

CONTACT US:
INFO@ZONTAL.IO

20
20

FAIR DATA

How data increase the value of Biotechs



ZONTAL
DIGITAL INTEGRATION HUB

DID YOU KNOW?

Research and development (R&D) is a major source of decreasing return on capital (ROC). Over the last seven years, the average cost to develop a drug doubled.

2020 Global Life Sciences Outlook. Deloitte Insights.
<https://documents.deloitte.com/insights/2020globallifesciencesoutlook>.

BIOTECH DATA CHALLENGES

When data is not properly documented or preserved, researchers can no longer easily interpret, communicate, or share information. Insufficient archive databases, metadata, and formatting standards make it difficult for both researchers and machines to easily find, search, filter, download, and integrate data. To ensure that an increase in data production generates an increase in knowledge, a modern biotech company needs to practice effective data management.

GOOD DATA MANAGEMENT IS THE KEY TO SCIENTIFIC DISCOVERY

Most biotech companies already apply data management solutions like ELNs, Data Lakes, repositories, LIMS, fileshares, and sharepoints. However, these strategies don't often integrate with each other nor can they capture all datasets or types.

AN EMPIRICAL ASSESSMENT OF METADATA QUALITY IN WELL-KNOWN ONLINE REPOSITORIES

The NCBI BioSample:

- 15% use attribute names that do not exist in BioSample's attribute dictionary.
- 68% of ontology-term attributes do not correspond to actual ontology terms.
- Only 27% of Boolean-type attributes were valid.

The EBI BioSamples:

- 89% use custom attribute names.

TGonçalves, R. S. & Musen, M. A. The variable quality of metadata about biological samples used in biomedical experiments. *Scientific Data* 6, 190021 (2019).



DEFINING DATA MANAGEMENT

A lack of data harmonization and data descriptors requires a continuous and costly investment in updated software. If current solutions are inadequate, how do we define and measure good data management?

The FAIR Guiding Principles: A prerequisite for good data management

From 2009-2019, the average R&D cost per product was **\$985 million**.

Wouters, O. J., McKee, M. & Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. JAMA 323, 844-853 (2020).

As an answer, industry stakeholders and academia have developed a community-approved set of principles to help researchers easily find, access, integrate, cite, and reuse data. Published in 2016, the FAIR Guiding principles provide concise and measurable mileposts that apply to data, algorithms, tools, and workflows.

FINDABLE

Data includes unique identifiers and rich metadata to enable automatic discovery.

ACESSIBLE

Using the assigned identifier, data can be accessed by humans and machines. Open access protocols are clearly defined.

INTEROPERABLE

Using a common vocabulary and formatting, data can be integrated with other workflows for analysis, storage, and processing.

REUSABLE

Data includes clear usage licenses and detailed provenance. Metadata and data meet domain-relevant community standards.

The FAIR principles support a variety of applications due to their distinction between data and metadata. Every company will prioritize and apply principles differently, but in all situations, FAIR solutions should enable data discovery and support third-party reuse.



IS GOING FAIR WORTH THE COST?

To apply FAIR principles, many changes are required in processes, technology, data, and people. Initial implementation costs can be significant depending on the amount of scientific data and the effort required. However, the benefits far outweigh the costs.

Effective management strategies

TRANSFORM DATA INTO A STRATEGIC ASSET

To advance knowledge discovery and innovation, biotech companies need to make good data management not only a goal, but a requirement. In the last few years, FAIR principles have become an expectation within larger pharmaceutical companies; now it is time for CROs to catch up. Those who apply FAIR principles will improve data integrity, simplify the research process, and accelerate R&D productivity.

Error-elimination

CLEAR CONTEXTUAL CLUES. ACCURATE INTERPRETATIONS.

Poor metadata causes data misinterpretation and avoidable human errors. The FAIR principles prevent this through data governance, master data alignment, and controlled vocabularies to improve readability for both machines and humans.

CASE STUDY

The NIH Data Commons project developed a cloud-based platform to store, share, access, and interact with biomedical research data.

Their successful iterative approach to implementing FAIR resulted in:

- More collaboration and outreach
- FAIR assessment tools
- Infrastructures to search and index digital objects
- New big data analysis methods
- Use case library that detail cross-disciplinary workflow
- Interoperability across multiple software
- An API registry
- Metadata documentation
- Identifier registry

Data Commons | NIH Common Fund. National Institutes of Health
<https://commonfund.nih.gov/commons> (2019).



High-quality data and metadata

TRANSPARENT, REPRODUCIBLE, REUSABLE

FAIR data supports effective data stewardship. The proper collection, annotation, and archiving of digital assets ensure high-quality data and metadata throughout an organization. When applied early in the research process, the FAIR principles improve external collaboration, maximize data value for publications, and transform data into valuable research contributions.

METADATA AND THE COVID-19 PANDEMIC

The lack of genomic metadata has made it challenging for researchers to reuse data in order to address the pandemic. For example, less than 33% of experiments in the International Nucleotide Sequence Database Collaboration are tagged with environment metadata. Researchers spend more time searching for metadata than they do analyzing data and finding answers.

“With COVID-19, the time and place a biosample was collected has suddenly become a life and death issue...poorly described data are still all too common across genomic and metagenomic studies.”

COVID-19 pandemic reveals the peril of ignoring metadata standards | Scientific Data.
<https://www.nature.com/articles/s41597-020-0524-5>.

Productive Collaboration

REUSABLE RESOURCES. OPEN COMMUNICATION.

FAIR minimizes the gap between research, development, and production. With community-supported rules of engagement, FAIR improves communication between data users and creators. Efforts to “go FAIR” can be coordinated with external stakeholders, leading to productive collaborations. Stakeholders become a part of a broader community as they discuss the meaning and application of FAIR and solve problems along the way. Centering the e-Science community around the common goal of FAIR data will improve efficiency and productivity for all stakeholders.



Artificial intelligence and machine-learning

GAIN INSIGHTS. MAKE PREDICTIONS.

The FAIR Guiding Principles promote the capacity to interact with data and metadata without human intervention. In-Silico experiments are the greatest hope of addressing the scope and complexity of drug development processes, but they require machine-to-machine communication.

With generalized interoperable technology and standards, a FAIR corporate data strategy will streamline analytical insights by helping machines recognize and integrate data automatically on a large scale.

Accelerated innovation

FLEXIBLE, EFFICIENT, ADAPTABLE

Accessible data enables primary and secondary reuse, provides new value in multiple contexts, decreases the time needed for clinical trials, personalizes data strategies, and supports collaboration across the industry.

From technology implementation to data accessibility and integration, FAIR data advances the development of effective and replicable solutions. By applying FAIR principles, data producers, managers, and publishers expedite innovation.

FINTECH AI

Innovative technologies like artificial intelligence are driving growth in the FinTech sector. More consumers than ever can access bank services and data due to increasing internet and smart device coverage. Better data analysis procedures enable financial institutions to turn unstructured data into valuable information.

FinTech and Digitalisation Report, April 2020.
<https://www.mnb.hu/en/publications/reports/fintech-and-digitalisation-report/fintech-and-digitalisation-report-april-2020>.



More findable, accessible, and machine-readable data will save researchers time and money, enabling better products, faster timelines, and more innovative solutions.

HOW TO GO FAIR

Although there is no one-size-fits-all FAIR implementation guide that applies to the entire data science industry, the following ideas can get you started:

Assess and create a plan

- Evaluate your current metadata and data policies, processes, and culture.
- Develop an implementation plan that addresses the needs and challenges of all stakeholders throughout the data lifecycle.

Invest and reformulate

- Budget with FAIR in mind and invest in the necessary costs.
- Reformulate your metadata to be as machine-actionable as possible.
- Shift business processes to be FAIR-compliant.

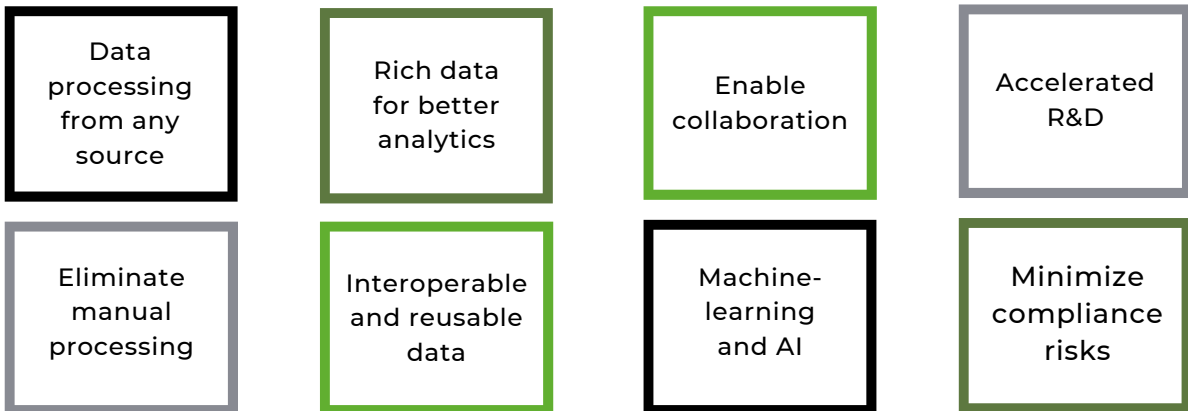
Measure and incentivize

- Use FAIR metrics to measure your progress.
- Adapt your technical infrastructure as you gain more insights.
- Continue to discuss, challenge, and refine your implementation choices.



ZONTAL SPACE WILL SIMPLIFY YOUR TRANSITION TO FAIR DATA

As a team of data engineers, architects, and scientists, ZONTAL is devoted to solving the biotech industry's data challenges. We are the first enterprise class, "good practices" (GxP) validated data platform able to manage data for the full product lifecycle.



Our products remove barriers between IT systems and lab devices, archive all digital assets on an enterprise scale, eliminate manual processes through regulatory automation, generate submission assessments, and provide advanced analytics to power ML/AI tools. Due to our 30+ years of industry experience, we deliver successful R&D integrations for outstanding results.

Our products enable compliant data centricity, efficient product development, faster time-to-market, and better patient outcomes.



TRANSFORM DATA INTO KNOWLEDGE

Find out how ZONTAL can
make your business FAIR!

We are excited to provide additional
insights into our products.

To learn more, please contact:
info@zontal.io

..ZONTAL..

DIGITAL INTEGRATION HUB

YOUR COMMITMENT TO FAIR WILL SHAPE THE FUTURE

References

AI: (Meta)data are retrievable by their identifier using a standardised communication protocol. GO FAIR <https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/>.

Dunning, A., Smaele, M. de & Böhmer, J. Are the FAIR Data Principles fair? *International Journal of Digital Curation* 12, 177–195 (2017).

Mons, B. et al. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use* 37, 49–56 (2017).

FAIR Principles. GO FAIR <https://www.go-fair.org/fair-principles/>.

Wise, J. et al. Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discovery Today* 24, 933–938 (2019).

Implications of F.A.I.R. on a modern data strategy in Life Science R&D. <https://www.linkedin.com/pulse/implications-fair-modern-data-strategy-life-science-rd-hemant-varma>.

Stall, S. et al. Make scientific data FAIR. *Nature* 570, 27–29 (2019).

Boeckhout, M., Zielhuis, G. A. & Bredenoord, A. L. The FAIR guiding principles for data stewardship: fair enough? *Eur J Hum Genet* 26, 931–936 (2018).

Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018 (2016).

BDO Biotech Briefing. BDO <https://www.bdo.com/insights/industries/technology/2018-bdo-biotech-briefing>.

Data Commons | NIH Common Fund. National Institutes of Health <https://commonfund.nih.gov/commons> (2019).