WHITE PAPER

Liaison's Integrated Research Collaboration and Data Analysis Platform

Executive Summary

Life sciences companies place their bets in drug discovery in the form of basic research, co-development, licensing and joint ventures. Many millions of dollars are required to take a promising formula to trial, where only a handful of the prospects will emerge. To succeed in the life sciences environment, accelerating research and development and enabling new insights hinges on the ability to collaborate with others. Providing scientists with rapid, reliable and secure access to the data and analysis tools needed to generate new insights can be the determining factor in being first to market.

Liaison ALLOY Health™ offers life sciences companies a cloud-based translational research collaboration environment that fosters an innovative approach to data sharing among different research partners. The space also provides vital applications in support of analytics and collaborative functions to explore diverse sets of omics, clinical and real world evidence data made available to the researchers. In addition, strong security, through integrated identity management controls and compliance with applicable regulatory requirements, is built into the environment.

TABLE OF CONTENTS

Objectives

The rise of precision medicine and the integration of high-throughput genomic analyses into patient care have increased the need for adequate tools for translational researchers to manage and explore data. Translational research is the new approach of translating discoveries in the laboratory into clinical therapies often described as research "from bench to bedside and back again."

Recent advances such as the sequencing of the human genome have brought mass amounts of high-content data. Translational research relies heavily on data science to extract insights and to find meaningful patterns from data. Researchers seek to effectively integrate techniques and theories from many fields to make reliable predictions. The success of the translational research requires a multidisciplinary team with wide-ranging expertise. As such, there is an urgent need for a collaboration environment that brings together dedicated people, robust processes and informatics solutions.

Liaison, building on these premises, has developed an integrated research collaboration and data analysis platform. This platform facilitates first-class, innovative and productive translational research within and across disciplines. It provides a bioinformatics infrastructure to submit, collect, share and analyze data in a secure manner.

Challenges

Translational research requires collaboration among many scientists in institutions across diverse geographic locations. The need for a high-throughput environment that enables submission, analysis, and sharing of data in a secure manner with numerous collaborators across academic or industries is a formidable challenge. Progress in bioinformatics is leading to new types of data, and the datasets are rapidly increasing in volume, resolution and diversity. Key challenges are:

- **Integration of low-dimensional clinical data with high-dimensional omics data.** The availability of high-throughput genomic analyses has increased the need for adequate tools to manage and explore these multiple scale, incongruent, incomplete and complex big healthcare datasets so that different questions can be postulated.
- **Substantial costs and the intellectual demands of managing computational complexity and data protection.** Open source analytic tools such as tranSMART require complex infrastructures (e.g. web and Java servlet servers, databases, solr, and rserve) that are most likely out of reach for the average translational clinician or researcher.

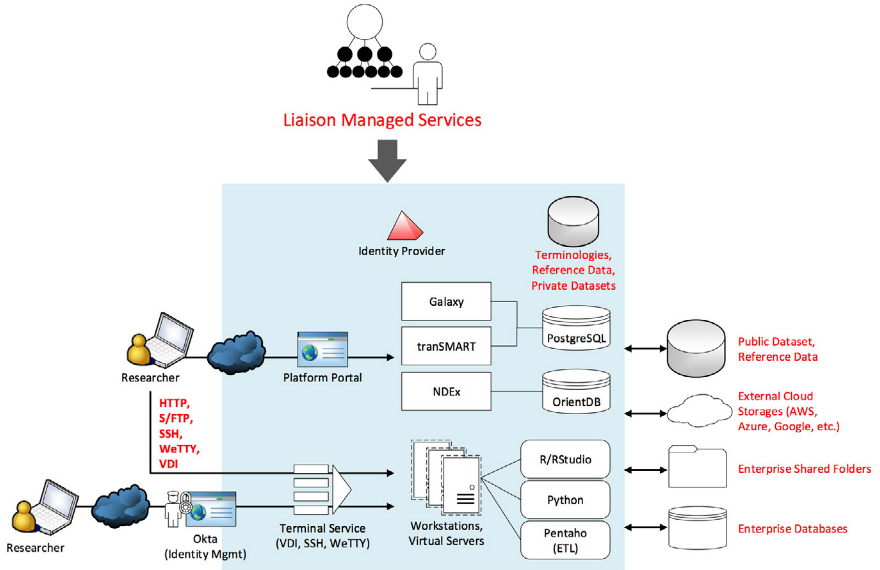
Requirements

Analyzing high-throughput omics data is a complex and compute intensive task, generally requiring numerous computerized tools and large reference datasets, tied together in successive stages of data transformation and processing. A translational research platform should be able to integrate researchers' workflows with best practice genomics analysis. Therefore, the platform should include:

- **A cloud-based infrastructure for interdisciplinary teams to collaborate within.** Liaison's translational research platform has a cloud-based computing infrastructure with high-throughput file transfer allowing clinical and omics big data to be moved efficiently in and out of the platform. There are built-in synchronization links to other public cloud storage sources ready for integrated clinical and omics data analyses.
- **An analysis framework that enables scientists to explore their data and generate hypotheses.** In addition to the built-in tranSMART analytic tools, Liaison's platform provides integration with the private Galaxy cloud for additional visualization, statistical and analysis support. More advanced users can use R, Python, and other tools inside the platform for advanced analyses.
- **Connected access to public databases and reference data.** From the Liaison cloud, researchers will have access to public repositories such as COSMIC, dbGaP, GEO, ArrayExpress, TCGA, or dbSNP, as well as public reference datasets such as GeneGo, Ingenuity, Entrez, or MeSH. Liaison has also provided standard terminologies and ontologies (e.g. SNOMED, ICD-10, RxNorm, UMLS).

Solution: Liaison ALLOY Health

Liaison's solution to the challenges of high-throughput translational research is to bring together a powerful cloud infrastructure, a set of high quality bioinformatics software, and a team of skilled personnel to manage the platform. The result is a first-class, innovative and productive translational research platform.


Liaison Managed Services
Identity Provider
Terminologies, Reference Data, Private Datasets
Galaxy
tranSMART
NDEx
PostgreSQL
OrientDB
Public Dataset, Reference Data
External Cloud Storages (AWS, Azure, Google, etc.)
Researcher
Platform Portal
HTTP, S/FTP, SSH, WeTTY, VDI
R/RStudio
Python
Pentaho (ETL)
Enterprise Shared Folders
Enterprise Databases
Researcher
Okta (Identity Mgmt)
Terminal Service (VDI, SSH, WeTTY)
Workstations, Virtual Servers

There are five distinct capabilities that set the Liaison ALLOY Health platform apart:

- **The Liaison platform is cloud based, allowing collaborative access to data across organizational boundaries.** Researchers have access to analytic tools from their desktops via web browsers, via remote desktops or via SSH remote terminals. Liaison's platform consists of Microsoft Windows desktop and Linux workstations. Researchers have access to thick-client tools such as Cytoscape and RStudio, as well as web-based tools such as tranSMART and Galaxy. For more advanced use, researchers can gain access to various ETL tools or scripting tools such as Pentaho, R and Python to curate, load or analyze data within the platform.
- **The Liaison platform provides intuitive clinical data and omics data import.** Data can be submitted via S/FTP, SCP or downloaded from cloud storages. For large datasets, it is possible to establish sync-folders to cloud storage or researchers' internal shared folders. Liaison infrastructure enables teams to define how their data will be submitted to the collaborative platform: via the web, via client tools or through a pre-defined sync-folder.
- **The Liaison platform supports delegated authentication with role-based security.** The collaborative platform is designed to run multiple research projects concurrently. Each project will involve a different team of collaborators. Liaison's platform includes an LDAP identity store. Researchers from different institutions will be able to use single sign-on (SSO) technology to log into the platform using their organization's login credentials. This delegated authentication via SSO allows researchers' identities to be implicitly established by their organizations' identity providers. Researchers who do not have such SSO capability can leverage Liaison's OKTA identity management cloud to SSO into the platform.
- **The Liaison platform delivers a set of robust processes for genomic analysis—from data collection, curation, loading, and analysis to archiving.** The platform supports a set of reproducible tools such as Galaxy, Rmarkdown/Knitr and Jupyter Notebook to allow for creating and sharing reproducible datasets.
- **The Liaison platform provides a set of translational research tools.** Overall, the platform includes the following set of tools:
    - Translational research platform: tranSMART
    - Biological network sharing: Cytoscape/NDEx
    - Big data cloud storage: SQL, NoSql, DocumentDb, NFS, etc.
    - Standard terminologies and mappings: SNOMED, RxNorm, ICD-9, ICD-10, CPT, UMLS, FDB, LOINC, MedDRA, etc.
    - Visualization, statistical and analytical tools: R/RStudio, Python, and Galaxy
    - Reproducible reports: Rmarkdown/Knitr, Jupyter
    - Remote access: Remote Desktop/VDI, SSH, and S/FTP

tranSMART

tranSMART is a knowledge management platform. It provides a framework for scientists to store and share curated phenotypic data and omics data. It enables scientists to develop and refine research hypotheses by investigating correlations between genotypic and phenotypic data, and assess their analytical results in the context of published literature and other work.

Clinicians, translational scientists and discovery biologists can explore data at various levels in tranSMART and interrogate aligned phenotype/genotype data to improve clinical trial design or to stratify disease into molecular subtypes with great efficiency.

Liaison tranSMART cloud has been enhanced and customized to support SAML 2.0 SSO with 2-tier deployment. The backend PostgreSQL and the front end App server can be independently scaled according to project sizes.

Liaison tranSMART cloud has also been enhanced with user-specific curation and ETL environments so that different project teams may be running their own curation and ETL without exposing their datasets.

tranSMART group based access control is enforced on all private datasets so that only permitted group members may view and access authorized datasets.

Cytoscape/NDEx

Cytoscape is an open source desktop application for visualizing biomolecular interaction networks with high-throughput expression data in a unified conceptual framework. NDEx is an online commons where scientists can upload, share, and publicly distribute biological networks. Researchers, for example, can enrich a biological gene-based regulatory network with more detailed transcriptional, post transcriptional and translational data, resulting in an enhanced network that more precisely models the actual biological regulatory mechanisms.

Researchers can analyze, visualize, annotate, and share biomolecular network information using these tools.

Galaxy

It is difficult to design a "one size fits all" application to integrate every available genomic data and perform every possible analysis. Many researchers are forced to manually integrate data and outputs from multiple resources. Galaxy is a web-based toolset that puts together a workbench for general data manipulation or for advanced NGS data analysis. It provides a uniform web interface, integrating multiple, independent external applications into a persistent user workspace. This serves as a workbench for researchers to share tools, workflows or collaborate on datasets.

R/Python/Java/Groovy

Scientists can run their own analytic scripts on the platform using programming environments such as R and Python to rapidly develop workflows for statistical inference, regression, network analysis, machine learning and visualization at all stages of a project from data generation to publication.

RStudio/Knitr and Jupyter Notebook tools are available for researchers to compile reproducible analytic packages.

Auto User Provisioning

When building a shared collaboration platform for use by multiple project teams, security is essential. A typical translational research team consists of 3-4 organizations from the pharmaceutical company and its external academia research partners. When provisioning new users to the platform, it is best practice to physically verify these new users before account credentials can be provided. This, however, is impractical. Instead, the Liaison platform relies on SAML SSO technology to delegate user authentication to the pharmaceutical company's internal identity provider.

When a user first logs into the platform with a trusted SAML SSO token, user accounts for the tranSMART and other applications are dynamically provisioned based on the "security assertions" provided in the SAML token. This "just-in-time (JIT)" provisioning enables us to provision new researchers as soon as they join the team without administrative overhead and without delay.

Conclusion

The advance of high-throughput and big data technologies and the dissemination of EHRs has lead to the explosion of omics and clinical data available for researchers. The exploration of big data for big science and the workflow to orchestrate the research process requires highly sophisticated tools and methods that are complex to deploy and costly to maintain. Liaison's cloud-based research collaboration platform, together with its managed services model, offers researchers in various settings a dynamic and cost effective environment in which to accelerate clinical research. The data driven platform addresses the desire of a functionally rich environment providing big data infrastructure, data integration, data curation, data harmonization, exploration tools support, secure access, and regulatory compliance. Most importantly, the Liaison platform relieves researchers from mundane data janitorial labor and focuses their valuable time on data insights.

About Us / Liaison Technologies is a recognized market leader in providing cloud-enabled data integration and data management solutions to break down complex information barriers, enabling enterprises to make smarter decisions, faster. Liaison's tailored offerings are designed to address the complex data needs of today's enterprise while laying the foundation to solve unforeseen challenges of tomorrow. Founded in 2000, Liaison serves over 7,000 customers in 46 countries with offices in the United States, Finland, Netherlands, Sweden and the United Kingdom. For more information, visit www.liaison.com.

Contact Us / Liaison Technologies / info@liaison.com / 1-866-336-7378

www.liaison.com