FAIR data

Transforming the life sciences industry

ONTOFORCE

Summary

Recent years have brought an explosion of data in life sciences which is used to drive decision-making, inspire research and development (R&D), and inform strategic plans, but organizations cannot leverage the full potential of their data without sound data governance[1].

Powerful techniques like machine learning, for instance, allow researchers to approach and analyze large data sets but the data access is predicated on the assumption of data reusability. Unfortunately, such reusability requires changes in the management and stewardship of research data - a broad transformation to align with what are known as FAIR Principles.

FAIR refers to a data[2] environment in which data is Findable (discoverable with machine readable metadata), Accessible (available and obtainable to both human and machine), Interoperable (synthetically parsable and semantically understandable to allow exchange), and Re-usable (sufficiently described and shared with the least restrictive licenses)[3,4].

First introduced in 2014 and born out of the collective experience of scientists working with data and metadata, the principles do not strictly define how FAIR is achieved but rather provide a roadmap. As a result, 14 exemplar universal metrics were developed with the goal of being universally applicable to all digital resources in research domains[5]. Professor Michel Dumontier, who published the FAIR Principles framework, describes it as more than just a data graveyard. Rather, he sees it as making sure the data are found and reused by others; if this cannot be demonstrated then the data are not FAIR.

Powerful techniques like machine learning, for instance, allow researchers to access large data sets but that access is predicated on the assumption of data reusability

FAIR refers to a data environment in which data is

FINDABLE, ACCESSIBLE, INTEROPERABLE, AND REUSABLE.

How FAIR helps businesses

FAIR principles have been taken up by many data-driven communities since its initial publication in 2016, with grant and paper submissions often having a requirement to explain how the data will be made FAIR. In this way, the data is subjected to rigorous checks and vetting processes that ensure a constant format and standard and are submitted to repositories of the specific data type, which makes it findable.

With FAIR data implemented and integrated, specific research queries can be answered more rapidly. Researchers can place more focus on data interpretation rather than collection or recreation, thereby driving research efficiency. Valuable datasets can be created in shorter timeframes, leading to innovation faster and less time mired in data handling[6].

One of FAIR's benefits is to clarify data licensing to ensure intellectual property protection, removing liability from the data producer. Making data FAIR is not incompatible with patenting. Indeed, open science resources can be licensed in ways that define the conditions for access[7]. Data can be FAIR without being made public.

Beyond benefits to research, there is an expected business impact measured in terms of return on investment (ROI), which in the longer term will be based on reduced costs and increased productivity.

An example of this in practice was an event at a pharmaceutical company that led to the cessation of production for six months; it was eventually realized that the problem had occurred three times previously and could have been remedied more quickly if FAIR had been implemented to allow data findability[8].

Cost-benefit analysis

Despite the benefits, there has been limited analysis to determine the economic and non-economic value of FAIR data, prompting one cost-benefit analysis report from the European Union[4]. To estimate these costs they defined seven indicators, five of which were estimated based on an assessment of the inefficiencies arising in research with the absence of FAIR data. These include time wasted due to unfindable, unstructured and incomplete metadata, costs of extra licenses required for access to data that was not freely available as well as additional storage, costs resulting from multiple copies of data - stored in information silos - produced by partners and collaborators when the data was inaccessible.

The absence of applied FAIR data principles also impacts downstream research through possible duplication of work and ineffective collaboration. The economic impact occurs through loss of innovation and potential growth.

Overall, an estimated annual cost of not having FAIR data was a minimum of €10.2 billion per year,

Although the actual costs could be much higher due to unqualifiable elements like the value of improved research quality and other qualitative metrics[4]. 60% of this figure is due to the impact on innovation, while time wasted, license costs, storage costs, research duplication and retraction accounted for the remaining 40%.

Further breakdown of costs shows over €3 billion lost by non-academic industry researchers in lost time, over €5 billion in unnecessary storage costs for siloed data and over €350 million in license costs due to the absence of open access. Double funding and research retraction were priced at €25 million and €4.4 million respectively[4].

FAIR benefits

1 Faster response to specific research queries.

2 Clearer data licensing to ensure intellectual property protection, removing liability from the data producer.

3 An estimated annual cost saving of minimum €10.2 billion per year.

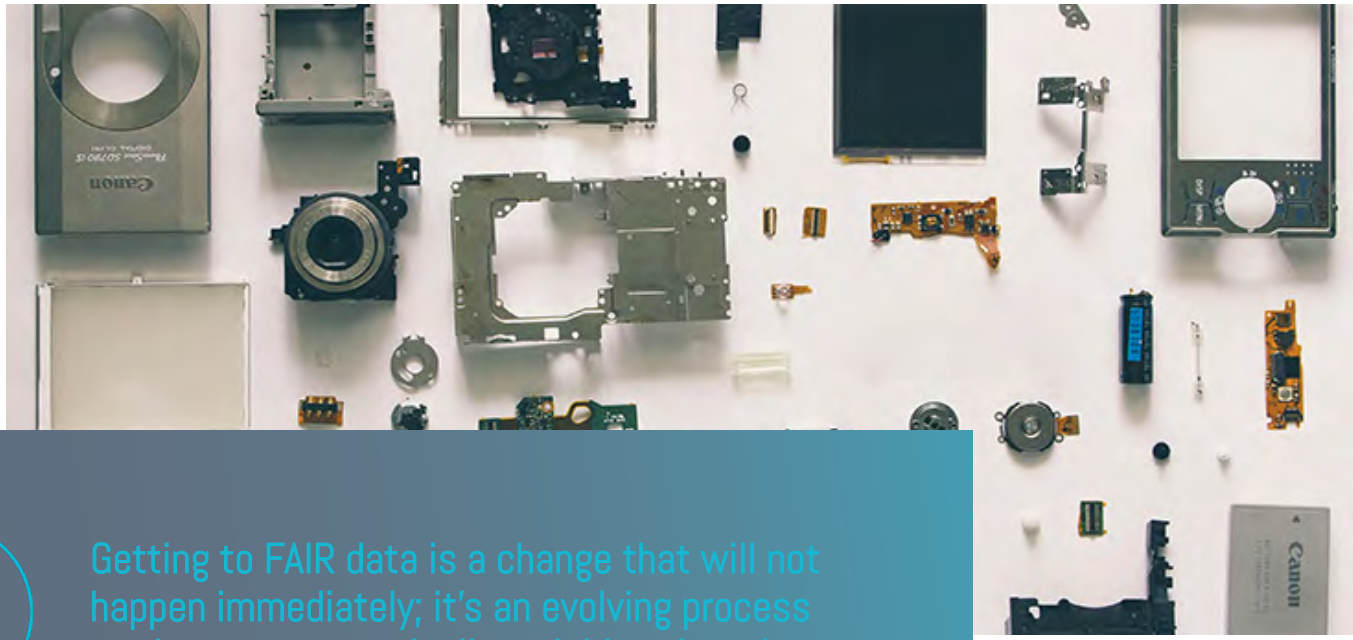
Getting to FAIR:
Challenges and misconceptions

There are many obstacles to achieving FAIR principles in an organizational context. Data producers are often reluctant to share their work for reasons including a lack of trust in the methodology, a perceived lack of recognition and the associated loss of credit, expertise, and career benefits. Indeed, while significant effort has been expended on crafting mandates and standards for data-sharing, considerably less has been spent on highlighting the value of such data sharing and its benefits to the researcher.

With the impact of research still measured primarily through the number of publications, the value of subsequent uses of data can be overlooked. Data-sharing can be incentivized through systems that track the use, reuse, and impact of shared research data. This would also help overcome a reluctance to curate data which the researcher believes is unlikely to be useful to others.

Other implementation barriers include costs required for training, process, technology, and data. Executive management may need to be convinced that FAIR implementation will generate a long-term return on investment[9].

A fundamental lack of understanding of how data can be made FAIR is another challenge. Stakeholders need to be aware that change will not happen immediately; it is an evolving process carried out pragmatically and driven by value. Rather than making all operations or processes FAIR simultaneously, companies can start the process and adopt a test-and-learn approach[10].



Getting to FAIR data is a change that will not happen immediately; it's an evolving process carries out pragmatically and driven by value.

Principles into practice

As the life sciences industry harnesses more data-driven activities, implementing FAIR principles allows organizations to make maximum use of advanced analytical tools such as artificial intelligence and machine learning which could in turn support their digital transformation, from research to the managing of clinical trials, regulatory compliance, and corporate governance. In the pharmaceutical industry for example, drug developers can assess all data available on a specific condition, including patents and existing medicines.

Translating FAIR principles into practice is essential to make full use of life science data. The benefits of doing so are that, it will firstly speed up innovation and secondly when data from both successful and failed experiments are available, this builds on existing knowledge and past missteps.

Convincing life sciences companies about the benefits of adopting FAIR principles requires both technological and cultural arguments. In terms of culture, data producers need to feel recognized and appreciated; they should understand how their work powers research in the larger science community and feel a sense of belonging to and being part of that wider mission. This means shifting from a 'me'-centric to a data-centric culture and an embrace of collaborations and sharing[11]. The cultural shift from protective and siloed data to one of data sharing requires both top-down efforts from senior management and bottom-up efforts from data scientists[12], best accomplished in the context of the organization's data strategy and a change in ideas on the value of data and its curation.

In terms of technology, it might not be feasible to make existing data FAIR as the absence of good data hygiene during collection and organization cannot easily be fixed after the fact; instead, help data producers to implement FAIR principles now, before data collection, initially on a small scale which can be built upon.

To successfully realize the FAIR principles, companies must be organizationally structured to assess key performance indicators (KPIs) by quantifying the actual implementation of FAIR. Professor Dumontier argues that someone else being able to find and reuse your datasets is a good indicator of whether the FAIR vision has been fulfilled. An example would be quantifying how many reuse scenarios a dataset has.

Translating FAIR principles into practice is essential to make full use of life science data.

This means shifting from a 'me'-centric to a data-centric culture and an embrace of collaborations and sharing.

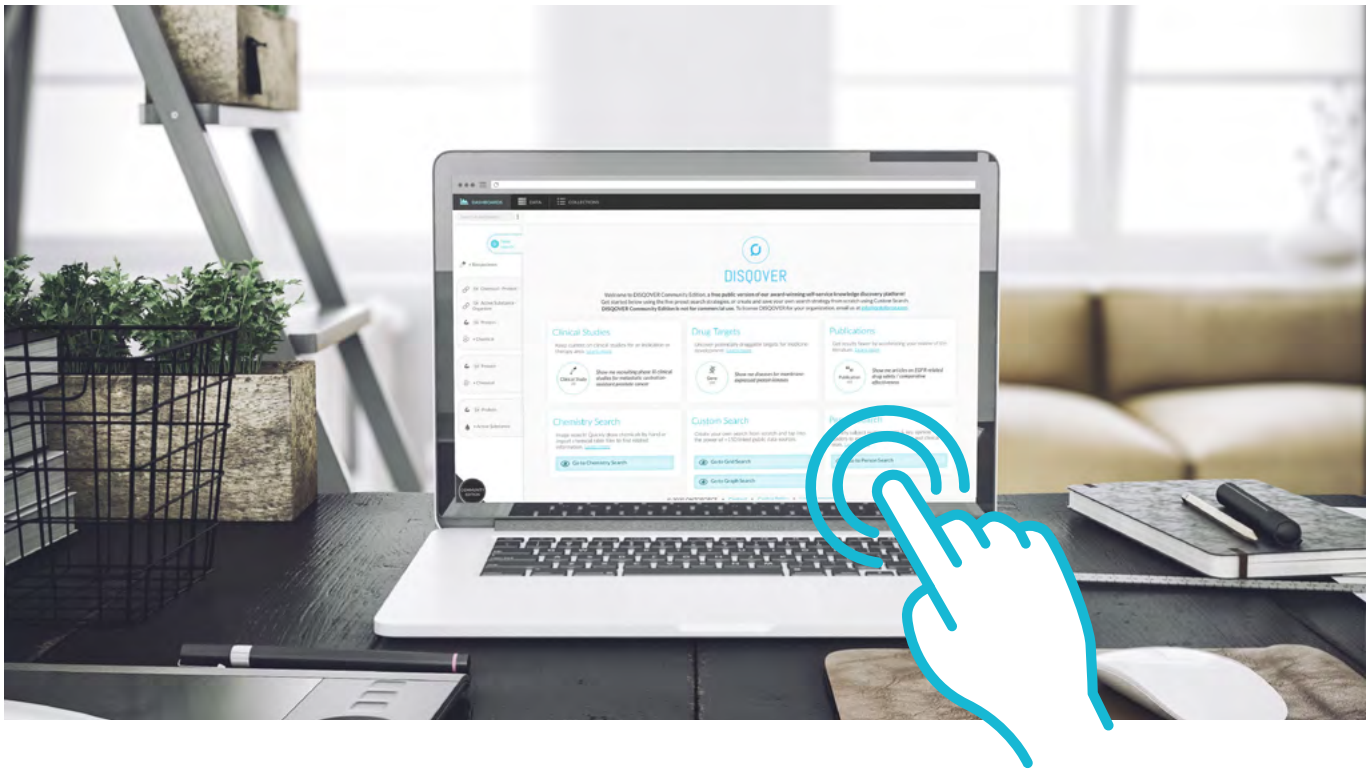
Convincing life sciences companies about the benefits of adopting FAIR principles requires both technological and cultural arguments.

How ONTOFORCE can help

Researchers often lack time to process all the data they have collected, but this can be made easier through data modelling and semantic searching, which is provided by ONTOFORCE, allowing the aggregation of data for future analysis. By providing semantic searching, provenance, smart filtering, comparisons of capabilities, along with intelligent visualization, our software makes data easier to find, filter and compare, freeing up time for developing treatment with the highest chances of success.

For example, an important part in the study design process is the ability to re-reuse data which had been generated for another primary use. This can save time and money because new experiments or clinical trials will not be required. Only 14% of all drugs in clinical trials are granted FDA approval[13], and ONTOFORCE provides the ability to optimize study design to improve the chance of success. Areas of optimization include effective selection of study endpoints, using historical data to select study sites and improve study outcomes while unlocking insights about competitors[14].

The application of ontologies - formal naming conventions used to organize information - gives a uniform resource identifier (URI) or digital distinguishing reference, to different categorizations with the same meaning.



Searching for clinical trial phases within the semantic search platform DISQOVER.

Ontologies can identify ambiguities in data and provide a means of standardizing and cleaning data. Consider for example the use of the terms 'non-recruiting' and 'recruiting' as descriptors for one clinical study, while another uses the terms 'not yet recruiting' and 'active'. Applying an ontology to this categorization means that the different terms will be discovered in the same search since they have the same meaning. The use of ontologies provides further benefits when different ones are mapped to each other, allowing integration of disparate data, and increasing access through linkage of data silos.

Applying an ontology to a category means that different terms that have the same meaning will be discovered in a single search.

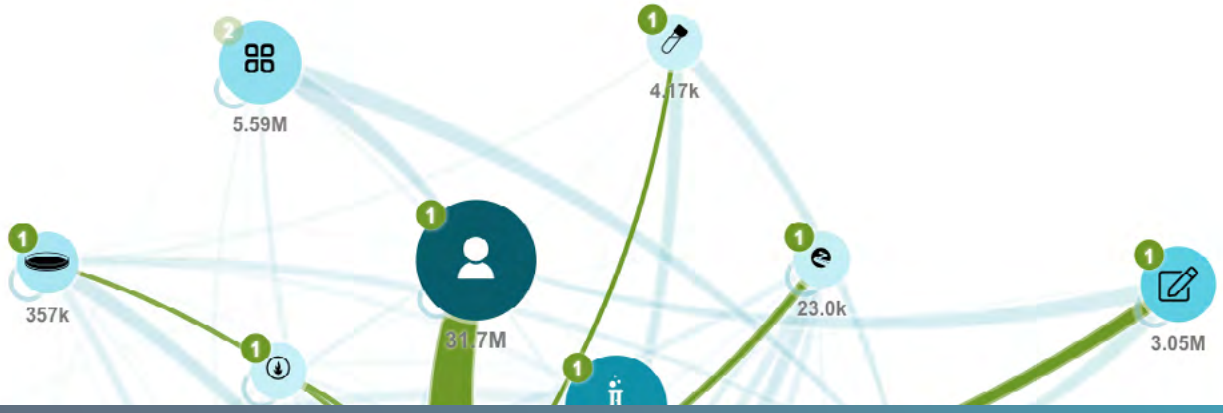
When performing a semantic search, results can be complex with patterns and insights often hidden in an inscrutable format. Effective visualization can make the data comprehensible and highlight important features. This is a useful tool when navigating through large amounts of data and allows users without a data science background to easily perform complex queries .

Effective visualization can make data comprehensible and highlight important features.

Some types of data such as clinical trials or early discovery data should only be visible by those who are a part of the project. ONTOFORCE's platform provides a level of data protection which allows boundaries to be defined and keeps details confidential. Data governance is tied to data protection, as companies need to ensure that their data has good management systems in addition to FAIR principles.

There is also value in sharing data across the company; information generated in the early stages of development or during the clinical trials might be of importance for someone working in medical affairs and after the commercialization of the drug. Having access to shared data across departments allows sharing of a diversity of knowledge which can help trigger new ideas, lead to new hypotheses, or make sense of challenges that might arise.

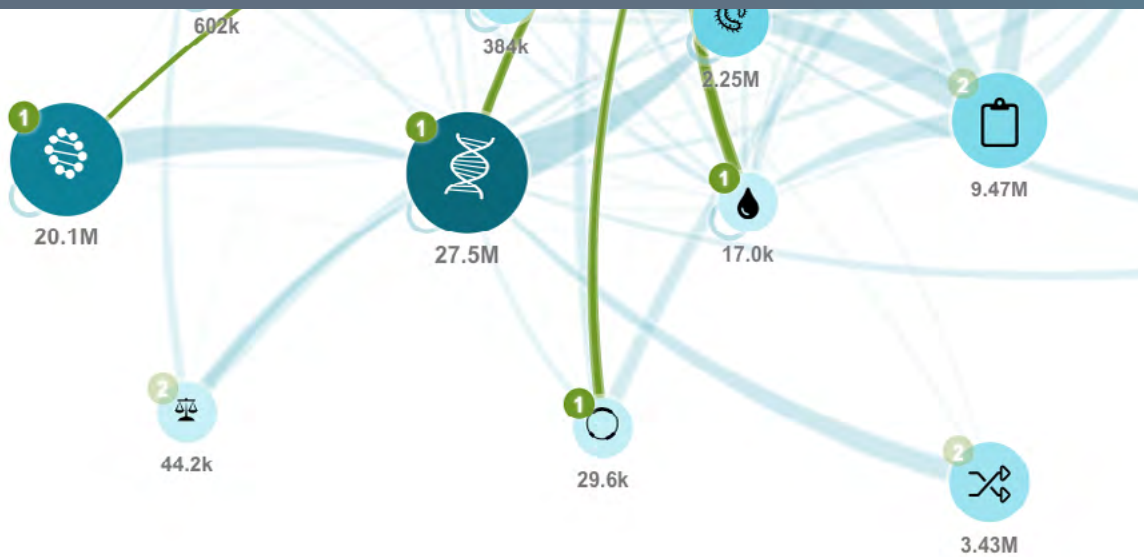
ONTOFORCE's platform provides a level of data protection in its software which allows boundaries to be defined and keeps details confidential.



Conclusions

EU research shows that collecting data that is not FAIR costs the European economy at least €10.2 billion annually. Implementation of FAIR principles can drive innovation by improving access to data and accelerating research and development. Significant barriers do exist, and cultural and technological shifts within organizations are required to ensure effective implementation. These changes include a shift from the siloed approach of data storage to one where data sharing is recognized as adding value to the organization, and therefore should be embraced. To help with this, data producers should be encouraged to implement FAIR principles in a pragmatic way before data collection, rather than attempt to FAIRify existing data. ONTOFORCE's platform holds technologies such as semantic searching, provenance, and intelligent visualizations that assist with implementing FAIR principles. Our software simplifies study design and improves outcomes by making data easier to find, filter and compare, freeing up time for designing studies with the highest chances of success.

Ready to innovate your clinical trial data with ONTOFORCE's transformational platform? Try DISQOVER today. www.disqover.com



Implementation of FAIR principles drive innovation by improving access to data and accelerating research and development.

Data producers should be encouraged to implement FAIR principles in a pragmatic way before data collection, rather than attempt to FAIRify existing data.

Sources used

Data governance:

[1] https://www.thehyve.nl/articles/fair-data-strategy-for-data-value-lifecycle

Data environment:

[2] Data refers to digital resources such as datasets, code, workflows and research objects, while metadata refers to data providing information about data, normally used for discovery and identification.

FAIR data:

[3] https://fairsharing.org/FAIRsharing.WWI10U
[4] https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1/language-en

14 Exemplar universal metrics:

[5] https://www.nature.com/articles/sdata2018118

Data handling:

[6] https://frontlinegenomics.com/a-guide-to-the-fair-principles-in-biopharma/

The conditions for access:

[7] https://ec.europa.eu/jrc/en/research/crosscutting-activities/intellectual-property

Data findability:

[8] https://www.sciencedirect.com/science/article/pii/S1359644618303039?via%3Dihub

Long-term return on investment:

[9] https://www.sciencedirect.com/science/article/pii/S1359644618303039?via%3Dihub

Adopt a test-and-learn approach:

[10] https://www.researchinformation.info/analysis-opinion/lack-fair-data-slows-innovation

Shifting from a 'me'-centric to a data-centric culture and an embrace of collaborations:

[11] https://www.europeanpharmaceuticalreview.com/article/157371/implementing-the-fair-data-principles-is-now-a-critical-endeavour/

Cultural shift from protective and siloed to one of data sharing:

[12] https://lp.frontlinegenomics.com/driving-fair-in-biopharma?utm_source=FrontLineGenomics&utm_medium=A+Guide+to+the+FAIR+Principles+in+Biopharma&utm_campaign=FAIRArticle

Only 14% of all drugs in clinical trials are granted FDA approval:

[13] https://www.centerwatch.com/articles/12702-new-mit-study-puts-clinical-research-success-rate-at-14-percent

Using historical data:

[14] ONTOFORCE White paper Data-driven clinical trial feasibility and study design

Navigating large amounts of data:

[15] ONTOFORCE White paper From big data to smart knowledge.pdf

About ONTOFORCE

ONTOFORCE TRANSFORMS DATA INTO KNOWLEDGE

For more than a decade, ONTOFORCE has addressed the problem many Life Sciences companies struggle with: bringing together structured and unstructured data to create new insights. These insights lead to accelerated drug discovery, more in-depth insights into real-world evidence, optimized clinical trial research, and faster go-to-market.

Do you wish to operate analytically, exploratively, or collaboratively? DISQOVER, the knowledge platform of ONTOFORCE, provides these insights quickly, clearly, and efficiently. Combine internal data or commercial data with the public data sources of DISQOVER, and you take the lead.

We already work for customers such as AstraZeneca, UCB and BMS, and numerous other life sciences colleagues. Thanks to the intense collaboration with renowned research institutes such as IMEC, VIB, UGent, and KULeuven and international research and industrial consortia such as ELIXIR, FAIRplus, and Pistoia Alliance, you have the guarantee of engaging with a global player that has made translating data into insights its primary objective.

ontoforce.com
twitter.com/ONTOFORCE
linkedin.com/company/ontoforce/

MAIN OFFICE
Moutstraat 108
9000 Ghent
Belgium

ontoforce.com
+32 9 396 80 07
info@ontoforce.com