

A unified Science Data Fabric for LifeArc : Benefits, Risks, and Considerations



Gintarė Bučaitė, PhD
Interview seminar, LifeArc
Data Engineer (Systems Engineer) – Data Science
2023 June

LifeArc and its current data challenges

LifeArc exists to help transform early stage, promising science into medical interventions to improve human health and benefit society.

Challenge: Growth of LifeArc has led to internal and external *data silos* limiting the availability of experimental data.



Data silos lead to a plethora of problems:

- **Inefficient data access and collaboration:** challenging for teams to access and collaborate on critical data, leading to inefficiencies, duplicated efforts, and delays in decision-making.
- **Reduced data visibility and insights:** without a unified view, valuable insights and patterns that could drive innovation and strategic decision-making remain hidden, limiting the organization's competitive advantage
- **Hindered data integration and analysis:** siloed data complicates the process of integrating and analyzing information from different sources as it requires significant effort and resources to transform and reconcile data
- **Impaired data quality and consistency:** this inconsistency can lead to data inaccuracies, redundancies, and conflicts, reducing the trustworthiness of the information and hindering effective data-driven decision-making
- **Increased operational costs and IT complexity:** managing multiple data silos requires maintaining separate infrastructure, data storage systems, and tools for each isolated data source

Unified Science Data Fabric to break Data Silos

Challenge: Growth of LifeArc has led to internal and external *data silos* limiting the availability of experimental data.

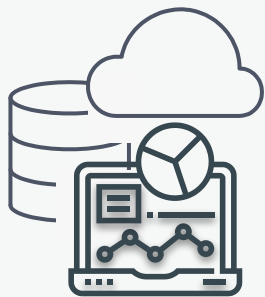
Solution: Data Fabric approach operating on FAIR (Findable, Accessible, Interoperable, and Reusable) principles



A Science Data Fabric is a *unified platform* or framework that integrates and harmonizes diverse scientific and experimental data from multiple sources and formats.

It operates on FAIR data principles, enabling *seamless data discovery, integration, and collaboration* while maintaining data integrity and security.

Potential Benefits of a Science Data Fabric



Benefits of Science Data Fabric approach

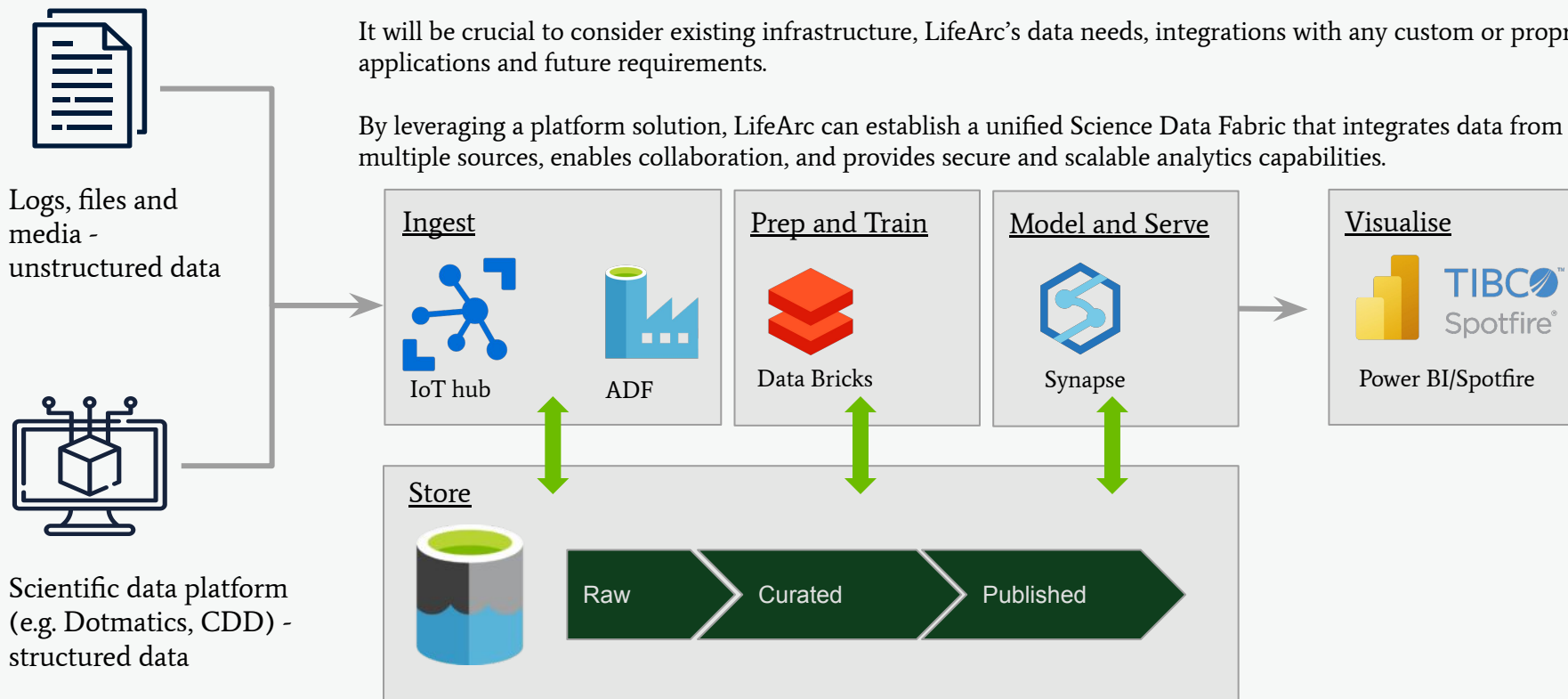
- **Enhanced *data accessibility* and discoverability:** provides a centralized, unified platform where data from multiple sources and formats can be easily discovered and accessed
- **Improved *collaboration* and knowledge sharing:** fosters an environment where data can be shared, reused, and built upon, enabling interdisciplinary collaborations, accelerating research, and generating novel insights.
- ***Accelerated innovation* and time-to-discovery:** scientists can leverage a broader range of data for analysis, hypothesis testing, and experimental design, ultimately accelerating scientific breakthroughs
- **Streamlined *data integration* and analysis:** simplifies the process of integrating and analyzing data from disparate sources
- **Robust *data governance* and security:** clear metadata standards, data stewardship roles, and access controls can be established to maintain data integrity, privacy, and comply with regulatory requirements

Science Data Fabric for LifeArc Data

Without further background information about the data structures and needs at LifeArc available, the initial proposal to consider could be a traditional data fabric implementation.

It will be crucial to consider existing infrastructure, LifeArc's data needs, integrations with any custom or proprietary applications and future requirements.

By leveraging a platform solution, LifeArc can establish a unified Science Data Fabric that integrates data from multiple sources, enables collaboration, and provides secure and scalable analytics capabilities.



Potential risks of implementing Science Data Fabric



Data Privacy and Security

Risk: Data breaches and unauthorized access

Mitigation: Strong access controls, encryption, regular security audits, compliance with data protection regulations.



Data Quality and Integrity

Risk: Inaccurate, incomplete or inconsistent data

Mitigation: Data validation, automated cleansing, quality assessments, data stewardship.



Integration and Interoperability

Risk: Challenges in integrating diverse data sources

Mitigation: Robust data integration processes, standardized formats, data governance practices



Technical Complexity and Scalability

Risk: Complexity in implementing and maintaining the Science Data Fabric

Mitigation: Training and skill development, cloud consultants or partners, scalable cloud infrastructure



Change Management; User Adoption

Risk: Resistance to new workflows and processes

Mitigation: Change management activities, stakeholder involvement, user-friendly interfaces and tools

Science Data Fabric for LifeArc: Conclusion

Embracing a Science Data Fabric operating on FAIR principles enables LifeArc to **overcome data silos** and harness the full potential of its experimental and collaborator-generated data.

- The Science Data Fabric promotes *data accessibility*, *collaboration*, and *innovation*, leading to faster discoveries and insights.
- By leveraging robust and secure platforms, such as Azure, LifeArc can *automate* data ingestion, integrate diverse data sources, and ensure data governance and security.
- *Mitigating risks* such as data privacy, quality, integration challenges, technical complexity, and user adoption is crucial for successful implementation.
- Adopting a Science Data Fabric empowers LifeArc to make *data-driven decisions*, accelerate research, and enhance scientific collaborations.

Q & A

...

Gintarė Bučaitė, PhD
Interview seminar, LifeArc
Data Engineer (Systems Engineer) – Data Science
2023 June

