astrix

Emerging Technologies Impacting a Digital Transformation of Life Sciences

(Part 1)

Executive Summary

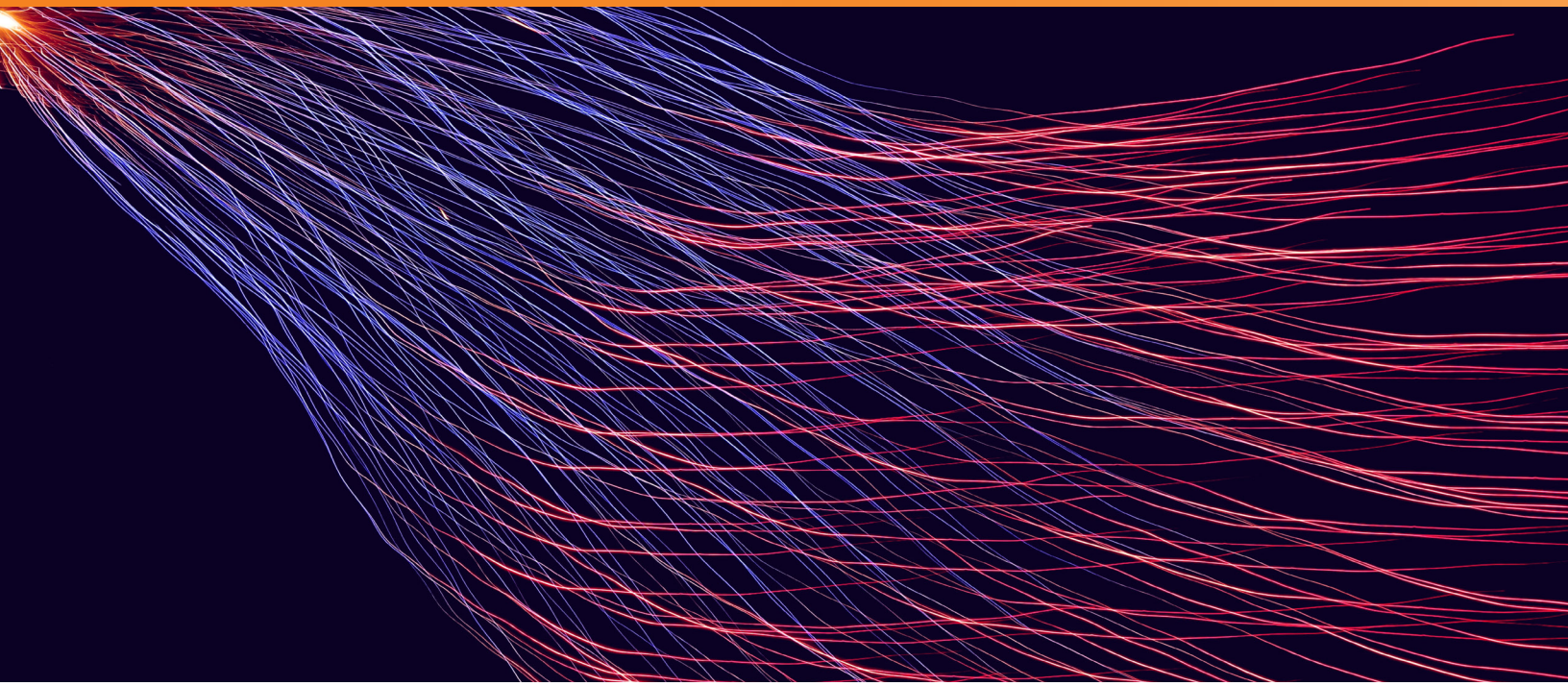
Many biopharmaceutical companies have increased their emphasis on innovative technologies and data sciences to accelerate a digital transformation across R&D.

Underlying successful digital transformations, is the critical need for R&D organizations to leverage information assets to enable enhanced insights, decision making, and process automation.

As part of the planning for digital readiness, emerging technologies must be considered in concert with existing infrastructure to optimize the technology footprint, increase operational effectiveness, encourage data findability, and reuse, and facilitate agile, harmonized, and optimized workflows to advance the pipeline, increase collaboration, enable rapid decision making, and accelerate time to market.

In this white paper, we discuss 5 important emerging technologies that are impacting a digital transformation in the life sciences industry.

1. **Enterprise Data Fabric**
2. **Hyperautomation**
3. **Big Data**
4. **Structured Content Authoring**
5. **Advanced Analytics and Simulation**



1. Enterprise Data Fabric

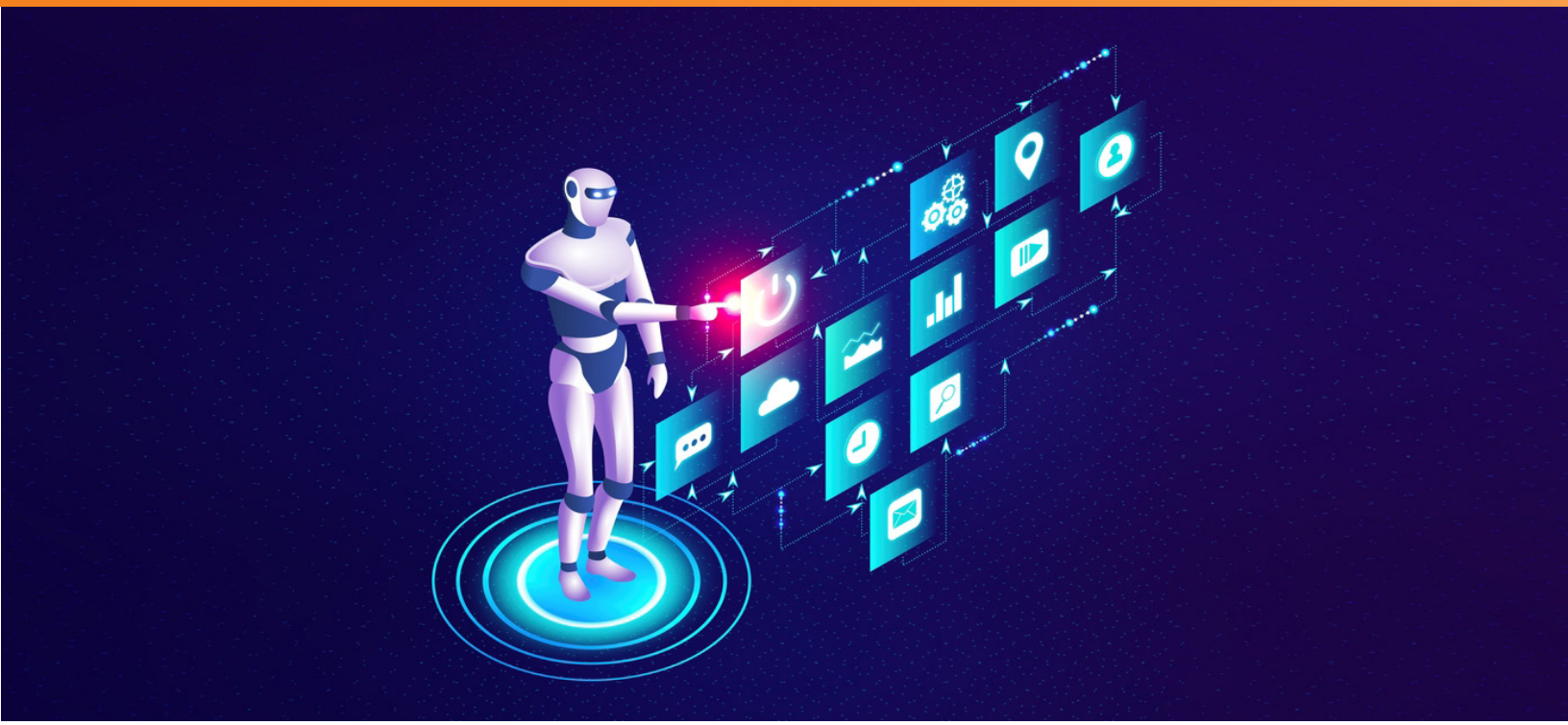
> *By 2024, data fabric deployments will quadruple efficiency in data utilization while cutting human-driven data management tasks in half.*[1]

Data Fabric is a hot topic buzzword today. We think of Data Fabric as the way of creating an abstraction from the underlying technology. Allowing the data to be exposed at the top level as business objects that are semantically related. This can be thought of as a weaving together of data from internal silos and external sources that creates a network of information to power the business' applications, AI, and analytics. The underlying plumbing allows for assembling and integrating data in somewhat intelligent and flexible ways to allow the integration of new data sources and to find the most efficient pipelines to deliver that data.

Data Fabrics includes ML to identify and connect business, technical and operational metadata. They include recommendation engines that help to identify the most efficient data pipelines for delivering new data as well as dynamic semantic modeling and data catalogs to minimize manual processing and enable rapid intelligent characterization of data while supplying coherent graphs for navigation across sources.

The following are some important components that need to be considered with respect to Data Fabric in the life sciences industry:

- **Repository Technologies** -Repository technologies, ingestion, and curation capabilities will need to keep pace with new and emerging data types, data collection and transmission methods, and the volume and speed at which R&D and real-world data is acquired. There will be a need to leverage a mixture of storage modalities (e.g., traditional data warehouses, marts, multi-modal hubs) to handle the breadth of data and query patterns, while accommodating specialty scientific, technology and emerging standards (e.g., semantics, graph DBs, blockchain for genomes and eHR)
- **Ingestion Archetypes** - Establishing ingestion archetypes (e.g., batch, streaming, data virtualization, ETL, etc.), underpinned by machine learning technologies to enable rapid data characterization and minimize the need for traditional labor-intensive model, map and move approaches. Using graph technology and semantic linked data models to explicitly encode context with the data, enabling inferential queries, while minimizing the technical debt that is typically incurred when new data sources are identified (industry leaders typically use graph databases to accomplish)
- **Data Federation** - Defining a data federation strategy, outlining when data will be moved to controlled environments, cloud or otherwise, or when to federate or keep data at the source. Leveraging data virtualization capabilities to access and integrate data at the source when dictated by business, regulatory, or technical requirements; virtualization services provide agility for exploration of data and relationships.
- **Data Marketplaces** - Catalogs power search and discovery through Google-like interfaces, enabling contextual navigation and retrieval (e.g., disease and molecular context). Data are automatically tagged, classified, cataloged, and integrated into dynamic data models, helping users to make sense of large or complex sets of data.



2. Hyperautomation

Hyperautomation is a key industry trend that will impact the Life Sciences industry. Hyperautomation is the application of traditional automation technologies augmented with artificial intelligence and machine learning. This automation has a significant impact on the full gamut of healthcare operations, including patient data management, claims processing, customer service, and patient-HCP interaction.

> *By 2022, the Hyperautomation market, according to Gartner, will be worth $596 billion[2] and by 2025, 70% of organizations will implement operationalized AI architecture[3]*

Hyperautomation involves the automation of processes both within and across functional systems using traditional automation tools such as BPM (business process management), BPA (business process automation), RPA (robotic process automation), and low code/no code development tools to automate and orchestrate end to end processes. These traditional tools are augmented with AI and ML tools that enable capabilities such as natural language processing and generation, machine vision, and intelligent assistants at the data processing end of the spectrum to the use of predictive analytics and probabilistic logic that may suggest next best steps, at the decision-making end of the spectrum.

In essence, Hyperautomation assists organizations in systematizing processes across the business, increases agility and optimizes workflows. It enables life science organizations to produce products more rapidly by leveraging intelligent insights, improving consistency, and improved human decision making. It also reduces costs by replacing manual tasks with automation.[1]

- The top business challenges that organizations face today that Hyperautomation can assist in fixing, are:
- Decision making is often made with partial information.
- Siloed legacy systems with internal workflows that are primarily manual in nature.
- Non-transparent processes for patients or HCPs, and non-compliance with the Audit Trail.
- Grueling, repetitive work that requires many worker hours and raises operating costs
- Manual processes that are prone to error that present an ongoing operational and regulatory problem.

With Hyperautomation, the AI/ML needs to be definable and trainable by end users. Automation is universally applied through workflow orchestration and process/task automation tools. The intelligent automation is adaptive to user role and task context. It leverages probability and logic to deal with uncertain situations.

3. Big Data

In Life Sciences we have data that we would classify as "Big Data". That being high in volume, velocity, or variety. This Big Data is fundamentally transforming the industry.

Some key examples of the impact in Life Sciences can be found with:

Genomics data

By 2025 genomics data (NGS and HiFi) will exceed the data totals of YouTube, Twitter and astronomy combined.[4]

Variety of sources of clinical data from streaming and mobile apps will continue to grow.

Our research indicates that there will be a significant increase in the R&D relevant data through social media, mobile apps, biosensors, multi-omics, imaging, single cell analysis, and real-world evidence.

Increase in high content analysis techniques being used in Research

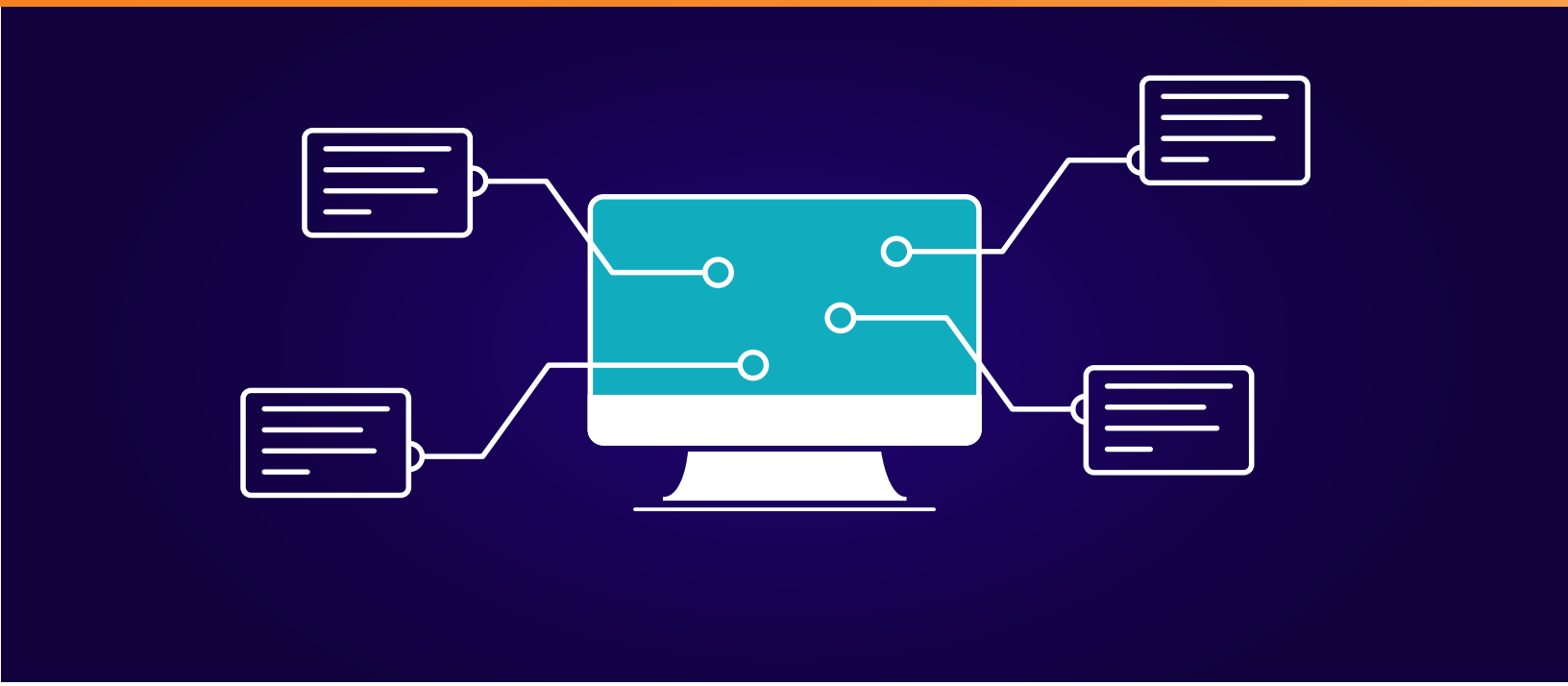
e.g., Imaging and single cell analysis

> *By 2025 genomics data (NGS and HiFi) will exceed the data totals of YouTube, Twitter and astronomy combined.4*

The explosion of health-relevant data enables new opportunity for R&D transformation; however, tools are needed to store, manage, and analyze this data. The following are key considerations:

- **Cloud & hybrid cloud architecture is needed** - to facilitate performant access and analysis of data and fit for data purpose storage modalities and formats that can best serve the analytics engines.
- **Clear data use, storage, and archiving strategies are required** – a strategy and roadmap to understand how data is being utilized throughout the organization and where it is stored, along with when and how it should be archived are requirements in the life sciences industry.
- **Strategies for data management are needed and include:**
    - Support for a variety of **data ingestion patterns**.
    - **Continuous prospective data curation** and data harmonization (to standardize vocabularies/ontologies and structure).
    - **Application of AI/ML capabilities** to data processing to increase speed and minimize human interaction.
    - **Parallel and distributed data processing engines** to process (e.g., Spark).
    - **Edge devices and systems** that push more of the data processing to this method.
- **Unique Storage requirements for big data**
    - Big data will require exabytes of space and data usage, partnering, and storage strategy with storage tiers at multiple price points coupled with effective compression technologies
    - A mix of on-premise and cloud-based storage with advanced and highly automated data processing (edge and distributed processing) and management capabilities are needed to ensure frictionless access and use of data, while ensuring compliance
    - There is a requirement for a data platform with fit for purpose storage modalities and formats with a clear data use, storage, and partnering strategy

- **Data Strategies for raw vs. cataloged vs. curated vs. data product (Insight)**
  - Distinct management strategies for core data vs. highly curated data are required as a best practice.
  - Human and machine-searchable indexes of data resources are necessary.
  - Proactive service to automate and manage updates, access, context, and indexes of data, using AI/ML capabilities to scan new data availability and consistently enrich key data as available.
- **Cleaning and the context of the data**
  - R&D data exists in a wide variety of formats and vocabularies, requiring extensive cleaning for utility and interoperability. There is a need to adopt well-established ontologies and industry standards (SEND, CDISC) where available and leverage ML technologies to characterize/classify data.
  - Automate and clean data, establish data context and deliver data through services and secure collaboration environments.
  - Requires R&D to establish and govern data and data standards.



4. Structured Content Authoring

Structured Content Authoring (SCA) is a standardized method where technical content is controlled by pre-defined guidelines. The technical writer marks up the content according to what it semantically represents and agreed guidelines. The writer is not concerned with editing styling or formatting. The idea is to develop discoverable, reusable, adaptable, reconfigurable templates for storing, tagging, standardizing, and translating critical content. The content should be auditable, publishable, individually authored, and approved. Moreover, it can also be used to automatically generate standardized documents from digital records and approved, reusable content.[5]

The Value of SCA

By creating and managing standardized and approved content entities that are ascertainable, compact, and reusable across the organization, it enables:

- Single source of truth – one standard copy of the information to be used across the organization
- Increased compliance and visibility/traceability of where the content is being used (even slightly different versions)
- Improved organization of content
- Decreased transcription errors through automated document generation and translation.
- Content Libraries and templates that can speed authoring through supporting technology such as:
    - Component Authoring tools
    - Smart contracts
    - Natural Language Generation
- Machine Learning closed-loop feedback to continually improve documentation, submissions, communications, and code - creating a self-improving ecosystem.

SCA Considerations

There is a shift from using documents to record scientific insights to viewing the content itself (layered intelligence, context, and knowledge within) as being the critical component. This is driving a move from document creation and management to development of core content. This allows for the assembly of content across sources to construct documents, publications, dossiers, and reports.

Emerging standards and open architectures,

- Darwin Information Typing Architecture (DITA), enables organizations to deploy tools to rapidly author and organize information into hierarchical, modular, and reusable objects. This allows for better compliance and traceability of approved and standardized information.[6]

Federated Content Management

- Gartner, and other tech consulting companies, advocate for content management strategies in recognition of the diversity of data resources in an enterprise ecosystem.

Document Repository Technologies

- Emerging technology brings emerging standards. Implementation of industry standards, such as Darwin Information Typing Architecture (DITA) will require a strategic position and evaluation of tools, either open source through DITA-OT or other commercially available solutions, to support these protocols.
- Content Management Systems (CMS) enable curation and storage of controlled documents and other standardized content. CMS software includes audit trails and workflow automation in order to enable template management, collaborative authoring, review steps, facilitate the approval of new content, maintain traceability, and uphold compliance with regulatory requirements.
- This requires environments to enable autonomous determination of relevant, reusable content, leading to intelligent authoring of documents, publications, and presentations.

Content Object Governance

- Disparity in lexicon must be alleviated by determining standard ontologies and establishing agreed dictionaries, standard terms, and templates.
- A single source of truth in current content must be determined to enable simplification and consolidation of information that will facilitate future use. This approved content should consist of the smallest, reusable pieces to enable rapid impact assessment and update of new documents.
    - Implementing new capabilities for tracking review and approval progress, as well as content object circulation.
    - Enact updates centrally (distributing changes as appropriate to all versions by embedded logic), based on a persistent review of outcomes and content effectiveness, of content and code.

5. Advanced Analytics and Simulation

Advanced analytics provides the foundation for automating decisions by applying next generation methods to solve business problems. The methods utilized produce quantitative, sophisticated insights that are unlikely to be obtained through traditional approaches like business intelligence (BI). Diagnostic, predictive, and prescriptive analyses leverage artificial intelligence techniques, such as machine learning to make sense of and learn from data.[7]

Through the traditional techniques of Descriptive Analytics and BI, we develop insights through reviewing reports and dashboards of historical data and trends that show us what has happened. With Advanced Analytics we leverage data science and machine learning technologies to analyze and model systems for simulation to surface root causes (Diagnostic Analytics) and predict (Predictive Analytics) and prescribe (Prescriptive Analytics) future outcomes or actions.

All are beneficial to the business for different reasons. Descriptive Analytics can drive new insights from already available data whereas Predictive Analytics can allow organizations to understand trends in the data and how those trends influence outcomes. Prescriptive Analytics can enable organizations to ask and answer the question "what do we do next?".

But this is just the surface, there are lots of changes happening in the market as companies bring traditional and advanced analytics together, organizing information in new and interesting ways

Descriptive Analytics *can drive new insights from already available data whereas* ***Predictive Analytics*** *can allow organizations to understand trends in the data and how those trends influence outcomes.*

One exciting way of leveraging these principles is through Simulation.

So, what is Simulation?

Advanced analytics identifies patterns, creates change scenarios, makes predictions, and proposes actions based on outcomes. Simulation provides a way to examine alternative outcomes and scenarios before, during, and after executing a plan.

If the outcome of an analysis leveraging advanced analytics points to the need to make changes, how that change is implemented is just as important as the analytics. Simulation provides a way to test adjustments to the system and to determine future impacts.

Both Advanced Analytics and Simulation are important in assisting Life Sciences organizations and the **Benefits of Leveraging Advanced Analytics in the Life Sciences Industry allow organizations to**:

1. Accelerate Drug Discovery and Development
2. Increase The Efficacy of Clinical Trials
3. Personalize & Create Targeted Medications
4. Reduce Cost and Increase Drug Utilization
5. Leverage Social & Search Engine Listening to Capture Data of Interest
6. Drive Effective Sales & Marketing Operations
7. Streamline Compliance
8. Improve Operations & Employee Training[8]

Trends regarding Advanced Analytics & Simulation

The following are the key trends driving emerging technologies across the industry:

- There is a strong desire to leverage legacy data, external and real-world data to drive increasingly advanced predictive analytics, models, and simulation activities, such as digital twins, trial simulations, signal surveillance, etc.
- A shift in thinking and technology to support:
    - Federated Learning (i.e., training models across decentralized systems to eliminate the need to bring raw data in-house)
    - Democratization of data sciences (i.e., creating citizen data scientists)
    - Automation of AI/ML model development, selection, training, parameter tuning, and deployment
    - Insight engines beyond just enterprise search to enable users to tap into previous outputs and decisions rapidly

Key Considerations when utilizing Advanced Analytics & Simulation

- Build predictive and prescriptive modeling capabilities with big data analytics powered by AI/ML across R&D.
- Advanced Analytics Infrastructure
    - Develop platforms and systems to enable an advanced analytics model
    - Setup data intake and data management architecture to streamline data availability and processing capabilities for the advanced analytics
    - Implement self-service analytical tools and automate the preparation and provisioning of data
    - Partner for federated modeling capabilities as well as for platforms and technologies
- Knowledge Management and Governance
    - Build skillsets in IT and data science, as well as core R&D functions, to increase AI/ML knowledge
    - Use insight engines to surface new information from already available data
    - Enable more users to discover and explore insights to help drive decision making with greater efficiency
    - Build data governance models to ensure 'FAIR' data to enable access, reuse, AI and/or advanced simulations from transactional systems, data warehouses or data lakes.
    - Leverage AI/ML technology to automate analytical tasks, such as
        - Data preparation
        - Insight generation
        - Analytical model development
        - Surfacing data to end users

Conclusion

Emerging technologies will play an important role in the future of the life sciences industry. It is therefore critical to understand these technologies and what they can bring to the organization. In this white paper we have discussed 5 important emerging technologies that the organization needs to review. We will continue to provide additional insight into other emerging technologies that can assist the industry.

Those organizations that optimize these technologies will ensure smooth operations and patient safety. With the right internal and external support along with a focused strategy and processes, the organization is prepared for future growth.

> *To achieve success with these emerging technologies, the business needs to ensure they have a solid strategy in place and a comprehensive plan specifically designed to transform end-to-end workflows and data flows, identify technology opportunities, and define user requirements.*

Moreover, it is imperative that the organization assess its current technologies, creating thoughtful platform-based and data-centric strategies with an emphasis on digital readiness and smart movement to the cloud, while leveraging these emerging technologies to support the business direction.

Additionally, it is critical that there are pragmatic strategies and frameworks for data governance, assessments of data quality, and planning for optimized management of digital assets across the organization.

Astrix's Strategic Consulting Services practice can assist organizations with optimizing their organization in this way. Leveraging years of experience and studying other organizations similar needs, Astrix can assist in an optimized digital transformation.

About Astrix

Astrix is the unrivaled market-leader in creating & delivering innovative strategies, solutions, and people to the life science community. Through world class people, process, and technology, Astrix works with clients to fundamentally improve business & scientific outcomes and the quality of life everywhere. Founded by scientists to solve the unique challenges of the life science community, Astrix offers a growing array of strategic, technical, and staffing services designed to deliver value to clients across their organizations

References

1. Gartner Report, "Top Strategic Technology Trends for 2022: Data Fabric", Mark Beyer, Ehtisham Zaidi, Guideo De Simoni, October 18, 2021.
2. ACL Digital, "Hyperautomation in Healthcare for Quality and Compliance", August, 11, 2021
3. Gartner, "The 4 Trends that Prevail on the Gartner Hype Cycle for AI", 2021, September 22, 2021
4. NIH STRATEGIC PLAN FOR DATA SCIENCE, 2022
5. Paligo, "About Structured Authoring"
6. Techwirl, "What is DITA"
7. Gartner, "What is Data and Analytics"
8. Polestar, "8 Ways Pharmaceutical Companies Ensure Success with Analytics", August, 15, 2019

astrix

Corporate Headquarters
125 Half Mile Rd Suite 200
Red Bank, NJ 07701 USA

732-661-0400
info@astrixinc.com

For more information, visit:
www.astrixinc.com/pharmacovigilance-services