

Do You Know Where Your Data Is?

If You Have Storage Silos, Probably Not.

In a recent Global Life Sciences Outlook study, analysts at Deloitte identified several challenges confronting the management of HPC and research data in the coming years. Topping the list were tech issues: specifically, the challenge of building and ensuring data integrity in order to maximize data value.

The Problem of Data Silos

Higher Education and Research institutions have been embracing information technology change and innovation in order to accelerate the acquisition of data and to drive advances in their areas of research. One consequence of this dynamic is the tendency for research and information to become trapped in technology silos that make data difficult to share or to leverage across research groups. Companies also confront the high cost of inefficient storage and the tendency of groups to isolate data on storage platforms and media that are not easily shared at all.

In addition to infrastructure management and data silo'ing, complying with the data integrity requirements of regulatory bodies remains elusive. Researchers need to provide data that is complete, consistent and accurate throughout its useful life – and to guarantee that this is the case to auditors and regulators. That guarantee must extend to all data, whether stored on premises at a research facility or migrated to public cloud storage services, and regardless of what technology is used to store the data.

In the face of these challenges, HPC and Research organizations need an approach to data management that facilitates authorized data access and efficient data sharing irrespective of location and storage technology. According to Deloitte, research organizations would profit from such an intelligent data management capability in three ways:

- 1. Improved inter-group collaboration:** With an intelligent data management capability, storage infrastructure (even of different types and multiple vendors) becomes a shared resource, so different groups can collaborate more effectively. Using storage as a shared resource by all measures is better than the practice of constructing separate storage silos for every research group and competing for

Research data often becomes trapped in technology silos that make it difficult to manage and collaborate across groups.

scarce resource budgets. Such a shared data storage infrastructure might also set the stage for the creation of cross-functional teams whose missions would be to leverage shared data to benefit the organizations, resulting in new outcomes, or better compliance strategies or more efficient supply chain or marketing processes.

- 2. Maximized data value:** Researchers are looking for some means to implement end-to-end (E2E) evidence management approaches that will enable the unification of data and processes from research and initial development stages through to publication and possibly commercialization. Such a strategy requires that data be properly classified, and annotated as through its life cycle.

This involves the application of metadata in a consistent and standards-based manner. Value is often non-intuitive, driving the need for analytics to understand relationships. By automating data classification, and reducing the reliance on researchers to remember to annotate their files, this helps ensure that data does not get orphaned. A simple metadata tag or combination of values that associates files with a project or a department means data value is not lost or forgotten when the researcher moves on to other things.

- 3. Effective Data Stewardship throughout the E2E Life Cycle:** Says Deloitte, data must also be provided with appropriate protection, preservation and privacy services. Given the volume of data being generated by research and trials, and the changing service requirements of the data over its useful life, surmounting this challenge requires some means to automate the management of the data/services relationship over time so that the hosting of data and its access policies will remain consistent with the organization's governance, risk and compliance posture, and consistent with the changing regulatory mandate.

The above challenges are complex and often off-putting to both technical and business planners in Research and HPC environments of all kinds – a reason why traditional silo'ed approaches, rather than enterprise level approaches, continue to dominate the industry, according to analysts. The good news is that technology for improving data collaboration, maximizing data value, and automating data stewardship is available today: StrongLink from StrongBox Data Solutions.

StrongLink is Strong Medicine

StrongLink is an automated data management technology that can be leveraged to provide an enterprise-wide data management solution, regardless of data hosting silos, storage topologies or media types, or physical locations. In the context of Deloitte's delineation of research data management requirements, StrongLink can deliver:

- ✔ **A Data Collaboration Engine:** Existing IT services and infrastructure, even silo'ed infrastructure, do not need to be dismantled or upgraded to achieve automated data management with StrongLink. StrongLink can manage data regardless of how it is platformed and include it in a common Global Namespace that can be used by cross-functional teams to conduct joint research and analysis activities, and to advise organizational decision-makers. Collaborative data use sets the stage for collaborative products and services.
- ✔ **A Data Value Driver:** StrongLink's powerful metadata management capabilities enables the data generated in different research groups to be shared with detailed descriptors of the context and content of each file or object. By applying management metadata schemas across disparate data sets and maintaining data provenance throughout its useful life, StrongLink can ensure that related data can be queried, discovered and cross-referenced to facilitate the integration of data regardless of its point of origin. This powerful integration of data sets the stage for big data analytics and the realization of greater value from the data that is being generated within the organization.
- ✔ **A Stewardship Guarantor:** StrongLink's file and object metadata management capabilities and policy-driven management functionality can be used both to define and to implement the organization's governance-risk-compliance strategies by automating the manner in which data is hosted, protected, preserved and kept private. Data management policies can be updated and applied dynamically whenever regulatory mandates dictate and whenever infrastructure changes occur, and what's more, StrongLink can automate the movement of data across infrastructure and the copying of data for purposes of production. If desired, these workflows can be automated in such a way as to minimize or eliminate the involvement of research project personnel.

StrongLink is Available Now

The time to unlock the value of your data is now. The next generation of E2E data management for research, HPC and other similar use cases can be realized today – using StrongLink from StrongBox Data Solutions.

For more on this software that can revolutionize how you control and manage your research data, check us out at www.stronglink.com