Automation, Automation, Automation:

Keeping Up with Exponentially Increasing Data Volume via Compliant, In-House Digitalisation

10-OCT-2023

OALT

12 October 2023

Novo Nordisk

Our journey to create an in-house automation group

- Development Data Science at Novo Nordisk
- Why build an Automation Group?
- What have we built already?
- What's on our roadmap?

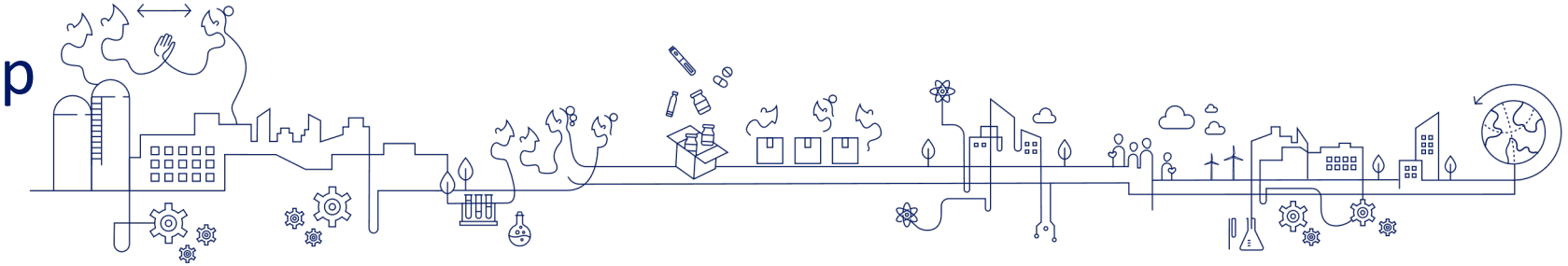
Views and opinions expressed are those of the speaker and not Novo Nordisk.

oalt@novonordisk.com

01

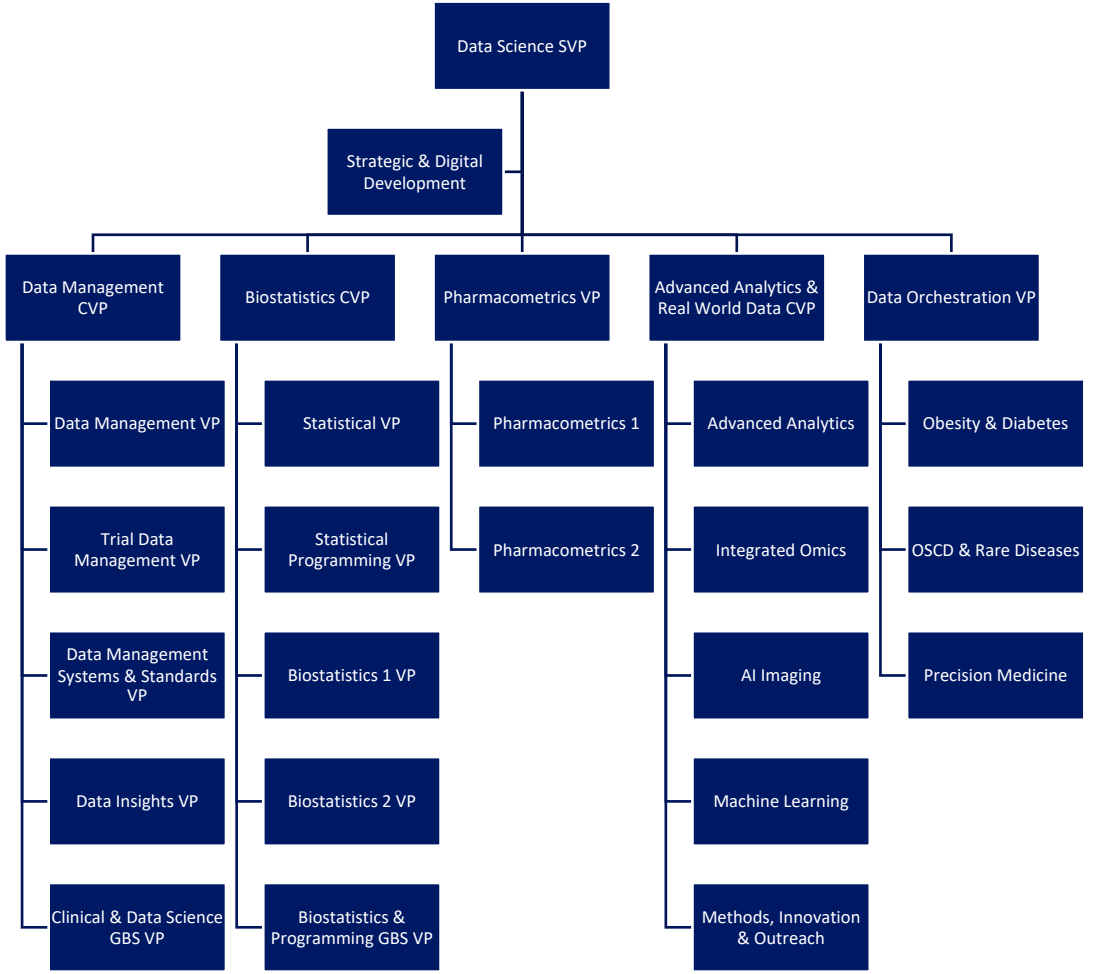
Development Data Science

DS Automation Group

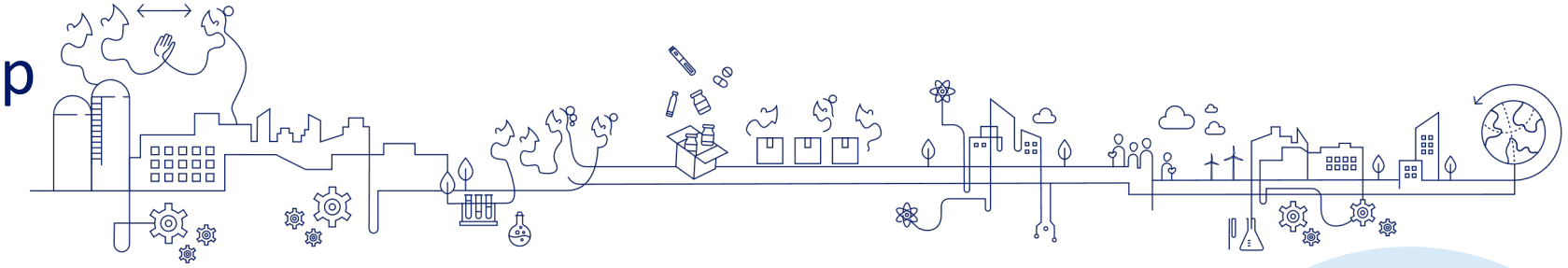


Development Data Science

- 1000-strong organisation, located in DK, IN, ESP, USA
- Data Science SVP area (since 2021)
- Data *collection, curation* and *analysis* for **clinical trials** all under one ”roof”
- **Data Science Automation Group (DSAG)** established June 2022 to coordinate, plan, develop and implement automation initiatives across the value chain



DS Automation Group



Exponential data growth leads to exponential resourcing needs

Complexity

2011
~6,000 Ph3 patients

2011
110,000,000 Data points

2016
~7,000 Ph3 patients

2018
13,000 patien years

2018
~20,000 Clinical AEs reported

2018
>275,000,000 Data points

2019
~10,000 P patient

2020
~38,000 Clinical AEs reported

2021
40,000 Patient years

2023
~86,000 Clinical AEs reported

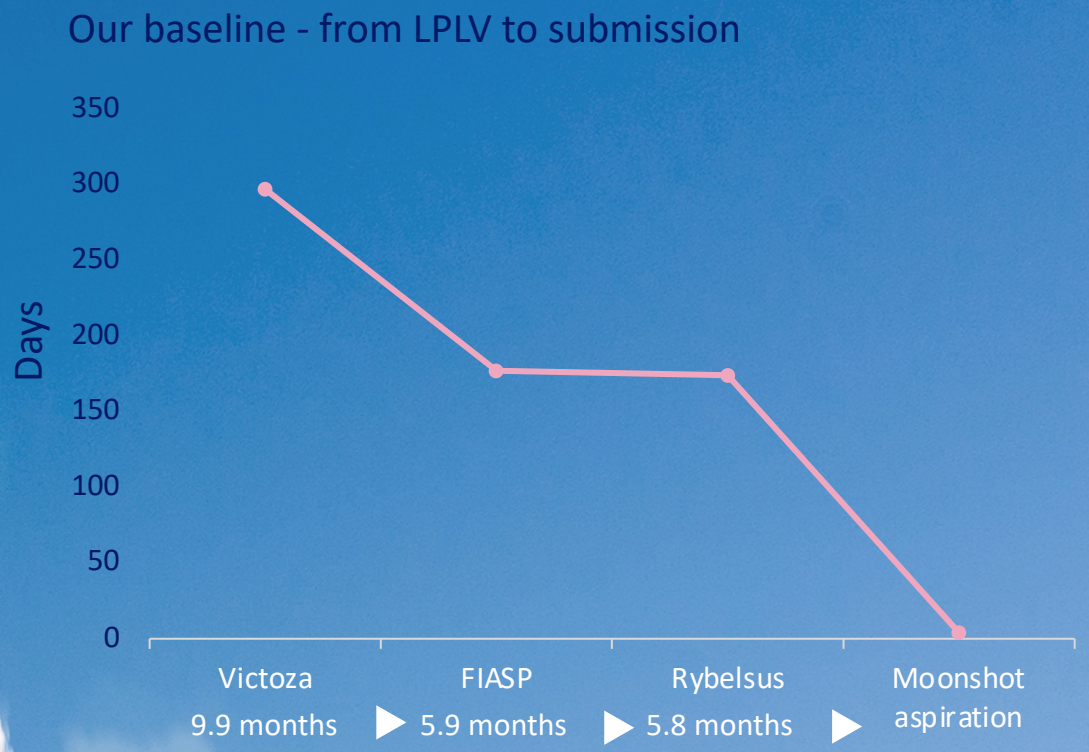
2023
56,000 Patient years

2024
>800,000,000 Data points

Time

Fit 4 Take-off

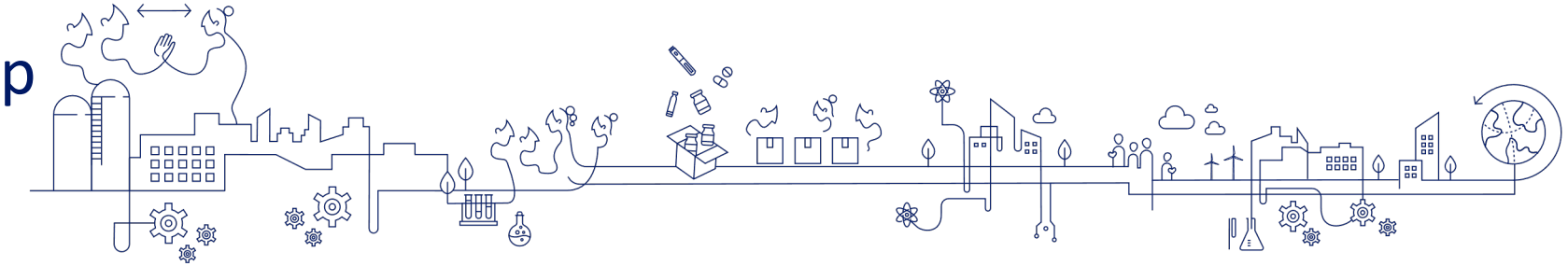
- LPLV to Submission in **4 days**
- To the **4 largest markets** (USA, EU, China, Japan)
- With successful access and **launch 4 days after approval**



02

Why build an automation group?

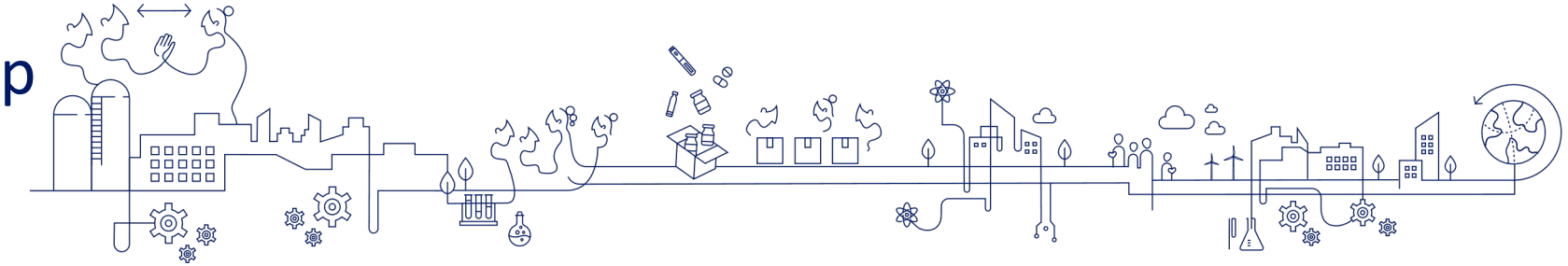
DS Automation Group



What is automation?

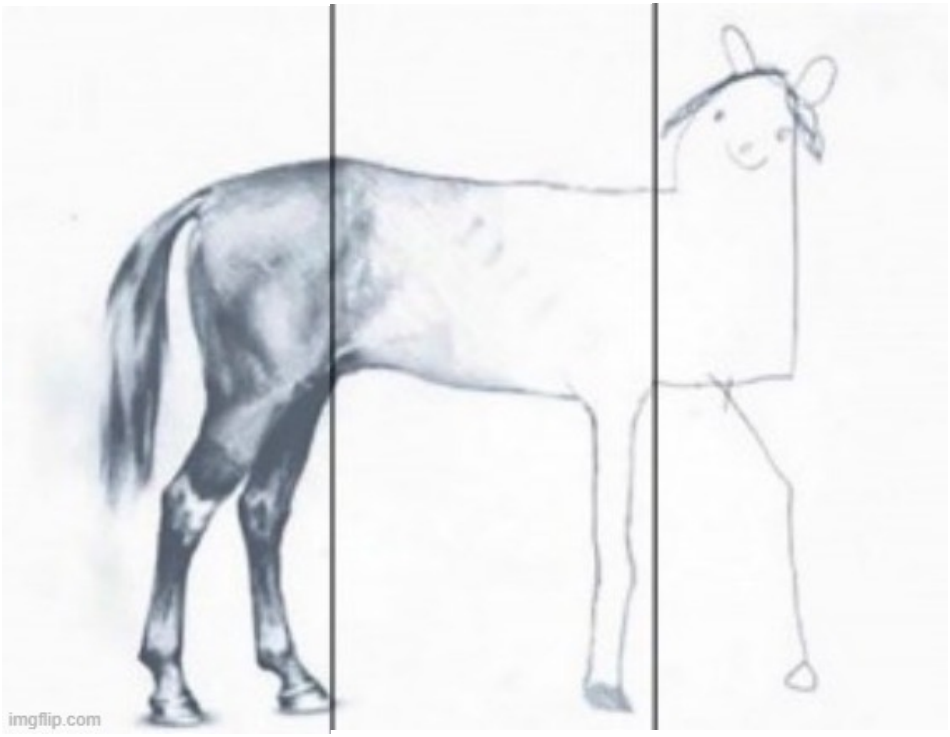
- Use of digital tools to speed up the execution of repetitive, rule-based tasks – accurately and consistently
- Depends on well-understood, well-oiled processes
- Not synonymous with AI, but AI plays an increasing role
- An essential tool to keep with our clinical ambitions!

DS Automation Group



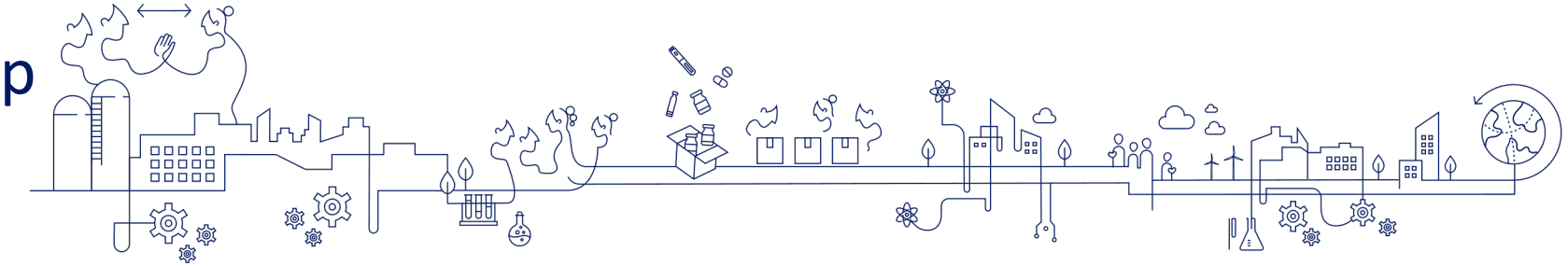
Why automate in-house?

Design specification | MVP | Production solution



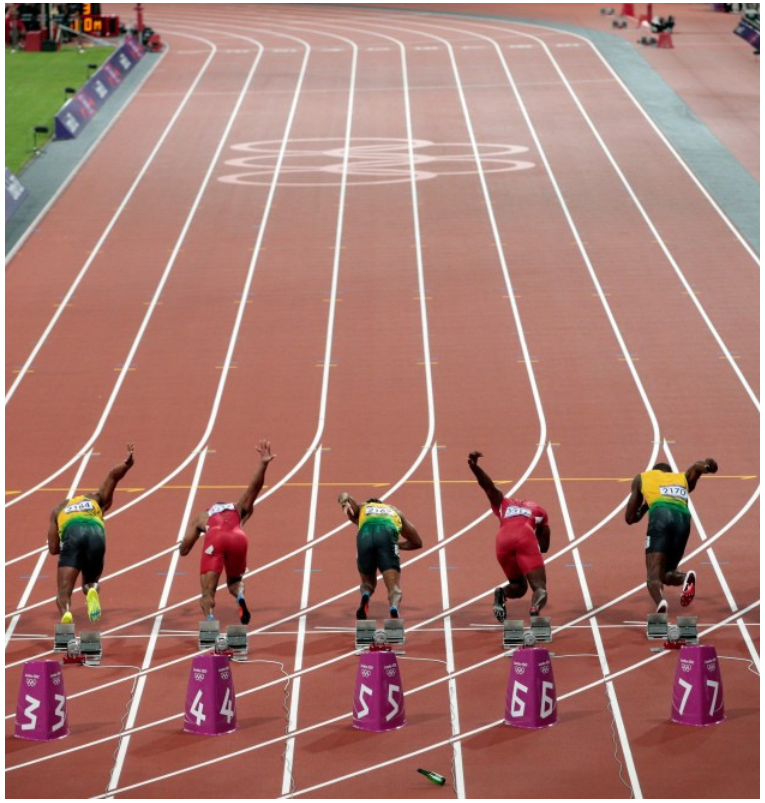
- **Rapidly prototype** highly tailored solutions which fit to NN process / system landscape and business priorities
- Control exposure to vendor lock-in
- Building competencies to *demystify* and *understand* vendor solutions, allowing us to provide technical inputs to NN leadership decision making
- Potential to build **Open Source solutions**, and thereby influence regulatory decision-making
- Support organisational transformation from Clinical Data Management to **Clinical Data Science**

DS Automation Group



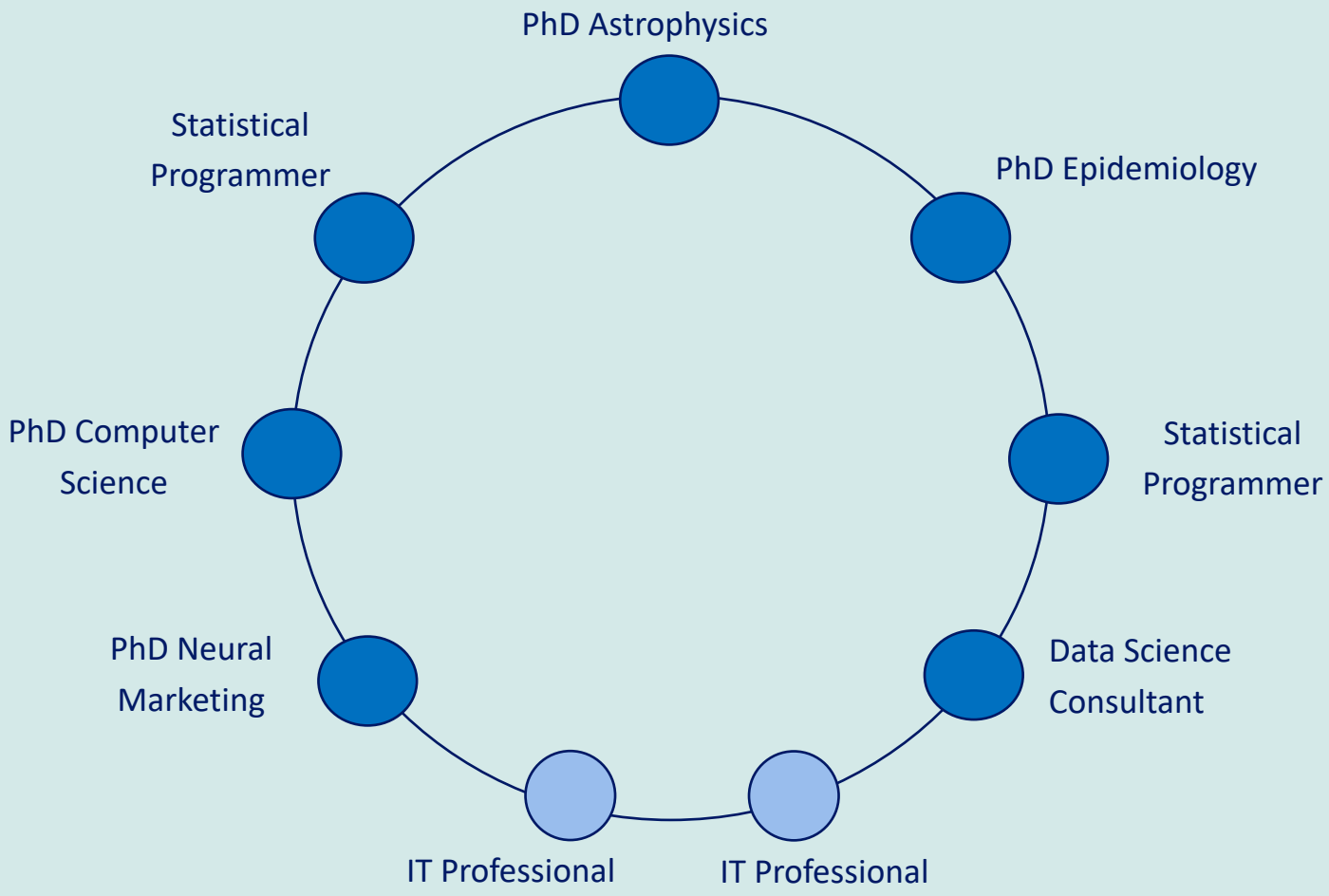
Why automate in-house?

- It’s a marathon **and** a sprint – happening at the same time!
  - ”Micro solutions” to address limitations of legacy systems: reports, dashboards, reconciliation tools, etc
  - Plan process improvements around the capabilities of new systems
- *How will we be doing Clinical Data Science 5+ years from now?*



Data Science Automation Group (DSAG)

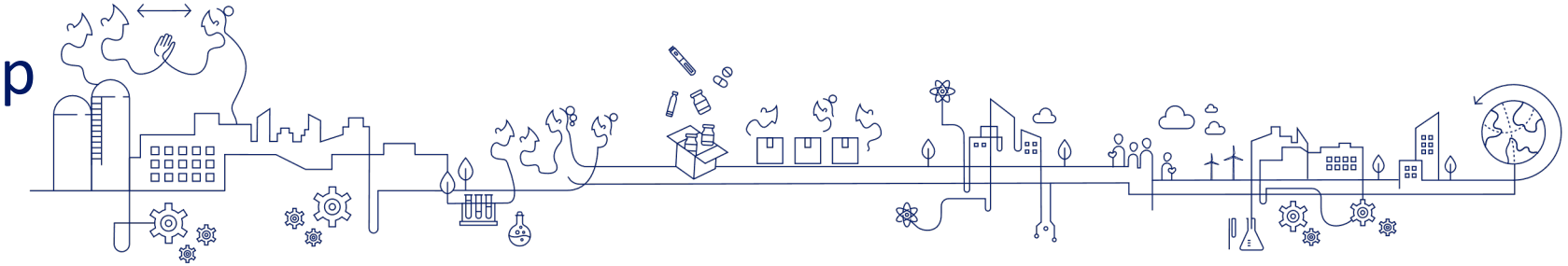
Profiles



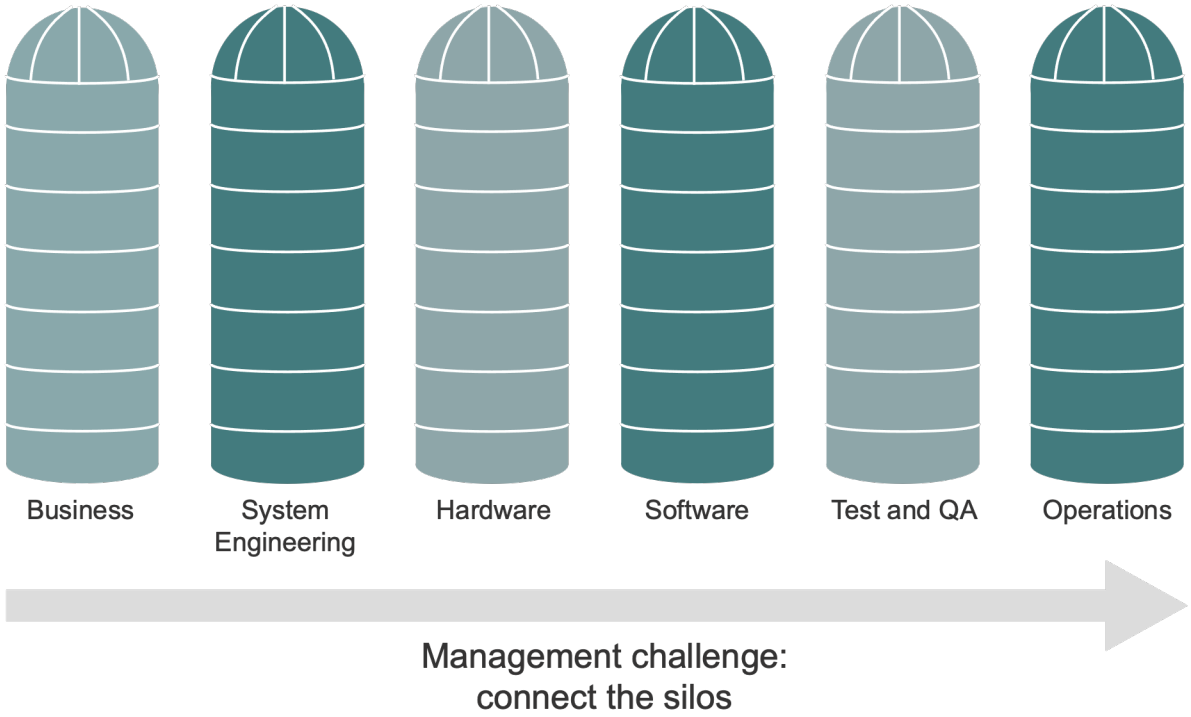
- Specialist-led group from DM / Biostatistics, with input from IT colleagues
- Skilled in: R, SAS, Python, AI/ML, Power BI, etc
- Drive automation initiatives i.e., ideation, development, change management and solution roll out, in collaboration with LoB
- Ensure alignment of automation initiatives, knowledge sharing and synergies across DM and Biostat
- Maintain short- and long-term outlook for automation opportunities across DM and Biostat in line with corporate strategy
- Provide status on automation initiatives to leadership on a regular basis
- Anchored in Clinical Analytics ART

Full time allocation Part time allocation

DS Automation Group

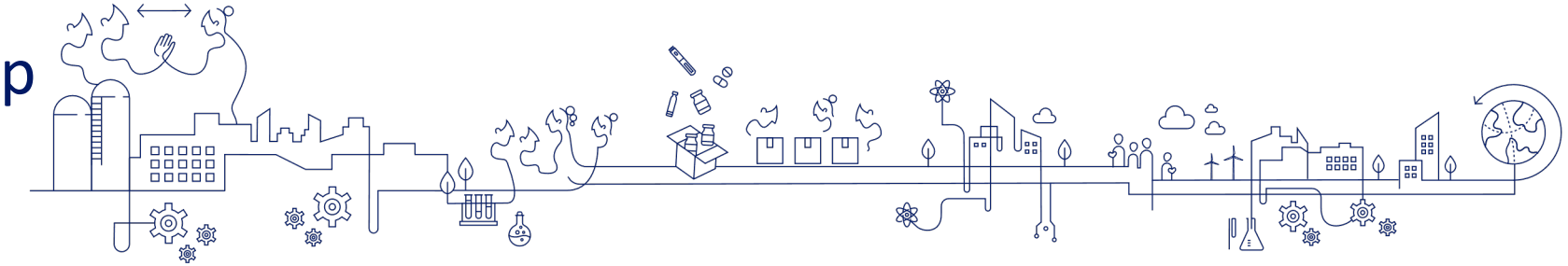


Product development: non-Agile

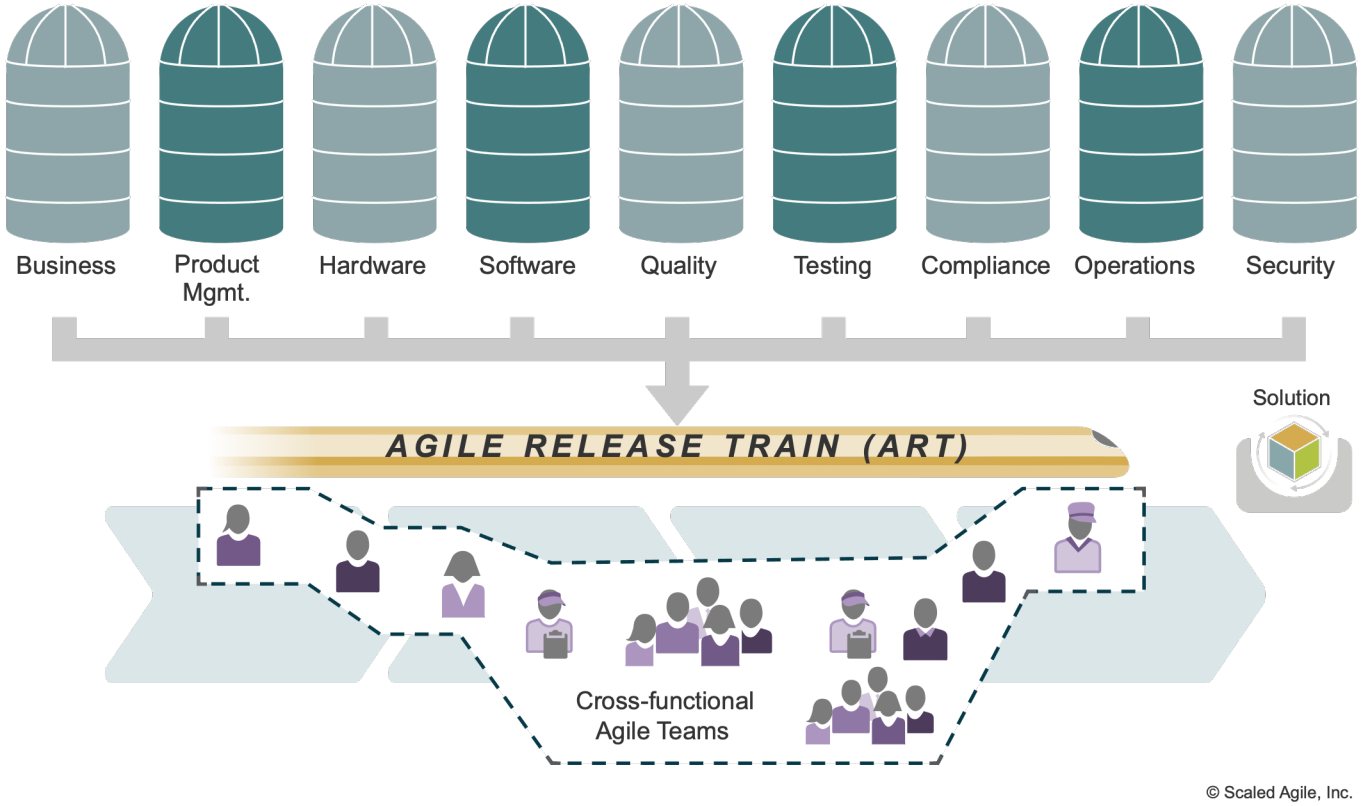


- Value delivery is inhibited by hand-offs and delays
- Political boundaries can prevent cooperation
- Silos encourage geographic distribution of functions
- Communication across silos is difficult

DS Automation Group

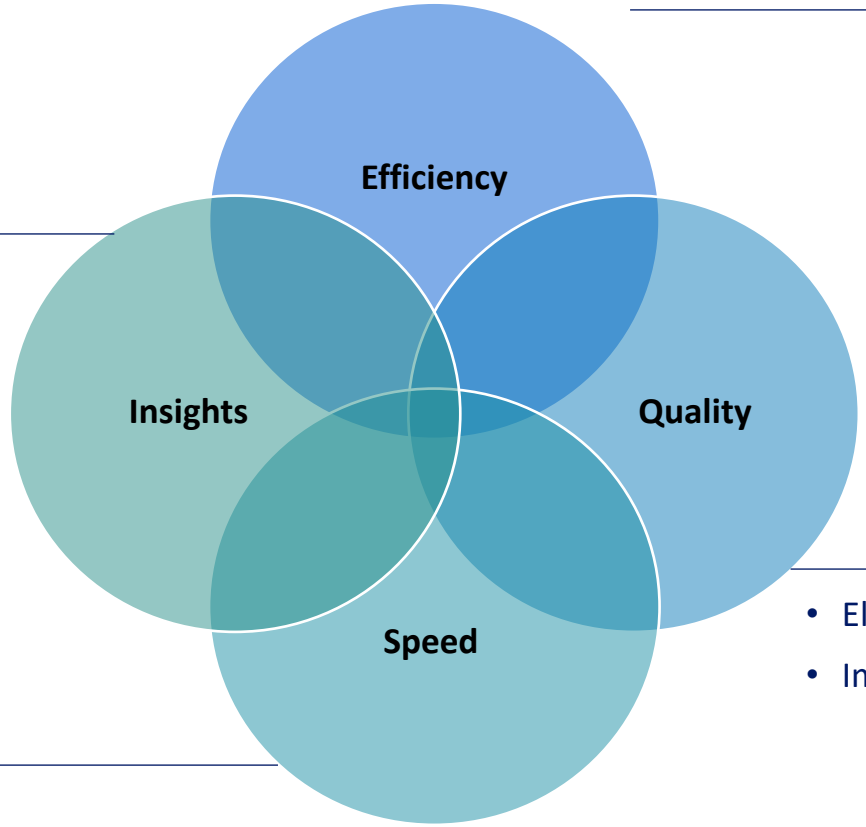


Product development: Agile



- Cross-functional teams, aligned around shared business and technology goals
- Established around organizational Value Streams
- Capabilities to define, build, validate and release solutions
- Continuous release of value by rapid release cycle: 10-12 week PI cadence

Four main value drivers of automation



- Breaking proportionality between # datapoints and # FTE
  - Essential to meet increasing data volumes without unsustainable growth!
- Potentially increasing output

- Creating more complex models
- Unlocking the hidden value in our data and enabling new discoveries

- Eliminating human error
- Improving consistency and accuracy

- Reducing cycle times
- Fit 4 Take-off

Feature requests in an ART

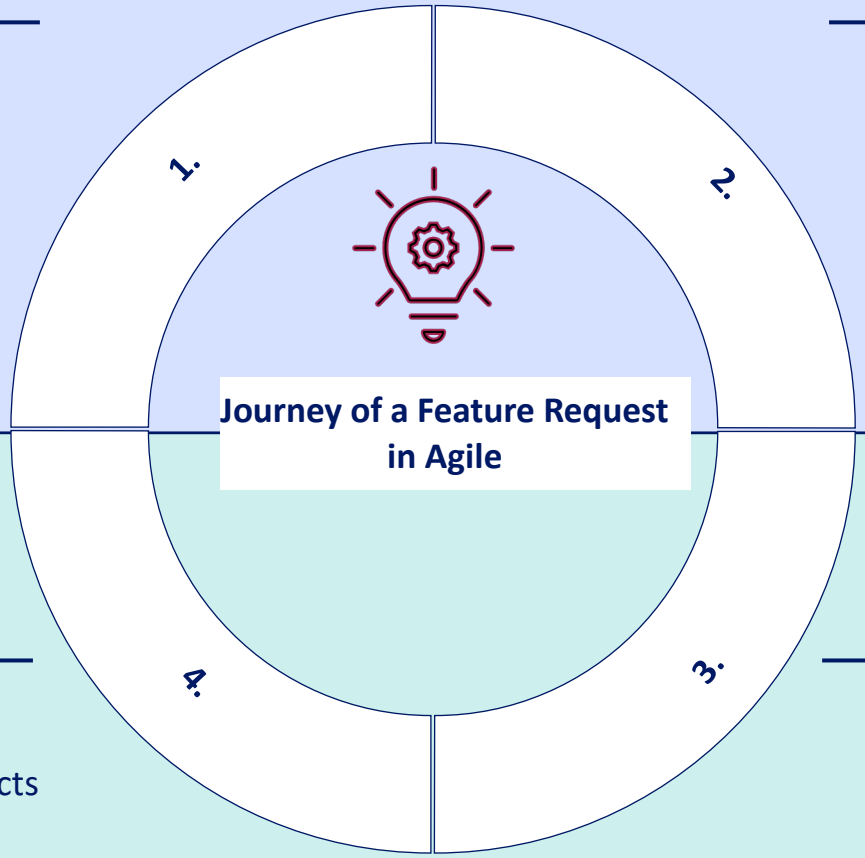
Ongoing

Idea (LoB)

- What is the issue?
- How many people will it benefit?
- How much value will it add?
- What systems are involved?

Dialogue (DSAG & LoB)

- **DSAG**: Ask for more details
- **DSAG**: Merge similar ideas
- **LoB & DSAG**: Sparring & Business case
- **LoB & DSAG:** Split idea into user stories



Every 10-12 weeks

Outcome (DSAG & LoB)

- Develop & Implement?
  - **LoB** will be co-owners and co-architects
- Park for later?
- Reject

Prioritisation (DSAG/BOs)

- Impact on **Speed**?
- Impact on **Efficiency?**
- Impact on **Quality?**
- Impact on **Insights?**

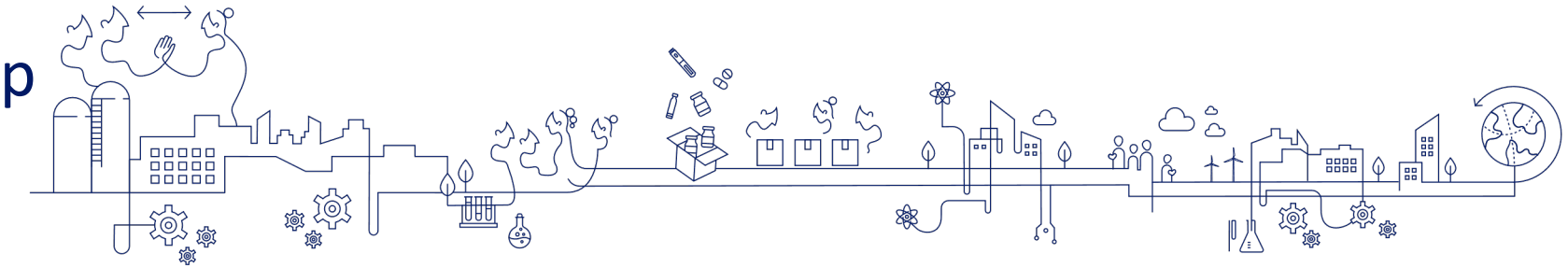
03

What have we delivered already?

03a

Safety Narratives

DS Automation Group



Safety Narratives

Business Context

- Medical writers create text-based narratives and TFLs describing patient AEs
- Requires extensive knowledge of SDTM and ADaM
- Created after DBL, i.e. on critical path wrt. F4TO

Approach

- SAS-based TFL template program
- Process re-anchored where ADaM expertise lies

Gain

- 2 days of manual work automated post-DBL
- Potentially 2 days faster to market

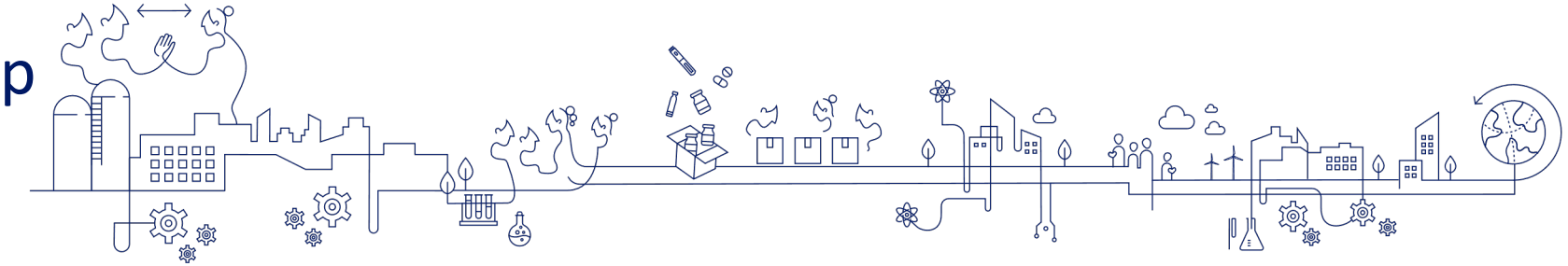
Status

- Tool supported OASIS 1 to F4TO submission; 17 trials in scope for extended roll-out in 2023/2024

03b

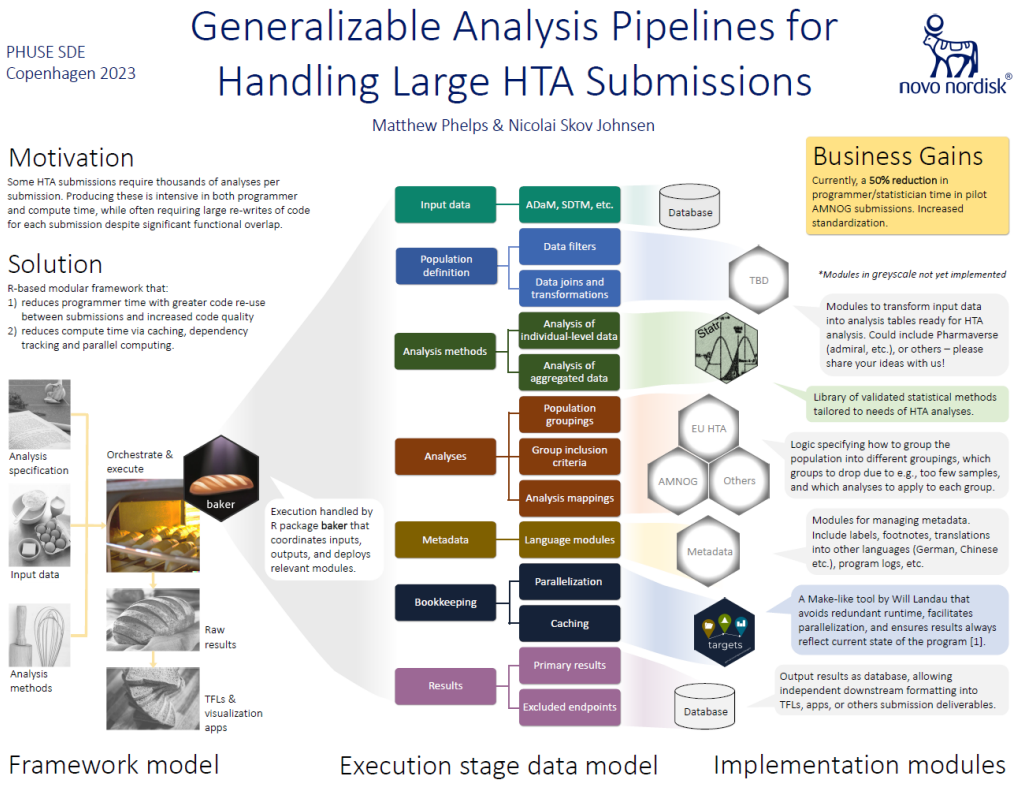
HTA Submissions

DS Automation Group



AMNOG

Poster advertisement:



Matthew Phelps

Statistical Programming Specialist

04

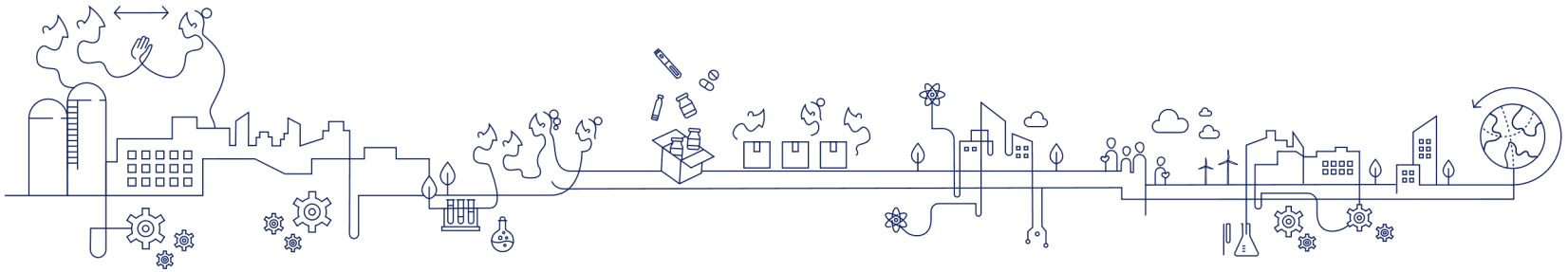
What are we working on now?

04a

Kassandra: automated test data for Fit 4 Take-Off

Kassandra

Automated test data for Fit4Take-off



What?

- Software and processes to enable seamless flow of automated *trial-specific* test data from EDC to SDTM (submission data format), ahead of FPFV

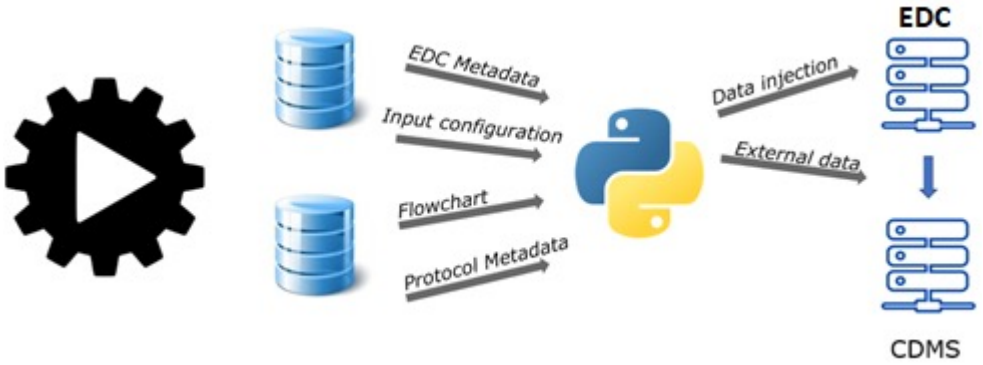
Why?

- Biostatistics programming on critical path to submission: typically not possible to do trial programming until *real* data arrives.

When?

- Technical PoC developed with external consultancy firm, June 2022
- Automation group refactoring code and extending functionality to enable organisational role out in Q4 2023

How?

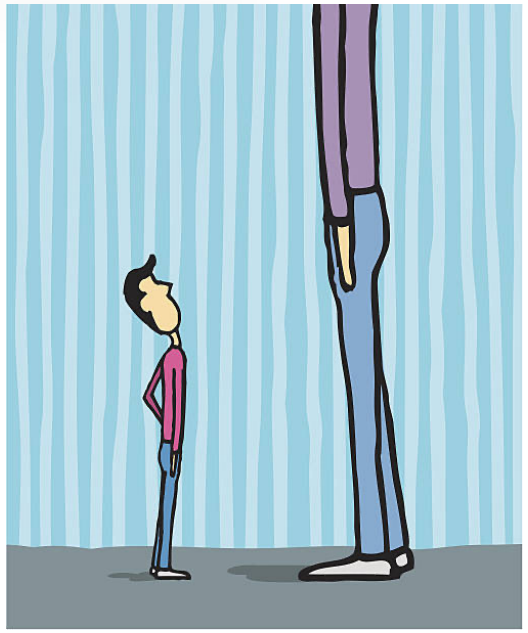


- Python engine: (i) reads study specification metadata (CSML); (ii) generates patients and populates all forms in EDC and other data models (e.g. lab data); (iii) delivers study-specific SDTM (submission format) test data to Biostat.
- Potential AI use case to make test data “realistic”

AI-enabled, Risk-Based data cleaning

Query Types

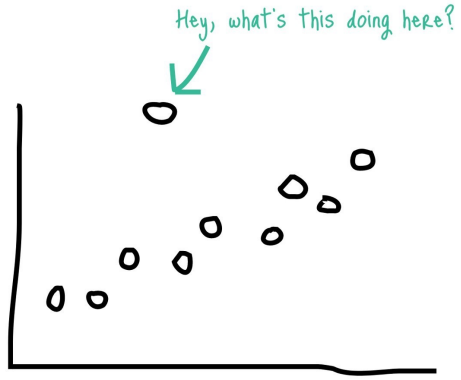
1-Dimensional data



Example: height = 5.0m

Currently automated through EDC Edit Checks

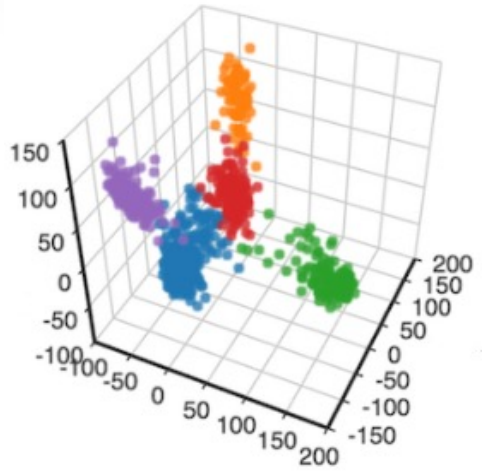
2-Dimensional data



Example: weight = 150kg, waist circumference = 26”

DSAG visualisations support DM decision making (e.g. D-VISTA)

> 2 Dimensional data

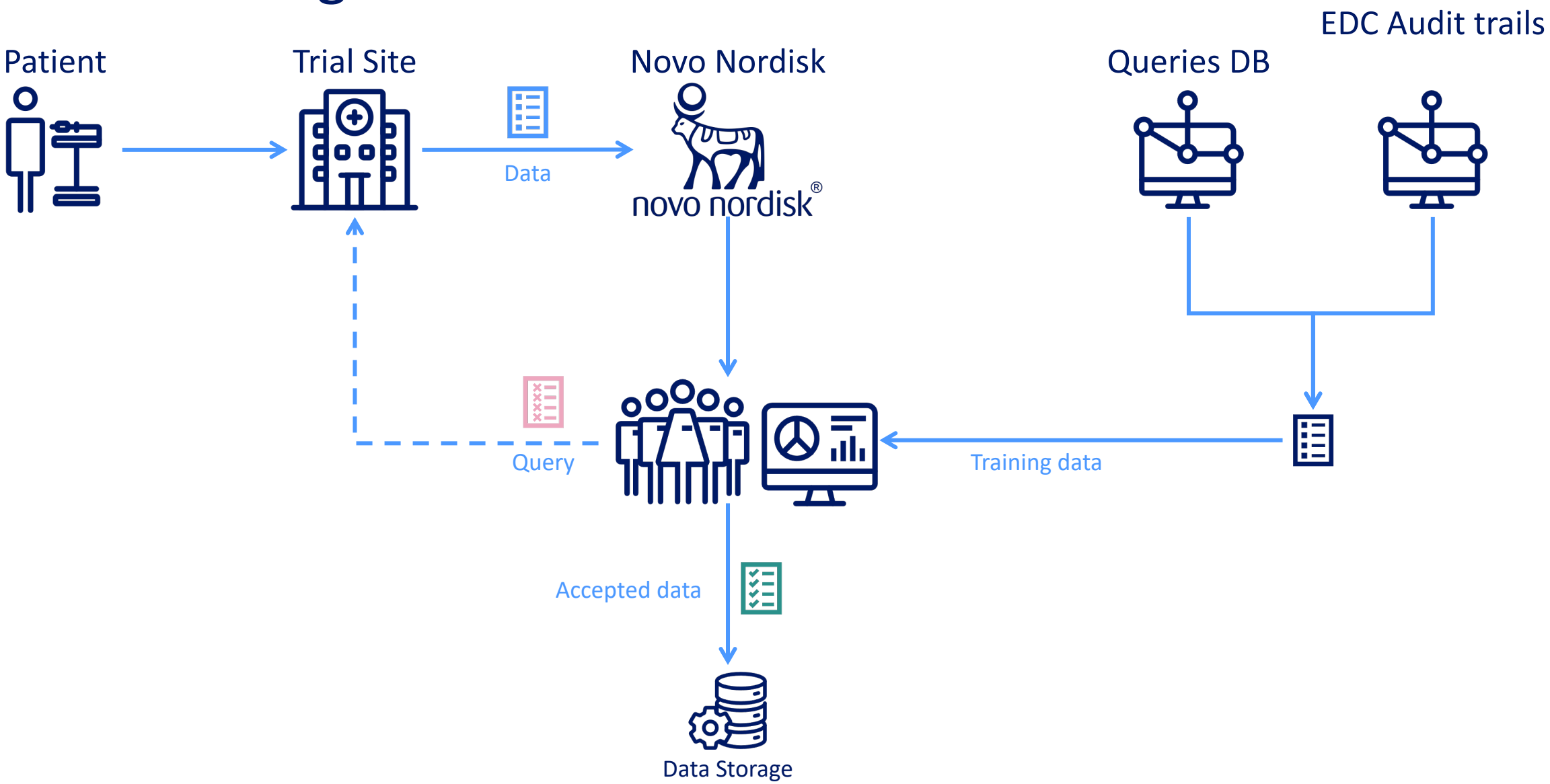


Example: subject status reconciliation across multiple forms/data sources (e.g. IWRS-EDC)

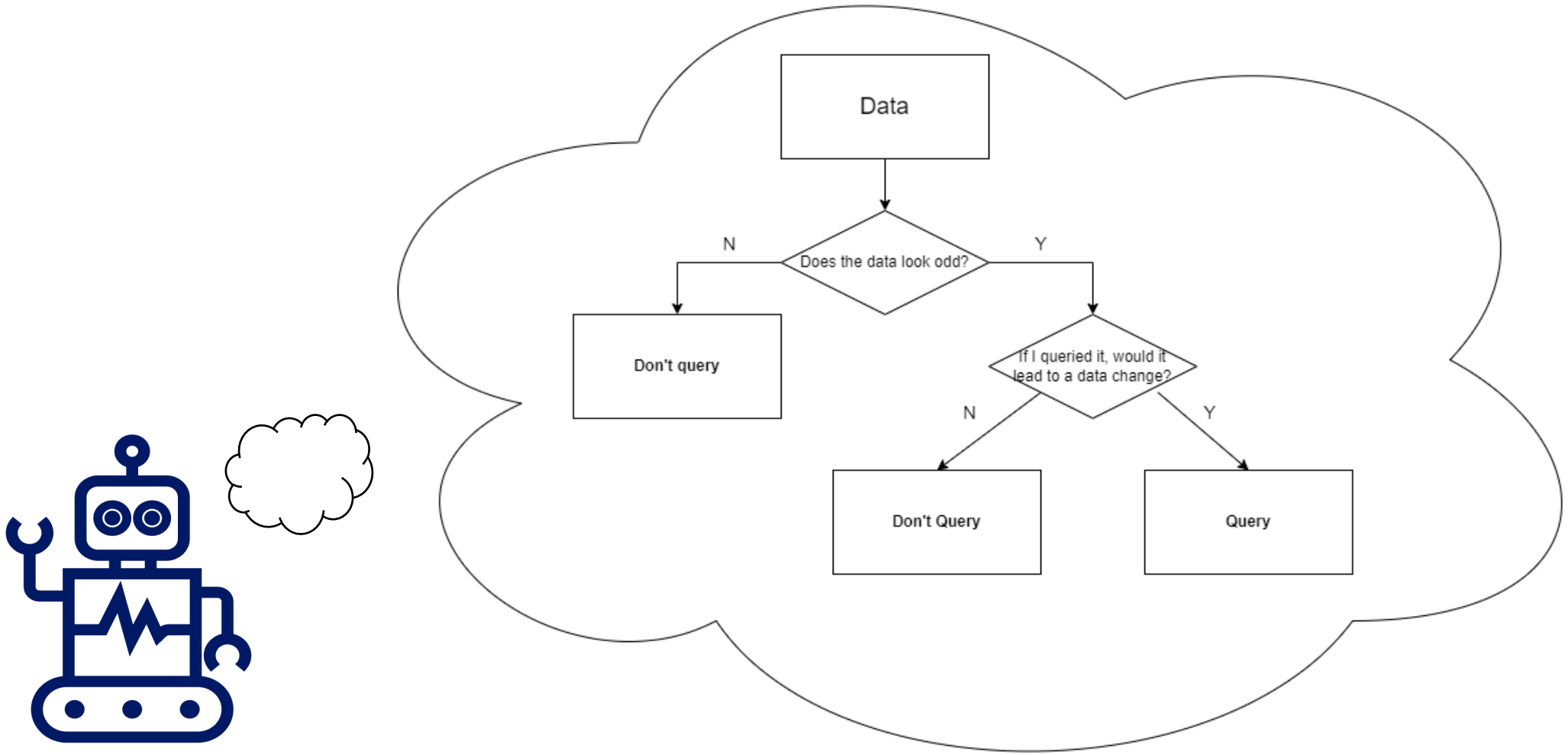
Identified through custom reports (Excel, Python, dashboards)

- *Need to define what you are looking for in advance*
- *Manual review of visualisations/reports not scaleable to expected data loads without enormous hiring growth*

AI-driven insights for DM



Supervised learning: can we spot *valuable* queries?



DS Automation Group

AI Queries Pt II: What not to query

Business context

- Want to know if a given query will impact the data or not so that we can eliminate non-value adding queries
- Saves time/resource and improves trust with sites

Approach

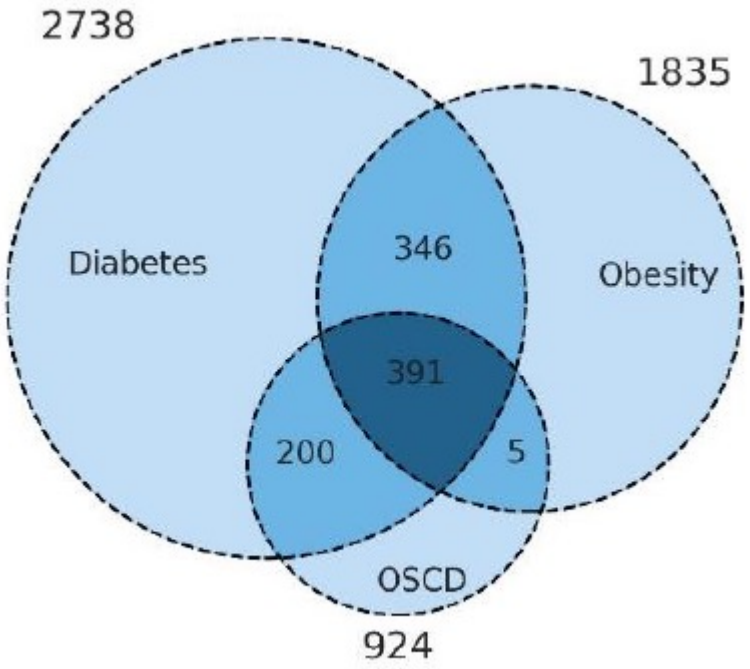
- Supervised learning connecting historic queries DB with EDC audit trails to identify data changed in response to queries

Challenges

- Different variables collected on different studies – difficult to define a training set which is applicable across entire trial portfolio
- Huge class imbalance: 0.4% of EDC data changes due to queries

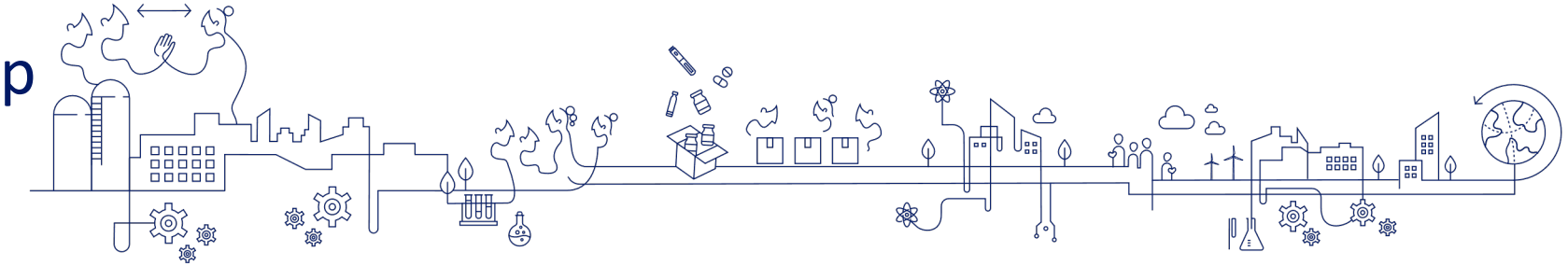
Status

- Initial assessment of LR, RF, SVM, EEXGBoost models
- XGBoost handles class imbalance well, and adaptively selects features of relevance – but at the cost of explainability



Common fields for trials in selected therapy areas

DS Automation Group



AI Queries Pt I: Outlier detection

Business context

- Want to identify anomalies in data to guide DMs to suspected data quality issues
- Current OD strategy relies on cuts based on fixed “out of range” criteria, and subjective assessment of data visualisations (unrepeatable, unscientific)

Approach

- Framework for repeatable statistical analyses on SDTM data
- Auto-identifies correlated variables in trial SDTM dataset: beyond “hand picked” visualisations.
- Considers: univariate, bivariate, multivariate, time series outliers
- Provides explanatory text to say *why* data were flagged

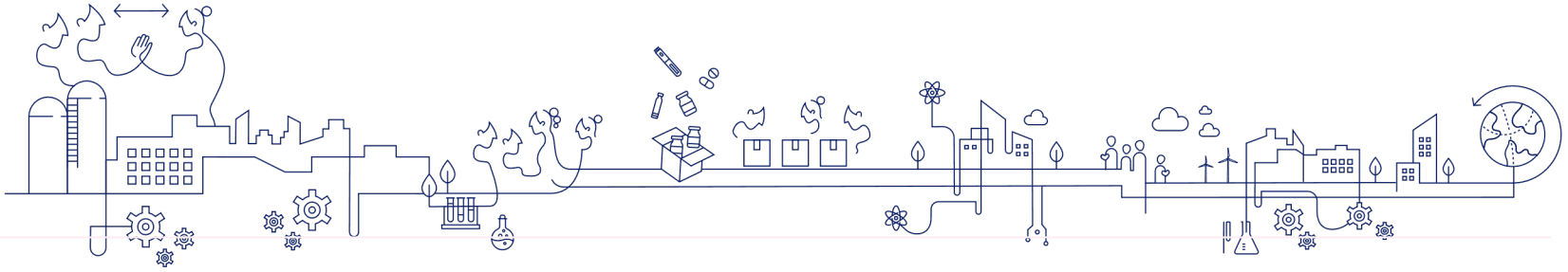
Challenges

- SDTM data model far removed EDC data models – connection to historic queries DB tricky

Status

- Majority of statistical tests coded; plan to implement in visualisations framework, D-VISTA in 2024

Conclusions



- If you’re not already thinking about automation, start now!
- **DS Automation Group:** specialist-led group of data scientists doing in-house digital innovation in Novo Nordisk to support clinical trials
- We maintain a short-term and long-term outlook for the needs of the DS community
- In-house automation can be *compliant* and *cost effective*
- Potential business value of solutions developed in first year:
  - *Efficiency:* >10s FTEs
  - *Speed:* potential 2 days faster to market
  - AI-enabled solutions coming to production soon (human in loop!)
- We are just getting started!!

oalt@novonordisk.com

LinkedIn alasdair-p-thomson