

Democratising Analytics:

Enabling the utilisation of re-usable and efficient analytics tools to generate high quality insights in pharma R&D.

Satnam Surai (CPO - satnam@aigenpulse.com) & Tobias Kloepper (CEO - tobias@aigenpulse.com)

INTRODUCTION

The deluge of biological data has seen the number of tools and software products with a component of advanced analytics, artificial intelligence and machine learning, increase dramatically in recent times. As such, (and given their nature) scientists are keen to experiment with and trial these new technologies. However, this leads to several key challenges, which we've narrowed down to three:

- Re-use
- Efficiency
- Quality

Aigenpulse provides insight into how organisations should think about addressing these three challenges to ultimately drive significant value from advanced analytics. Democratised analytics, or the widespread and effective adoption of advanced tools in the R&D pipeline, requires data to be interoperable and re-usable, scalable compute and storage to aid efficiency, and an environment for scientists to build, test, validate then productionise and archive their tools.

RE-USE

Legacy



- Data becomes siloed, irretrievable and unusable.
- Supporting meta-data and experimental parameters are unknown.
- Clinical data dictionaries are incompatible.
- Data is generated and used for single purpose.
- Scientists spend most of their time searching for data.

- Data policy should support centralised and accessible storage after generation.
- Controlled meta-data vocabularies and interoperability with LIMS/ELNs.
- Standardise clinical data dictionaries allowing for variance.
- Data policy should enable use in multiple contexts.
- Scientists spend most of their time analysing, modelling and generating insight.

Democratised



EFFICIENCY

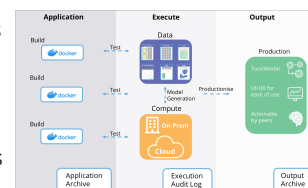
Legacy



- Research pipelines are highly fragmented.
- Analyses are manual which are time consuming and subject to human error.
- Data held and distributed in flat files on local machines.
- Analyses typically in Excel, R or Python and output in static graphics.
- Scientists spend most of their time wrangling and cleaning data.

- A unifying analytics environment for application building and execution, e.g. using Docker to allow iterative and safe tool/model building, access to data, and sufficient compute to enabling scalable execution.
- Moving new tools to production requires testing and software development (e.g. UX/UI) techniques to enable accessibility across the organisation.

Democratised



QUALITY

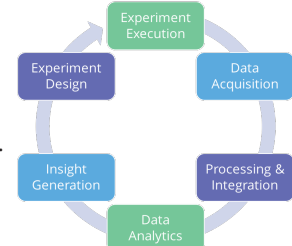
Legacy



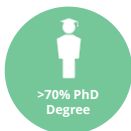
- Data is generated, processed for a specific question with little thought for generalised usage.
- Output are usually graphical and distributed in uncontrolled methods (e.g. powerpoint).
- Linking decisions to experimental datasets is lost over time.
- Rerunning analyses becomes tricky as specific data, software/tool version and parameters are not retained.
- Clinical data rarely used in R&D.

- Data policy should promote generalised usage of data.
- Audit trails for each step should be recorded - from the experiment execution to insight generation.
- A software and data deprecation strategy should be considered to ensure all outputs can be reproduced.
- Data access policies should be able to enable correct usage over clinical data in research.

Democratised



ABOUT AIGENPULSE



The Aigenpulse Platform

- Scalable, single-tenant, enterprise SaaS.
- Flexible, fully-visible licensing per named or concurrent user.
- Modular multi-omics & assay data processing and integration.
- Experimental meta-data management (ontologies and controlled vocabularies) and interoperable with LIMS, ELNs via API.
- External/public data sources (e.g. GTEx, TCGA, CPTAC, etc).
- Machine Learning out-of-the-box for non-experts.
- Full audit trails, federated and reproducible analytics.
- GxP-compliant deployments for clinical and manufacturing applications.

LET'S TALK:

aigenpulse.com
info@aigenpulse.com
+44 (0)2037 735431

