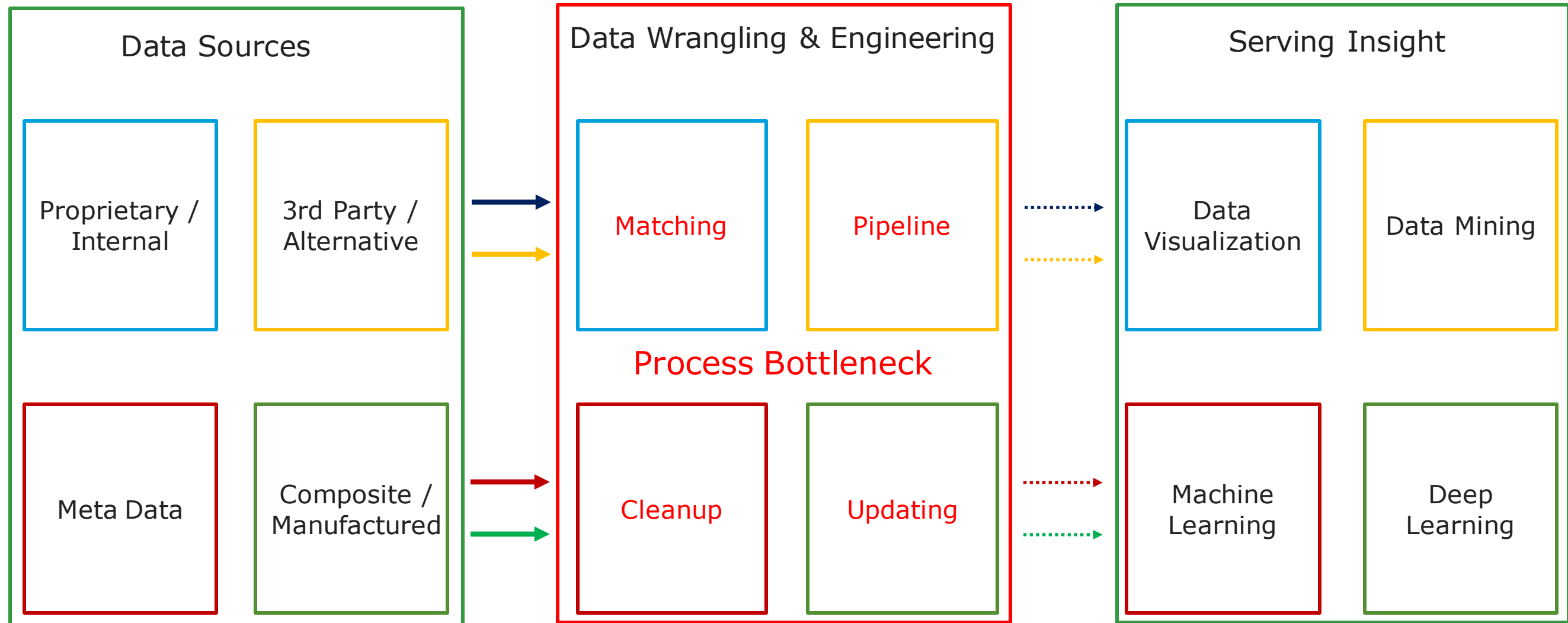
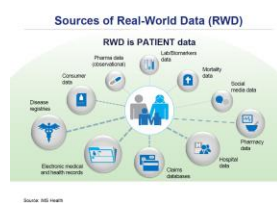
PHUSE EU Connect 2021

Sangeeta Bhattacharya
Axel Muehlig

The structure of HIV

Clinical and Statistical Programming





C&SP opportunity: Medical Data Engineering and Analytics partner for Data Science

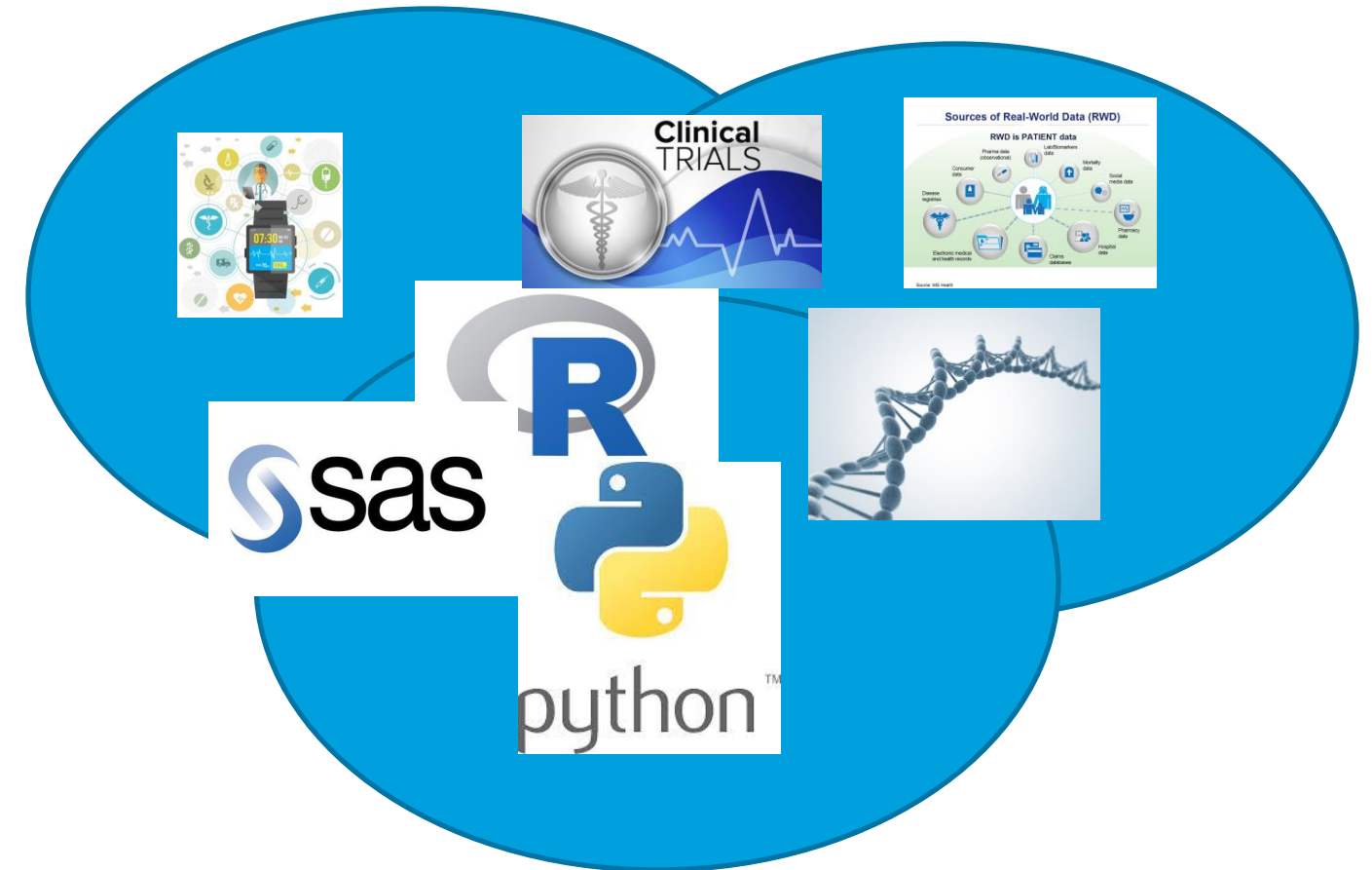
DE&A Approach

Cross-TA initiative

- Partner with stakeholders
- Deliver insights from data
- Learn as a group
- Develop Community of Practice

Broad skills & experiences

- Exposure to different data types
- Diversity of experiences and approaches
- Multitude of software (best tool for the challenge)
- **Mindset:** Analytical, curious, hacker, figure-things-out



Our Journey

2019

2020

2021



Initiation

Exploration

Consolidation



- Purpose of DE&A
- Job description



- Hiring of engineers (TA aligned)
- Build stakeholder relations
- First projects
- Initiate Community of Practice
- Ensured collaboration of C&SP staff



- Build on early success
- Grow stakeholder map
- Grow as team
- Learn from experiences and build sustained model
- Integrate DE&A work into TA

Initial Hiring Experience

Job description

... who provides **leadership**, and **strong technical, analytical** and **data wrangling** services in support of Janssen R&D (JRD) data sciences strategy and projects.

... to **actively contribute** to the development and maintenance of their **data science strategy**...

... **creating insights from the data** to **influence** decision making and/or strategy.

→ **Strategic roles**

Types of candidates:



Clinical Data background



(Bio)Informatics



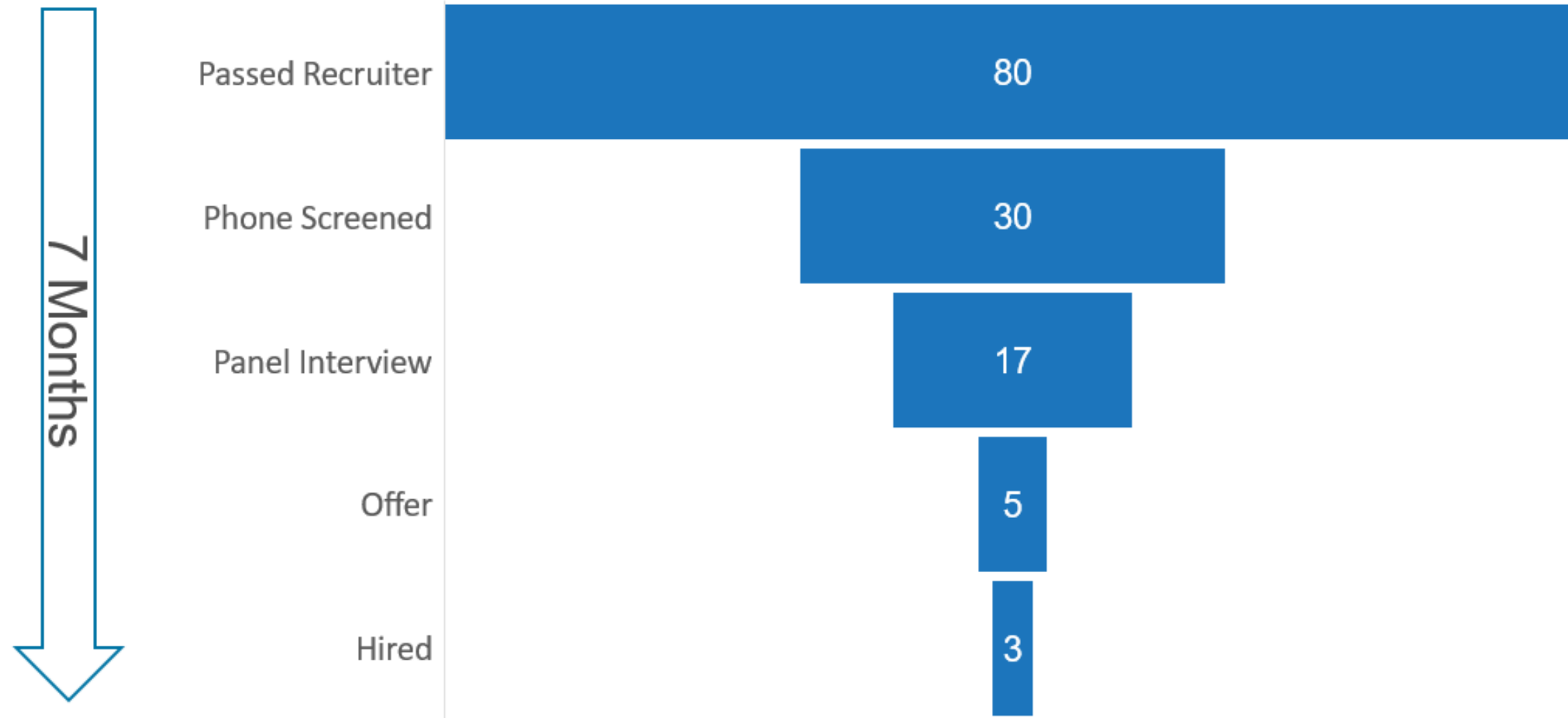
Banking industry (dashboards)

Industry experience



Outside thinking.

Initial Hiring Experience





DE&A Roles

Sponsor



Advise & provide direction on Strategy
Coach Initiative leads
Remove Road-Blocks

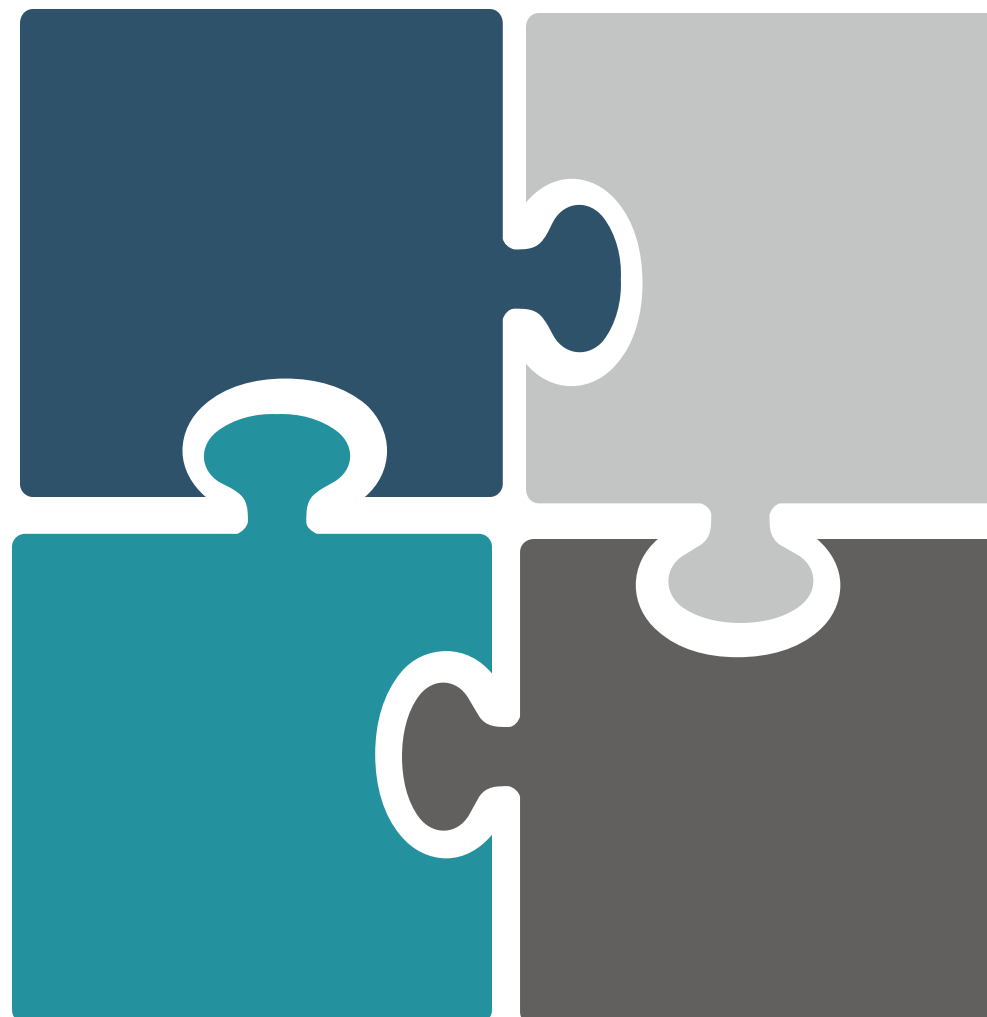


Matrix Lead

Set Strategy
Overall Accountability
Implement CoP
Expand Stakeholder Network

DE&A PL (8)

Stakeholder engagement
Assign TA project programmers
Ensure alignment on TA/project
Contribute to CoP
10-15% Commitment



Engineers (6)

Execution of Work
Stakeholder engagement
Contribute to CoP
100% commitment

SP&A Volunteers (9)

Temporary Assignment
Coached by Engineer

Benefits:

- Support TA in innovative way
- Learn & apply new skills

CoP = **C**ommunity **o**f **P**ractice

DE&A Activities

Business Areas

High opportunity
Business areas/trends
Methodology/guidance
Deliverables
Reusable code
Enabling new type of
work by C&SP x-TA

Community of Practice

- Group structure/roles
- Workflow
- Computing environment
- Information sharing

Business Problem (Feasibility): Are there investigators/regions with a higher placebo response rate?

Solution: Pool response data from several studies and visualize.

Placebo response analysis

Choose hierarchy level:

☐ Country

☒ Study

☐ Site

Choose placebo variable name:

☒ Response Rate

☐ Remission Rate

Choose type of plot:

☒ BARPLOT

☐ SCATTERPLOT

☐ TREEMAP

☐ BOXPLOT

Minimum placebo population (N):

1 494

Minimum Placebo RR (%):

1 100



Business Problem (Data Science): Identify predictors for clinical outcome in Prostate cancer patients. Need the data in format for model fitting (6 studies).

Solution: Develop workflow to identify rich enough data points and help communicate with stakeholder. Then pool the data.

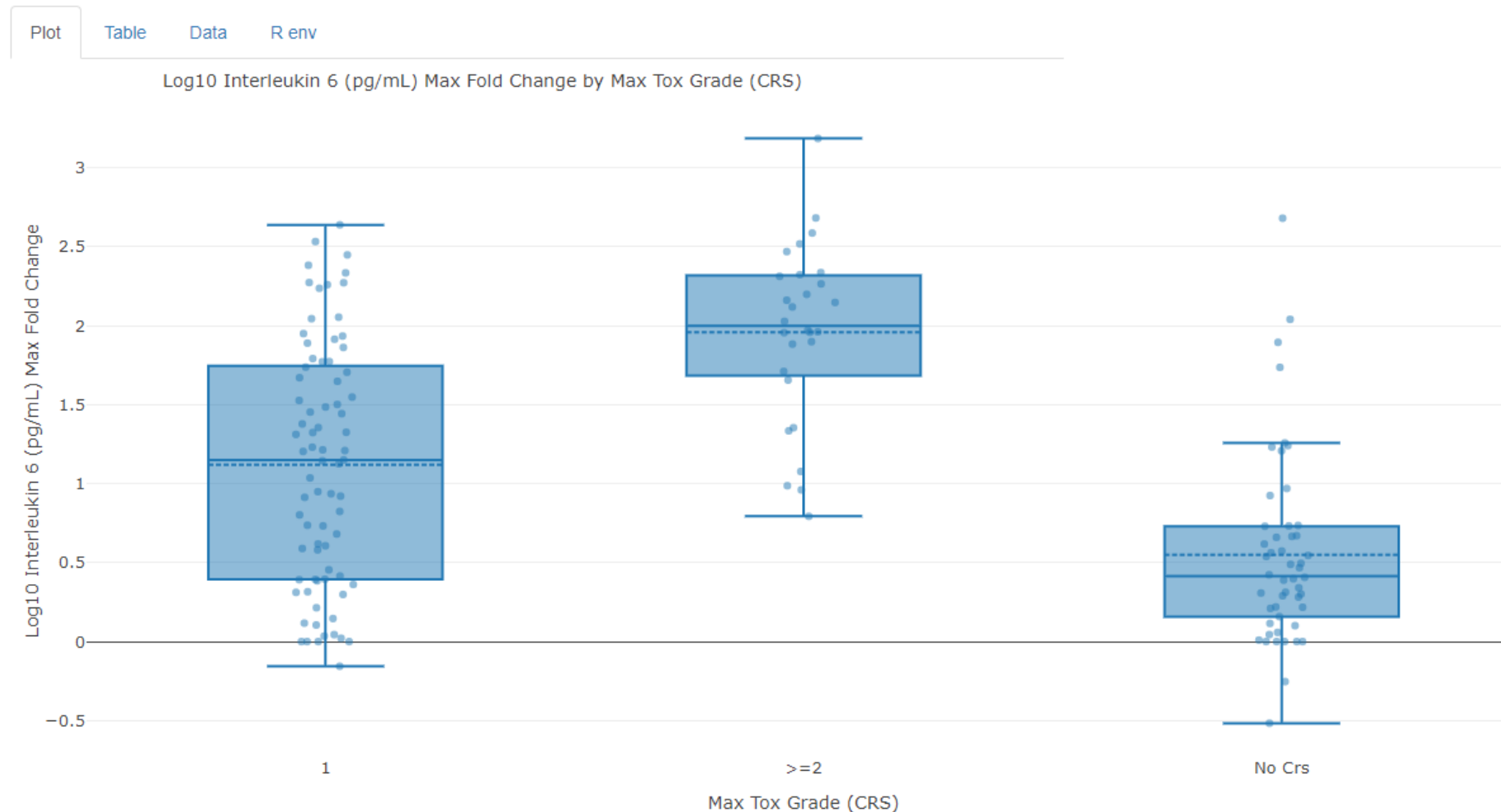
⇒ Below is a list of all derived and source data sets from the 6 studies that have “bone”, “lesion”, “recist”, or “target” in the data set info:

DATASET	ADRB	ADRB	ADRB	ADRB	ADRB	ADRB
DERIVED: ADBONE (Bone Progression Analysis Dataset)	—	—	Y	Y	—	—
DERIVED: ADBONE_OLD (Bone Progression Analysis Dataset)	—	—	—	Y	—	—
DERIVED: ADRECIST (RECIST Analysis Dataset)	—	—	Y	Y	—	—
DERIVED: NLA (New Lesion Assessment)	Y	—	—	—	—	—
DERIVED: NTRG (Non-Target Lesion)	Y	—	—	—	—	—
DERIVED: RECIST (Response Evaluation per RECIST)	Y	—	—	—	—	—
DERIVED: TRG (Target Lesion)	Y	—	—	—	—	—
SOURCE: SCRBONE	—	Y	—	—	—	—
SOURCE: XL (Bone Lesion Data)	—	—	—	—	Y	—
SOURCE: XR (Lesion Results)	—	—	—	—	Y	—

Business Problem (TM): Ability to pre-select biomarker analyses for CSR (Oncology study).

Solution: Harmonize existing Shiny apps. Create alternative data pipelines to feed biomarker and clinical data into the same app.

Outputs

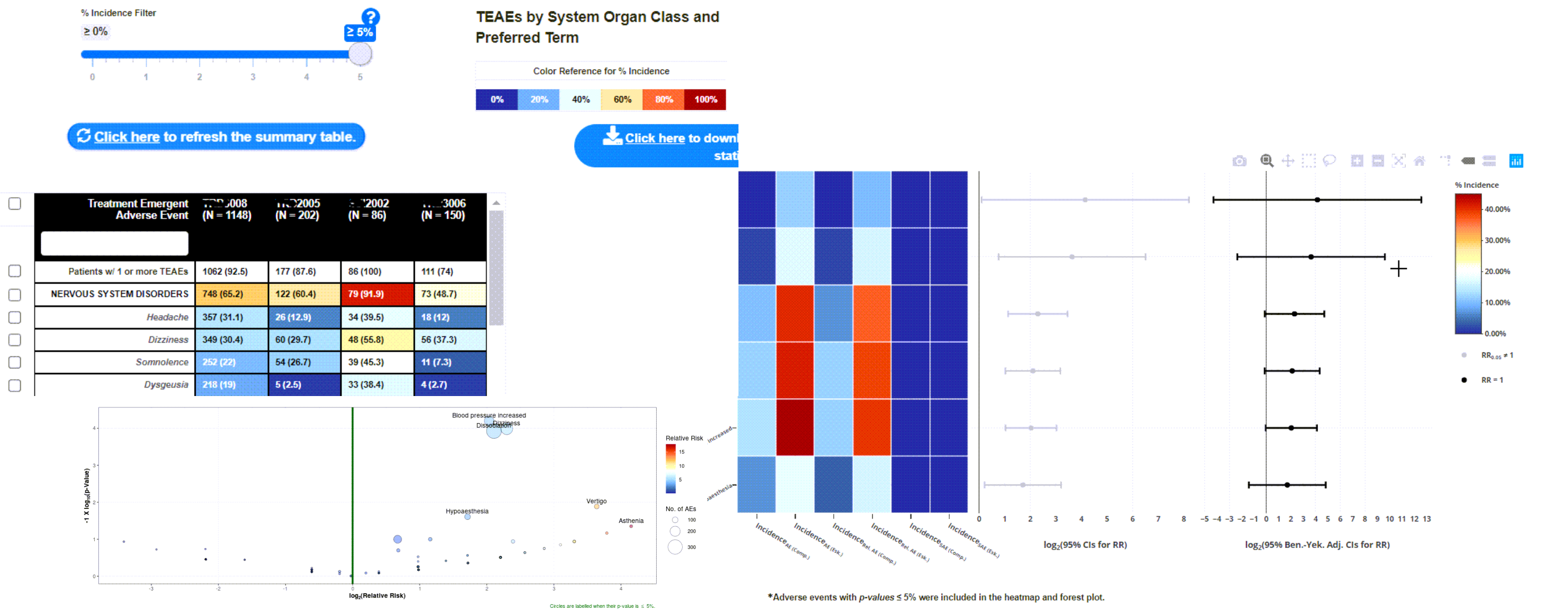


Warning: Input data were prepared for exploratory purposes only, please do not use for presentations or submissions.

Will be scaled up for wider use.

Business Problem (Safety Science): Efficient exploration of AE data across compounds. Ability to see trends (birds eye view) and drill down to patient level.

Solution: Data flow/pooling in data lake, Shiny app as end user interface.



DE&A Focus Areas Formed

FEASIBILITY

Harmonize visualization and analytics for similar requests across TA's to design better trials from Janssen Clinical trial data.

RWD

Create a guidance in the form of a Playbook for SP&A.
Aligned with the life-cycle of the study/project utilizing RWD/RWE.
Curate code to be used in future studies.

SAFETY

Build Partnership with GMS.
Provide support on visualizations & compound level proactive safety monitoring.

TRANSLATIONAL MEDICINE

Develop guidelines for integration of biomarker data with other types of data.

ACTIGRAPHY

Create guidance on analyzing actigraphy data for use in studies.
Identify reusable code and make it available to users.

DE&A Book of Work Topic Synergy*

Feasibility/ DS



Enterprise	4
Immunology	10
NS	2
Oncology	2

Translational Med



CVM	1
ED/CP	4
Enterprise	1
Immunology	4
NS	1
Oncology	5

Safety



CVM	1
ID&V	1
Immunology	1
NS	1
Oncology	2
Grand Total	6

RWD



Immunology	1
MA/EP	1
NS	1
Oncology	5
PH	1

* Requests originate from & worked collaboratively with various stakeholders : Data Sciences, Feasibility COE, Clinical, GMS , C&SP TA, QS

Learnings/Challenges

- Reliable process of tracking, resourcing and delivering on stakeholder requests.
- Flexibility of shared resources and building knowledge experts in key focus areas, and transferring solutions to TA programmers to use, thereby elevating collective department skills.
- Continuous training in the field of data engineering and visualization/analytical skills.
- Robust Stakeholder Map, and ability of data engineers to connect with key TA stakeholders.
- More opportunities for automation.
- Learning curve for out of industry talent.
- Help stakeholders understand their requirements/translate into solution.
- Data engineering projects deprioritized by TA Head over submission projects.

Data Engineering & Analytics – Evolution Proposal



Contact: SBhatt11@its.jnj.com, AMuehlig@its.jnj.com

