INTRO

Research, experimentation and evidence are the cornerstones upon which modern medicine is being built. Research generates important benefits for both individual patients and populations through improving our understanding of health and wellbeing and the complex factors that lead to disease.

Such research depends on access to both individual and population level data, and sharing complex datasets from sequencing, databanks and repositories, quality-assurance efforts, public health measures and patient-generated data.

Aridhia's AnalytiXagility service redefines what's possible in research analytics for biomedical and life science research, providing a secure online environment to pool, interrogate, analyse, and transform these datasets into actionable information in support of knowledge transfer from clinical research to clinical practice.

ENABLING ADVANCED HEALTHCARE RESEARCH

The fusion of the –omic sciences, systems biology, and biomedical data mining has generated an explosion of complex data, and it's growing at an exponential rate. However, while managing and analysing these datasets is key to generating improved health outcomes and reducing costs, most organisations are unable to take advantage of the insights such data holds.

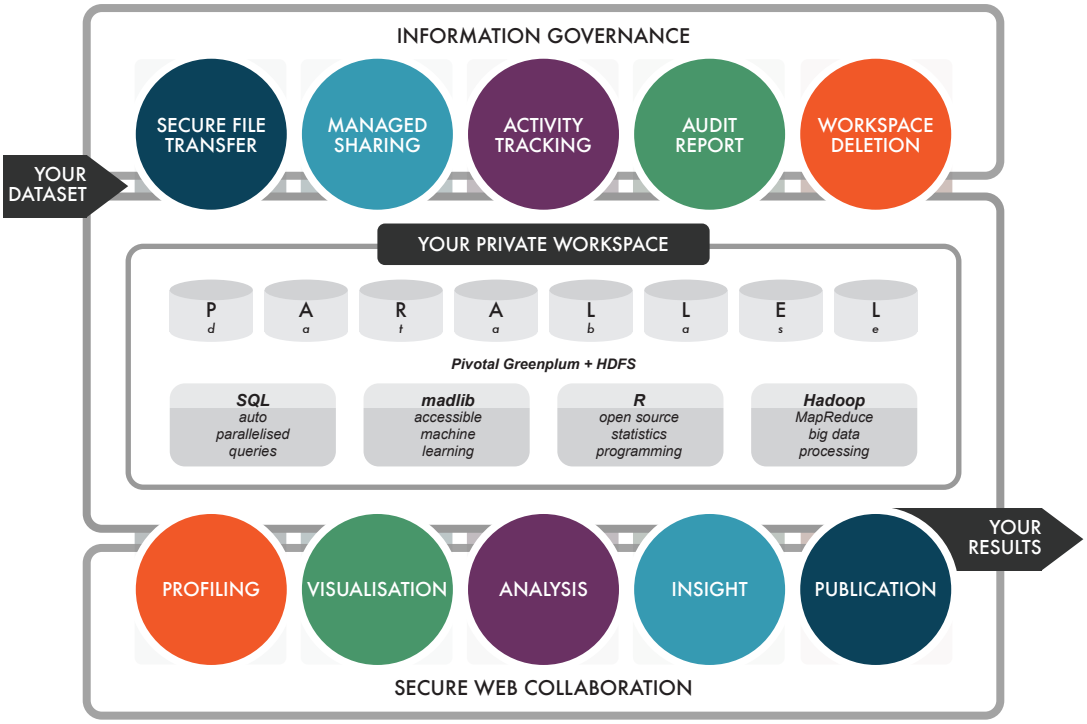
The modern research setting now relies on the utilisation of reproducible analytics derived from highly diverse data sources, many of which are public or curated/controlled by a variety of third party organisations. Within healthcare and healthcare research there are a number of observable issues in facilitating a collaborative analytics environment to make use of this data, including information governance, the adoption of shared, high-performance compute infrastructure and an environment that enables diverse domain expertise to collaborate on solving both research and operational priorities.

Aridhia's AnalytiXagility platform adopts the best architecture practice from data-rich,

consumer-grade companies such as internet businesses in terms of performance, open source and open standards, and has added the concepts of collaboration, information governance, infrastructure independence and an information commons approach to build a consumer-grade biomedical informatics platform.

THE ANALTIXAGILITY PLATFORM

Although the technology is advanced, the concept underlying our AnalytiXagility platform is simple. The web-accessible nature of the platform and its built-in analytic toolkit means that clients can log into their dedicated workspace at any time to view the available data assets, communicate with colleagues, and perform collaborative analysis in a secure multi-tenant environment.



Easy collaboration and integrated analysis

AnalytiXagility has been designed to ensure that team members can easily discuss, share, modify and create analysis in a single, secure online location. The 'owner' of a project is given access to a private workspace configured with compute, storage and analytical tools, and can invite collaborators to join the project and define the data sharing and information governance models for the particular project or study. The newly formed team have embedded social media tools within the platform to post and share updates, comments and insights.

Integrated analysis from upload to publication

An embedded R web console gives access to a vast array of cutting edge, open source, statistical and machine-learning analysis packages that enable you to quickly construct and customise advanced analytical workflows across complex and varied data types, and support the whole process from data preparation, to visualisation, analysis, and reporting through to publication.

Access to open source and international standards

From a compute and storage perspective, the emphasis is on open source and international standards (Linux, Centos, VMware for virtualisation, ITIL Capacity and Availability Management processes) and at the database, analytics and application layers on open standards and open source technologies (Postgress, SQL, Madlib, R). Wrapped around this has been the development of ETL, Data Sharing, Auditing and Collaboration functions for health and biomedical markets.

Maintain safe, secure data sharing

AnalytiXagility is built to the ISO27001 information security management model, and undergoes regular penetration testing by independent security consultants using the PCI Security Standards Toolkit (credit card payments) as a benchmark. By definition, the system is a collaborative environment, but role-based access ensures that users only have access to specific databases and other resources through partitioned workspace environments for separate research groups.

FEATURES & BENEFITS

Secure File Transfer

Users import their datasets into the AnalytiXagility platform via a secure and reliable file transfer mechanism that utilises the industry standard, SFTP. A dedicated and private SFTP server area is created as part of the configuration of a workspace for a subscriber, from where it is automatically processed and migrated into the associated workspace area.

De-identification Service

Prior to SFTP import, subscribers can elect to make use of the AnalytiXagility de-identification service which can modify incoming datasets in order to de-identify records and anonymise at the field level to ensure their dataset is stripped of personal identifiers.

Pivotal Greenplum

Greenplum is a massively parallel SQL-compliant database from Pivotal. It is based on the open source relational database PostgreSQL and has been highly optimised for processing datasets. Internally Greenplum automatically segments and distributes datasets across a cluster of compute nodes and efficiently auto-parallelises SQL queries when processing the data. Each workspace has its own associated federated area within the database in which imported datasets will be stored in their equivalent database table representation. Datasets maintained in Pivotal Greenplum are accessible by workspace users directly through writing SQL scripts, and via the R web console. As part of the security considerations the system has been created such that internal database system connection details are not exposed to workspace users.

R Web Console

R is an open source statistical programming language and is rapidly becoming the lingua franca of data science. The AnalytiXagility platform provides an embedded R programming and script execution environment embedded within its web portal interface. The user interface supports script import, editing, syntax highlighting and console based execution. The R environment is pre-configured with a wide array of packages enabling data science across a spectrum from data manipulation, visualisations including scatterplots, histograms and run charts to complex genomic data processing and annotation.

Our approach to using R advocates the dynamic document for reproducible research where a document embodies both the analysis and the code that executes the analysis. Our R web console also provides direct and simple access to the scalable computational resource within the platform for accelerated R-based analysis of big data. Subscribers are also able to request additional R packages and configurations.

MADlib

MADlib is a high performance open source implementation of a set of data profiling, mathematical, statistical and machine learning algorithms optimised for execution on data in situ within the Pivotal Greenplum database. It supports a range of machine learning and predictive modelling techniques: supervised learning (Bayes, Logistic regression, Decision Tree, Support Vector Machines), unsupervised learning (k-means clustering) as well as advanced descriptive statistics. New techniques are added in each release. The MADlib functionality is directly accessible to both SQL and R scripts.

Hadoop

Hadoop is an ecosystem of open source big data processing technologies originating from the Apache foundation. It is becoming the standard for storing large datasets of structured and unstructured data, and supports an evolving range of computing strategies. AnalytiXagility supports standard Hadoop MapReduce, a framework for reliable and performant execution of long running, large-scale data processing and Hadoop Pig, a scripting language for data processing workloads.

This facility is typically applied to extracting information unstructured or non-relational data, such as text, images or raw genomic sequencing files. That information can then be added to the database to enrich the analysis undertaken. Subscribers can obtain Hadoop as an add-on service to the base AnalytiXagility platform subscription, with additional support available in creating and running specific Hadoop-based analytics jobs.

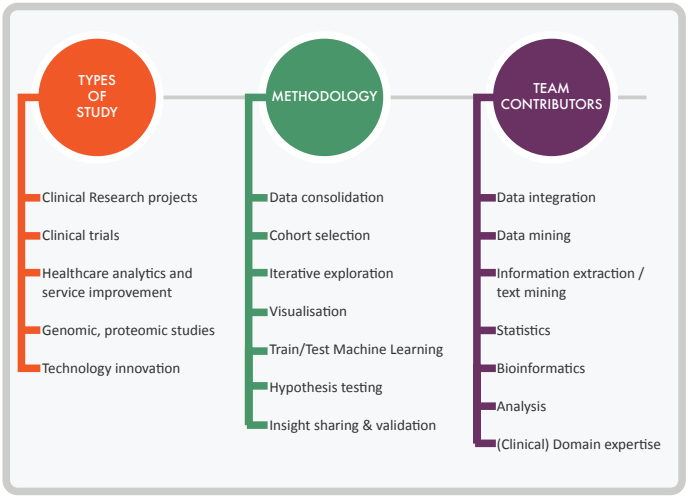
SUPPORTING MULTIDISCIPLINARY COLLABORATION

Change only occurs when groups of people work together to pool their knowledge, experience, and crucially, their data.

Within AnalytiXagility the balance between collaboration and governance can be set according to the specific circumstances of your project. Projects typically involve multiple stakeholders coming together to collaborate, often from different organisations. What's consistent is the need for diverse groups to collaborate with diverse data while ensuring that relevant governance and legislation is followed.

AnalytiXagility provides the social sharing, audit tracking, workfile alert and notification tools necessary to facilitate collaboration easily, quickly and cost-effectively, reducing time to insight and ultimately time to value – whatever domain your project lies within.

This expansive toolkit allows you to bring a multidisciplinary team together in a space which allows team members to easily discuss, share, modify and create analysis in a single, secure online location.



AnalytiXagility supports a range of study types, methodologies and skillsets to enable fast, efficient collaboration

The web-accessible nature of the platform and its built-in analytic toolkit means that they can log into their dedicated workspace at any time to view the available data assets, communicate with colleagues, and perform collaborative analysis in a secure multi-tenant environment.

The types of analysis which can be undertaken are wide-ranging; AnalytiXagility makes it possible to facilitate your analytic study needs by engaging different methodologies and the differing skillsets of team contributors to ensure that tasks are performed quickly and efficiently, productivity is increased, and cross-discipline collaboration and sharing of vital scripts and code is improved.

HOSTING ENVIRONMENT

The AnalytiXagility platform is hosted in the UK on cloud Infrastructure-as-a-Service (IaaS) provided by Skyscape Cloud Services Ltd, whose IaaS offering holds pan-Government accreditation to store data up to and including Business Impact Level 2 (IL2). The IaaS element of the service is independently certified to ISO20000 (IT Service Management), ISO27001 (Information Security Management), and ISO9001 (Quality Management) standards.

All changes to systems or configuration within the application environment are handled by Aridhia's operations team under ITIL-based change protocols.

All elements of the application and its data are securely backed up within Skyscape's infrastructure to a secondary site in the UK to facilitate disaster recovery.

TECHNICAL REQUIREMENTS

The following browsers are supported in the current release of the AnalytiXagility platform:

- Google® Chrome™ browser release 32.
- Microsoft® Internet Explorer release 10.