SciBite from Elsevier

Whitepaper

Enablement of Effective AI

A Practical Guide to Getting Data “AI-Ready”

SciBite from Elsevier

Enablement of Effective AI

A Practical Guide to Getting Data "AI-Ready"

Overview

Enterprise-wide structured and FAIR data is foundational to realizing the value of data assets within life science organizations. Establishing a set of data standards and deploying these across an enterprise takes planning, and there are several considerations in planning and implementing enterprise FAIR. Here, we articulate the benefits of successful implementation and the technical and operational considerations.

Introduction

According to Morgan Stanley, Artificial Intelligence (AI) has the potential to revolutionize and speed up drug discovery, creating a $50 billion industry over the next decade[1]. This is reinforced by recent Pistoia Alliance surveys, which highlighted both a strong belief in the potential of AI to expedite R&D cycles[2] and that it is set to be the number one technology investment for Life Sciences companies over the next two years[3]. Similarly, the 2024 GlobalData report puts AI as the most impactful technology for pharma for the fifth year running[4].

The rapid rise in Generative AI, or 'GenAI' – enabling computers to generate new content – has further fueled excitement for the possibilities that AI can bring to the industry. The McKinsey Global Institute (MGI) has estimated that GenAI could generate $60 billion to $110 billion a year in economic value for the pharma and medical-product industries, potentially streamlining every aspect of the pharma value chain[5]. It is important to be realistic, though. In highly regulated industries like pharmaceuticals that rely on evidence and traceability, the progress of GenAI is incremental. The practical implementation and value of GenAI currently lags far behind the hype cycle. Therefore, this optimism is tempered with caution – Gartner's Hype Cycle positions GenAI at the 'Peak of Inflated Expectations,' about to enter the 'Trough of Disillusionment'[6].

McKinsey also counters their prediction by stating that GenAI cannot deliver results unless a proper data architecture is in place[7].

Despite significant technological advances, poor data quality poses a significant barrier to the widespread use of machine learning (ML) and AI. The decades-old axiom 'garbage in, garbage out' still holds true – good quality data is essential to ensuring that ML and AI give accurate and true outputs. As the title of a 2018 Havard Business Review article puts it: 'If your data is bad, your machine learning tools are useless'[8]. As noted within the same article, poor data quality is a two-fold issue when it comes to ML – first, in the historical data used to train the model, and second, in the new data used by that model to make future decisions. The experiences of pharma industry leaders reinforce this. For example, according to Novartis CEO Vas Narasimhan, "The first thing we've learned is the importance of having outstanding data to actually base your ML on"[9]. Well-structured, standardized data is a prerequisite for the adoption of the latest cutting-edge AI technologies.

ChatGPT has played a pivotal role in raising awareness of GenAI as a technology and lowering the barriers to accessing large language models (LLMs). ChatGPT has also exposed a number of the shortcomings of the technology and the need for caution. This was illustrated by a 2023 legal case, where a lawyer representing a client in a personal injury case against Avianca Airlines used ChatGPT to research and prepare supporting documentation[10]. The ten-page brief generated by ChatGPT and submitted by the lawyer to the court contained quotes and legal precedent from more than half a dozen relevant court decisions. However, none of these quotes or court cases actually existed; ChatGPT had invented them. The court case was thrown out, and the lawyer was fined for submitting false documentation to the court. This case highlighted a fundamental misunderstanding of what ChatGPT can do.

ChatGPT is not a search tool, it is a tool for generating content based on similar content it has been trained on. If ChatGPT cannot access content that contains an answer to a question, it will do what it is designed for, it will generate that content, leading to grammatically correct but factually incorrect answers or hallucinations.

Improving data quality – the FAIR Principles

Aligned with an ambition to become more data-driven, there is industry-wide interest in adopting the FAIR principles[11]– ensuring that data is Findable, Accessible, Interoperable, and Reusable, so-called 'FAIRification' – to increase data quality, reduce repeated work, and enable more informed decision-making.

Like all good initiatives, the FAIR principles are grounded in common sense – after all, what use is data that you can't find, access, integrate, or reuse? Yet, for many companies, that is the reality for a large proportion of the data at their disposal. For example, PharmaIQ's 2022 State of Digital Lab Transformation in Biopharma report, identified that almost 50% of scientific data is neither FAIR nor sufficiently prepared and available for analysis and data science[12]. Similarly, Novartis found that "...people underestimate how little clean data there is out there, and how hard it is to clean and link the data"[13]. Dark data, as defined by Gartner, is the "information or assets organizations collect, process and store during regular business activities, but generally fail to use for other purposes"[14]. In a global survey of 1,300 business and IT decision-makers, Splunk found that one-third of respondents estimated that 75% of their organization's data was dark data[15]. Significant investment into uncovering this dark data needs to be made to deliver on the promise of AI, and this should occur ahead of big AI investments.

A European Commission report estimates that inefficiencies and other issues related to not having FAIR research data cost the European economy at least €26 billion each year, taking into account the resulting impact on research quality, economic turnover, and machine readability[16].

However, meeting the exacting standards laid out by various industry bodies[17] and achieving fully FAIR data can be an unrealistic goal for many organizations, regardless of their size. While FAIR data may be the nirvana, for many pharma companies, a benchmark of 'FAIRer data' or data that is 'suitably FAIR' or 'FAIR enough'[18] is a more pragmatic goal. For some use cases, the mantra of 'document your data in the state you'd want to receive it from others' can be a simple, initial guiding principle[19].

Regardless, one of the underlying principles is that when data is well described with metadata, which is aligned to established, agreed data standards in the form of ontologies, it becomes interoperable – with both internal data and external data sources – and this is typically at the heart of a majority of the commonly found data challenges in any organization.

The International Data Corporation (IDC) predicts that organizations that can analyze all relevant data and deliver actionable information will achieve significant productivity benefits over those that are less analytically oriented. Organizations that lack good management practices are putting themselves at a disadvantage compared to their more 'data savvy' peers.

While this may seem daunting, there are active, practical steps that any organization should take to ensure their data is AI-ready and to maximize the success of any AI implementation.

Steps to take for AI readiness

1. Define and communicate the vision

To be successful, any initiative needs to have a clear long-term vision with tangible objectives and outcomes, and adopting FAIR is no different. High-quality, FAIR data is often cited as a pre-requisite to effective AI, but a data strategy based purely on an ambition to be able to 'do AI' is unlikely to have a positive outcome. It is important to have a well-defined overall strategy and achievable goals that are clearly communicated and understood by all. Implementing data standards may not be the most exciting topic for a typical scientist, so it is key to give everyone the vision of the 'art of the possible' once data standards are in place and to draw upon examples of what other similar companies have achieved to help bring things to life for the wider scientific audience.

For example, most people are familiar with the scenario of not being able to understand or learn from data generated by others without having access to the original creator or author of the data, sometimes necessitating repeating experimental work. When terminology is standardized, non-technical data consumers can more easily understand data across the organization. Decision-makers can make quicker business decisions, such as which projects to prioritize and which to put on hold/cancel.

To support the vision, there needs to be senior leadership team support and buy-in, ideally communicated from the top. FAIR initiatives need to be budgeted and resourced appropriately.

Benchmarking against peer organizations can also be helpful, as can drawing parallels with other industries, such as banking or telecoms, where a lack of data standards would be detrimental to us all. The reality is that while some organizations have started on this process and made significant progress, they are all still on the journey, and no one has '100% FAIR' data. There is no obvious correlation with the size of the organization – the common denominator is that the companies having the most success are those that recognize the value and benefits of FAIR and are receptive to change. Companies need to acknowledge that FAIR is not a one-off exercise but rather an adoption of these principles and their investment and planning need to be reflective of this mindset.

In August 2023, CB Insights[20] published an index on pharma 'AI readiness' and ranked the leading pharma companies. This report assessed pharma's AI readiness in terms of patent applications, partnerships, and licensing agreements, as well as their ability to attract and retain talent. Roche and Bayer topped this chart with significant investment in AI talent and the acquisition of AI capabilities. The index also assessed the execution of AI initiatives by bringing products to market or signing partnerships with AI companies. While huge investments are being made into AI, it is still too early to assess whether these investments deliver the expected returns and whether the data and data quality are sufficient to support these initiatives.

2. Develop a robust business case

Instead of pursuing the FAIR ideal for the sake of it, the focus should be on improving data management practices and demonstrating tangible business benefits, focusing on the 'why' and not the 'what' and prioritizing the most impactful use cases for the business. Well-structured and well-managed data enables organizations to be more agile and make use of state-of-the-art technologies, such as AI, to extract value from their data.

When developing a business case for implementing FAIR, it is important to include metrics that will give an indication of progress, success, and value. However, it can be difficult to quantify these benefits without putting a lot of work into recording baseline figures before data standards and FAIR principles have been introduced. Commonly used metrics include:

- The degree of reuse of existing data, e.g., the proportion of data that is used by groups other than those who created it
- Time spent aggregating data
- The time spent reformatting and preparing data prior to performing analytics, reporting, or decision-making and/or the reduction in turnaround time for data analytics requests.

Our experience supporting customers to implement data standards is that this latter metric will be dramatically reduced, and AI is likely to exacerbate this. However, some of the main benefits to an organization may be hard to quantify in terms of time savings as they are often related to being able to do things that were not previously possible, such as adopting new technology or gaining insights that were previously hidden.

3. Define and communicate a roadmap for FAIR

Any organization that starts by trying to standardize all its data will quickly become overwhelmed. When starting a journey towards implementing data standards or using AI, an iterative, agile approach, starting small and identifying well-defined 'quick wins' with limited scope, has a far greater chance of success. In turn, the resulting success stories can help educate the wider organization about the possibilities. They can also act as valuable proof of concepts for applying similar principles to more complex systems or use cases.

Nevertheless, implementing FAIR should be an enterprise-wide endeavor rather than thinking about making individual data sources or applications FAIR. A data-centric approach is key – any activity should happen within a common framework with clear and realistic timelines. We will come back to the topic of governance in more detail later, but data standards initiatives need to be centrally coordinated. Otherwise, data silos will remain.

The project can be broken down into different workstreams focusing on different areas of standards (e.g., research, clinical, non-scientific/business); responsible teams can be given some autonomy so that they can move at a speed that works for them without being held back by others, so long as this occurs within and aligns with the overall framework whereby progress can be monitored and supported; it should not be left to individual teams to try and make their data FAIR independent of an overall strategy.

When defining the roadmap, we need to consider how and when data standards will be applied across the business. For example, will standardization only be applied to the capture of new data, or also to aligning legacy data? It is typically more straightforward to start with new data and then address legacy data in a later phase once the standards and systems are in place and working, but this will depend on the specific use cases that are prioritized.

While it is useful to draw upon expertise and support from outside the organization, such as external consultants, to help with the initial groundwork and to get things started, it is crucial that the roadmap should include a gradual transition to reducing such dependencies. Companies need to own their own data standards and governance – many of the key decisions that occur along the way cannot be outsourced.

SciBite from Elsevier

4. Define, document, and prioritize the use cases

A common theme of many failed AI initiatives is an over-focus on technology rather than clearly defining the question(s) to be answered and the data required, including whether it involves only existing data, data to be generated in the future, or both. In short, it is vital to define the use case(s) clearly, the pain points to be addressed, and the benefits that doing this will bring. For example, high-level use cases might focus on faster data analysis turnaround, improving data exchange with partners, or streamlining data collation for regulatory submissions.

This brings us back to the concept of 'FAIR enough' – different use cases will have different data requirements, both in terms of the source and purpose of the data and in terms of how well it needs to be described. For example, a use case that will utilize a few purely internal data sources containing legacy data may require a different degree of 'FAIRness' to one that involves the integration of many external sources.

Once the use cases and the associated data requirements are understood, they can be evaluated and prioritized based on their value to the business and the impact on different stakeholders. Consider a project to implement data standards in an Electronic Lab Notebook (ELN). This can help lab scientists by providing centrally managed picklists to streamline data entry, increase data quality and consistency, and make it easier to compare results across experiments.

Data generators are often unaware of who relies on what they produce. It is frequently cited that data scientists spend 80% of their time as 'data janitors', collecting, cleansing, formatting, and linking data instead of analyzing it[21]. Data scientists who are consumers of the data in the ELN may benefit even more than the lab scientists themselves. What may initially appear to be a lower-priority use case based on the value to the data generators may actually be addressing a major pain point to the consumers of that data. An effective means to communicate this is to have data scientists demonstrate the steps they need to go through once they receive data generated by the lab – scientists are often not aware of this.

Ultimately, when defining a use case, it is valuable to take a holistic view of the flow of data throughout the organization – what people and which systems does it 'touch' along the way and communicate the rationale and benefits, accordingly, tailoring the message to the different audiences.

5. Put together the team

Any organization embarking on a FAIR journey needs a dedicated core team who can support the initiative. Specific roles and responsibilities, such as data stewards who are responsible for the implementation of data standards in each of the source data systems, need to be established. The people assigned to those roles need to have the capacity to support and deliver the vision – this is not something that can happen in people's 'spare' time. It is also important to understand that implementing data standards and governance is an ongoing practice, not just a project with a defined endpoint. It requires a long-term commitment by the organization to make it successful. This requirement for ongoing data standards and governance has led to the concept of 'data products.' Product owners define the data standards and governance requirements, and processes are enabled at an enterprise level to support these data products in the long term.

The team requires a blend of both ontology experience (or a desire to learn) and, more importantly, domain expertise/understanding, for example, to help develop and maintain data standards. This is likely to involve an investment in appropriate training – according to Gartner, when it comes to digital transformation, "70% of employees have not mastered the skills they need for their jobs today, and 80% of employees do not have the skills needed for their current and future roles"[22].

The team should be responsible for identifying the data standards that are relevant and appropriate for the organization and identifying gaps.

Industry experts and external consultants can help accelerate this and provide practical, real-life experience in implementing FAIR solutions rather than relying on theoretical knowledge. However, it is critical to have a core group of employees who are involved from the outset and who understand that they will ultimately be taking over operational responsibilities as the project progresses.

6. Establish governance processes

Before starting to implement data standards, it is critical to establish a governance structure and processes for managing them, including a good change control process.

Successful Data Standards & Governance

Define Standards

- Define scope for standards (entities & attributes)
- Identify existing standard terminology
- Agree on future standard reference data / terminology

Also consider:

- Identify external ontologies & standards to leverage
- Define standard naming rules for identifiers

Implement Governance

- Build a team to own the data standards
- Define a governance structure
- Implement a process (SOP) for managing change control

Also consider:

- Define Data Stewards for all relevant systems
- Identify metrics / KPIs to measure performance

Implement Supporting Tools

- Evaluate capabilities of existing systems to manage data standards
- Create single source of truth for data standard (Data Standards Catalog)
- Integrate or align end user systems with published data standards

Also consider:

- Evaluate / vendor selection of dedicated data standards system (ontology management or RMDM system)

The key responsibilities of a governance group should include:

1. **RESOURCES: define data stewards and data owners**
2. **SCOPE: agree what data & metadata should be managed**
3. **DEFINITIONS: consensus for business terminology/glossary**
4. **REFERENCE DATA: agree what reference data will be used**
5. **POLICIES: define policies for the management of data sources and reference data**
6. **PROCESS & STANDARDS: define the processes and standards that will be used**
7. **TECHNICAL IMPLEMENTATION: agree on the technical tools and rules for integration/consolidation of standards**
8. **QUALITY: consensus on data quality rules**
9. **METRICS: agree on metrics to be used to measure quality and for reporting purposes**
10. **COMMUNICATION: communicate, educate, and promote the initiative and associated critical data projects**

The Data Standard and Governance team should meet regularly, with team members reporting the progress of all relevant activities along with targets for the next period both within the team and to the sponsor(s) in the leadership team. This should also include checkpoints to demonstrate progress and value to the business. Transparency is key to seeing progress and ensuring that the project does not stagnate.

The governance team should also be responsible for evaluating external standards. There is an increasing prevalence of standards in the life sciences industry, which can present a challenge in selecting the right one for your business. For example, ensuring a standard has good coverage of the data types that are relevant, is established, suitably managed (for example, that it is version controlled), and free from errors – all of which can plague public ontologies.

7. Implement technology to manage standards

To support the ongoing management of data standards and ontologies, the standards need to be managed in one place to provide a single point of truth for the organization. While an Excel file may have been used historically to manage simple lists, such as lists of project codes, etc., this will not be a long-term solution. Dedicated systems are essential in the same way an ELN or Laboratory Information Management System (LIMS) should be used for experimental data capture and management. There needs to be a mechanism for domain experts, data stewards, and scientists to contribute to standards, suggest improvements, highlight gaps to be addressed, etc.

Similarly, define your existing data and technology landscape – understanding which systems and applications in your organization capture and use data, such as ELNs and assay registration systems, in the form of an 'application register,' the ownership and management of which can form part of the overall governance process. For each application, understand what type(s) of data it manages and which use case(s) it is involved in. The degree of structure/interoperability – a simple check of the current 'FAIRness' can help prioritize where to focus attention.

Rather than have a situation where each application manages its own ontologies, it is more effective and efficient… not to mention a far lower administrative burden… if picklists, etc., can be populated from a centralized ontology management system. As such, when selecting such a system, it is necessary to find out what off-the-shelf integrations exist, the configurability of destination tools, and whether established integrations actually meet the desired use case(s) and flow of data. Regardless of whether interfaces exist or need to be developed/configured, the work to implement and maintain such connections needs to follow a common framework and not be addressed ad hoc by different development teams using different technological solutions. It also needs to be factored into a resource plan and budget and, of course, incorporated into the roadmap – integrations need to be prioritized and phased.

8. Communicate progress & maintain momentum

Once a FAIR initiative is underway, the project leader needs to keep momentum by ensuring regular communication of project progress and key tangible achievements on a regular basis. Providing bite-sized, non-technical 'explainer' articles or videos, both of the accomplishments within the company and of interesting applications in the wider pharma industry, to spread the key benefits or FAIR data more widely. This will not only help maintain interest but also enable people to draw parallels with their own part of the business and spark ideas for new use cases.

FAIR should be integrated as an ongoing topic in existing team meetings and company forums. Senior management should continue to make it clear that this is a key goal for the organization in their own company-wide communications.

Some of the terminology associated with FAIR, such as URL references, can be impenetrable to lab scientists, so it is important to ensure messaging is tailored accordingly. Regular communications, keeping things simple and clear and avoiding sharing technical details of ontologies with end users that they do not need to know about is key to success – keep everyone informed, but not overwhelmed or confused with irrelevant details.

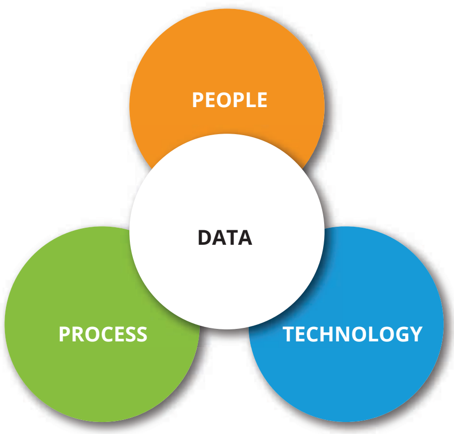
The foundations are in place… What's possible now?

As mentioned earlier, with robust data management practices in place, organizations can be more agile and use state-of-the-art technologies to extract value from their data. For example, many pharma companies are taking inspiration from the popularity of tools such as ChatGPT to apply similar principles to develop pharma-centric 'Chatbots' trained using both proprietary and public life sciences data.

Such tools are based on LLMs, which excel at 1. converting natural language to syntax and 2, summarising large volumes of data. While such technologies are extremely powerful, their output still needs to be verified as truth. This is where good data management practices and ontologies, in particular, help. Ontologies represent scientific knowledge as agreed upon by human experts in each domain: that a term has a specific meaning, is a specific type of thing, and its relationships with other types of things. Ontologies complement LLMs by providing the backbone of scientific knowledge that the LLM can use and making the output explainable and reproducible.

So, if ontologies are incorporated into your AI application, you get the best of both worlds: the AI can focus on searching through and summarizing large volumes of data while having underlying 'scientific truths' built into it.

This need for verification has only become more pressing with the emerging field of GenAI. Ontologies can provide the control at the most important step of the initial identification of relevant data, and consistency in information retrieval, ensuring the same results are returned each time via an explainable process and in a way that scientists can understand and verify by going back to the source data.



Summary

The potential of AI and GenAI in revolutionizing the pharmaceutical industry is vast, with up to an estimated $110 billion of value being added to the pharmaceutical and medical device industries. However, the practical implementation of GenAI is tempered by the need for high-quality data, requiring a data management strategy and the resources and tools to deliver it. The importance of good quality data for accurate ML and AI outputs should not be overlooked, and poor data quality remains a significant barrier to ML and AI use.

The FAIR data principles offer a framework for improving data quality and enabling more informed decision-making. This paper presents some practical steps to AI readiness, including defining and communicating the vision, developing a robust business case, defining a roadmap for FAIR, defining and prioritizing use cases, establishing a team, implementing technology to manage standards, and maintaining momentum.

The potential of AI applications, such as pharma-centric 'Chatbots,' is massive, offering new ways to interact with data and lower the barriers to accessing insight. However, AI is not a magic bullet and still requires high-quality and well-curated data to train models and extract insight from them. Therefore, the true potential of AI will only be achieved in conjunction with robust and well-resourced data management processes, as outlined in this paper.

SciBite
from Elsevier

References

1. Morgan Stanley. Why Artificial Intelligence Could Speed Drug Discovery, 2022. https://www.morganstanley.com/ideas/ai-drug-discovery Accessed 04.07.2023.
2. Pistoia Alliance. Research finds that while interest in blockchain and AI is increasing in the life sciences, barriers to widespread adoption across the sector persist, 2021. https://www.pistoiaalliance.org/news/survey-finds-ai-will-lead-to-faster-rd/ Accessed 04.07.2023.
3. Pistoia Alliance. Press Release: Pistoia Alliance Launches Lab of the Future Report, Finding AI is Set to be Top Technology Investment for Life Sciences Companies Over Next Two Years, 2023. https://www.pistoiaalliance.org/news/press-release-pistoia-alliance-launches-lab-of-the-future-report-finding-ai-is-set-to-be-top-technology-investment-for-life-sciences-companies-over-next-two-years/ Accessed 04.07.2023.
4. GlobalData. State of the Biopharmaceutical Industry, 2023. https://www.globaldata.com/store/report/state-of-biopharmaceutical-industry-analysis/ Accessed 04.07.2023.
5. McKinsey; Shah, B:, Viswa C. A.; Zurkiya,D.; Leydon, E.; Bleys. Generative AI in the pharmaceutical industry: Moving from hype to reality, 2024. https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality Accessed 04.07.2023.
6. Gartner; Perri, L. What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle, 2023. https://www.gartner.com/en/articles/what-s-new-in-artificial-intelligence-from-the-2023-gartner-hype-cycle Accessed 04.07.2023.
7. McKinsey; Shah, B:, Viswa C. A.; Zurkiya,D.; Leydon, E.; Bleys. Generative AI in the pharmaceutical industry: Moving from hype to reality, 2024. https://www.mckinsey.com/industries/life-sciences/our-insights/generative-ai-in-the-pharmaceutical-industry-moving-from-hype-to-reality Accessed 04.07.2023.
8. Harvard Business Review; Redman, T. C. If Your Data Is Bad, Your Machine Learning Tools Are Useless, 2018. https://hbr.org/2018/04/if-your-data-is-bad-your-machine-learning-tools-are-useless Accessed 04.07.2023.
9. Forbes; Shaywitz, D. Pharma's Desperate Struggle To Teach Old Data New Tricks, 2019. https://www.forbes.com/sites/davidshaywitz/2019/05/23/pharmas-desperate-struggle-to-teach-old-data-new-tricks/?sh=339dfa3f3434 Accessed 04.07.2023.
10. Forbes; Bohannon, M. Lawyer Used ChatGPT In Court—And Cited Fake Cases. A Judge Is Considering Sanctions, 2023. https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/ Accessed 04.07.2023.
11. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 2016, 160018. doi.org/10.1038/sdata.2016.18 Accessed 04.07.2023.
12. PharmaIQ; Watts, B. 2022 State of Digital Lab Transformation in Biopharma, 2021. https://www.pharma-iq.com/pre-clinical-discovery-and-development/reports/2022-state-of-digital-lab-transformation-in-biopharma Accessed 04.07.2023.
13. Forbes; Shaywitz, D. Pharma's Desperate Struggle To Teach Old Data New Tricks, 2019. https://www.forbes.com/sitesdavidshaywitz/2019/05/23/pharmas-desperate-struggle-to-teach-old-data-new-tricks/?sh=339dfa3f3434 Accessed04.07.2023.
14. Gartner. Information Technology Glossary. https://www.gartner.com/en/information-technology/glossary/dark-data Accessed 04.07.2023.
15. Splunk. The State of Dark Data. https://www.splunk.com/en_us/form/the-state-of-dark-data/thanks.html Accessed 04.07.2023.
16. European Commission, Directorate-General for Research and Innovation, Cost-benefit analysis for FAIR research data – Cost of not having FAIR research data, Publications Office, 2018. https://data.europa.eu/doi/10.2777/02999 Accessed 04.07.2023.
17. GoFAIR. FAIR Principles. https://www.go-fair.org/fair-principles/ Accessed 04.07.2023.
18. Pistoia Alliance. How FAIR is my data: The FAIRe(nough) benchmark at AstraZeneca, 2022. https://www.pistoiaalliance.org/eventdetails/fair-enough-data/ Accessed 04.07.2023.
19. Frey J.G. FAIR enough: A sideways look. Patterns 2021, 2, 100331. https://doi.org/10.1016%2Fj.patter.2021.100331 Accessed 04.07.2023.
20. CBInsights. AI boom? 2023. https://www.cbinsights.com/research/ai-readiness-index-pharma/ Accessed 04.07.2023.
21. Forbes; Press, G. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says, 2022. https://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#3d3e39a96f63 Accessed 04.07.2023.
22. Gartner; Baker, M. Gartner Says Only 20 Percent of Employees Have The Skills Needed For Both Their Current Role And Their Future Career, 2018. https://www.gartner.com/en/newsroom/press-releases/2018-09-06-gartner-says-only-20-percent-of-employees-have-the-skills-needed-for-both-their-current-role-and-their-future-career Accessed 04.07.2023.

About SciBite

SciBite from Elsevier offers data-first, semantic analytics software for those who want to innovate and get more from their data. Leading the way by pioneering the combination of the latest in machine learning with an augmented ontology-led approach, SciBite's semantic infrastructure answers business-critical questions in real-time by releasing the value and full potential of unstructured data.

Supporting the world's leading scientific organizations with use cases from discovery through development, SciBite's suite of fast, flexible, API-first technologies empowers customers, making it a critical component in scientific, data-led strategies.

Take the first steps on your journey – contact one of our experts today at **contact@scibite.com.**

SciBite
from Elsevier

To learn how SciBite can enable the effectiveness of your data, speak to one of our experts today or email us at contact@scibite.com

Head Office:
SciBite
BioData Innovation Centre
Wellcome Genome Campus
Hinxton, Cambridge CB10 1DR
United Kingdom

www.scibite.com
contact@scibite.com
LinkedIn: SciBite
Twitter: @SciBite
+44 (0)1223 786 120