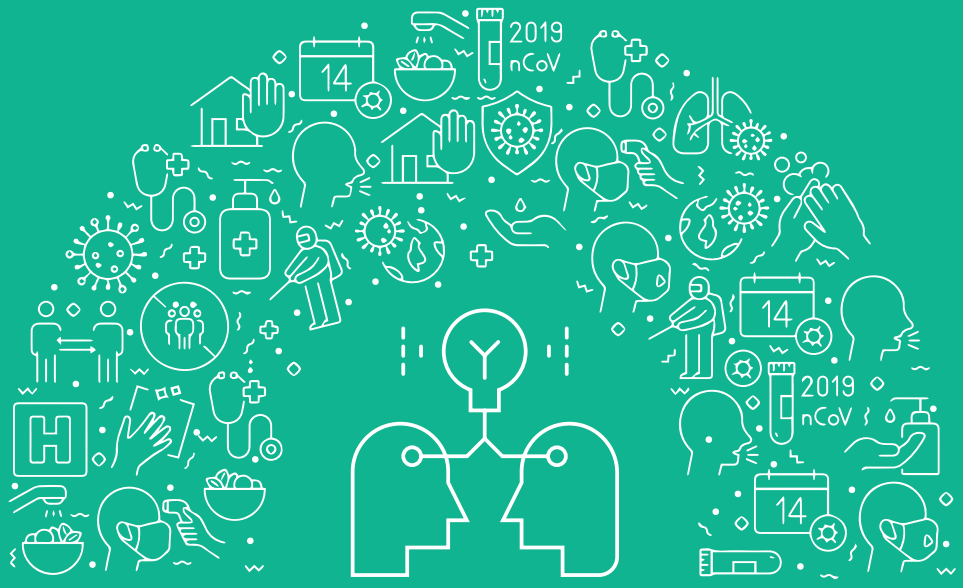


Trend 7: Knowledge sharing and data integration build stronger healthcare ecosystems

- Companies are shifting their perspective from treating data and clinical evidence as proprietary and private to joining more open and collaborative ecosystems, sharing to convert data into insights and evidence.
- Researchers can use artificial intelligence (AI) and other technologies to more effectively discover, access, and navigate data in the healthcare ecosystem.
- We expect that data and evidence integration across the healthcare ecosystem will lead to transformative new products, drugs, and treatments, and will bring commercial success for more participants.

Data is vital to pharmaceutical companies across discovery, patient enrollment, clinical trials, and commercially, to prove and improve the real-world impact of medicines.

Complex and increasingly diverse data sources are accessed across the value chain for data science and research and commercial practitioners to convert into actionable insights and clinical evidence. Increasingly, a significant amount of previously proprietary data is now shared in an open, collaborative, and federated healthcare ecosystem, as discussed in a recent Financial Times panel

with Subhro Mallik, executive vice president and global head of life sciences, Infosys.

Companies need the capability to link internal insights and evidence across firms. This is essential for greater drug discovery, commercial success, improved drug development, and efficacy across the ecosystem. However, inter-enterprise flows increase complexity, as parties collaborate and compete at the same time, sometimes in the same space. Life sciences firms need strong governance and data-sharing policies to negotiate permissions for introducing and accessing external data sources.

R&D and commercial data integration extend beyond individual enterprises. AI and other tools will be useful in this open environment to rapidly normalize and join up data as it enters from disparate sources before it is used in data models and analyses.

Open and federated healthcare

Competition traditionally drives company innovation and operations excellence. As a logical result, businesses keep private plans, proprietary knowledge, and intellectual property as competitive differentiation against their rivals. While seeming locally optimal, this approach fails in the aggregate because life sciences innovation and healthcare improvements require data from multiple sources across the healthcare industry.

The development of an early Covid-19 vaccine — in months, not years — was the result of the collaboration between Pfizer and BioNTech and multiple healthcare and population health ecosystems. More recently, public–private partnership between the Pharma Proteomics Project, a pharmaceutical consortium, and UK Biobank, a large-scale biomedical database of genetic, lifestyle, and health information, worked together to share results of a large study of genetic variation in proteins that will accelerate the development of biomarkers, predictive models, and therapeutics.

The need for competitor collaboration and coordination is driven by three factors happening in parallel. First, available data in the sector is surging. Second, there is a strong demand for real-world evidence (RWE) of product safety and efficacy, backed up by real-world data (RWD). The RWE solutions market is expected to more than double in the next five years, from $2 billion in 2024 to $4.5 billion in 2029. And third, the need for diversity and inclusion in trials is increasing.

Healthcare stakeholders, from physicians to pharmaceutical companies to regulators, increasingly recognize the importance of RWE and RWD. The US Food and Drug Administration created a regulatory framework using RWE to "support new indications for already approved drugs or biologics and post-approval studies." Similar efforts are happening in Europe, seeking to use RWE to improve regulatory decision-making. This momentum has spurred corporate partnerships, increased private-public partnerships, and created the need for new guidelines for using RWE and RWD.

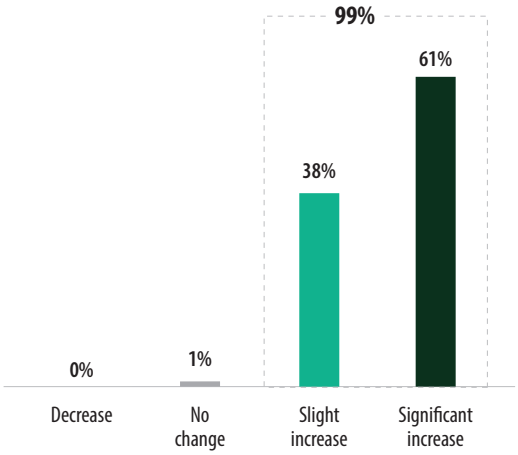
Even so, RWD is only a fraction of the data that the healthcare industry generates and consumes. The World Economic Forum estimates that in the next seven years, 30% of the world's data will be generated by the healthcare sector, especially with the rise of wearables (see also Trend 2 in this journal). The sources of data the healthcare and life sciences consumes is also increasing. They include, for instance, social determinants of health, more diverse and inclusive clinical trials, personal datasets, and behavioral science.

Behind all of this is the recognition of the ways that a wide array of data will make treatments more personalized, more precise, and ultimately more effective. This enormous amount of data will remain underutilized without a strong sharing ecosystem among R&D organizations, patient service settings, and commercial companies — and the potential for greater drug discovery and development, commercial success, and transformative healthcare outcomes.

Pfizer executive John Pastor describes companies as "going through a big mind shift" and realizing that they need to be able to analyze data where it resides. To do that, companies can leverage their partners.

Indeed, life sciences executives say they will increase spending on data integration. Data integration has long been important to global enterprises, as for decades they have slowly linked internal data repositories and applications to gain a comprehensive view of their enterprise. Now they realize

Figure 1. Nearly all executives plan to increase spending on data integration in the next five years



Source: Infosys Knowledge Institute

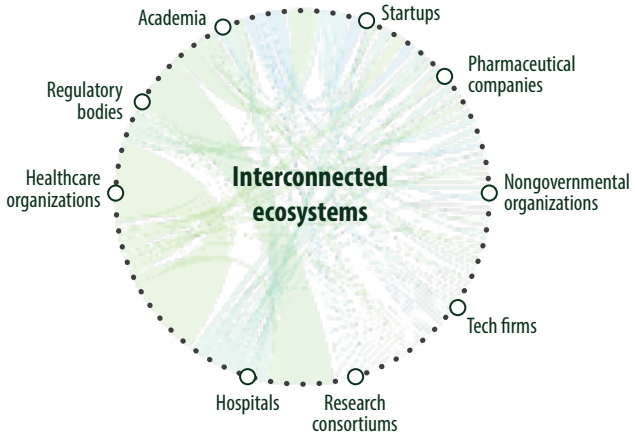
the importance of doing the same across companies. The Infosys Knowledge Institute survey of life sciences executives found that 61% intend to significantly increase spending on data integration in the next five years (Figure 1). Most others anticipate some increase in data integration over that time.

Sharing goes beyond data

Beyond data, other pre-competitive information should also be shared, according to panelists at a recent forum organized by the Financial Times. Data assets, products, models, evidence, and many other evidence-generating assets must be integrated and shared across the ecosystem.

Executives from Infosys, MSD, and Novo Nordisk discussed the need for stakeholders to share general lessons with peers and among biotech, tech giants, and startups;

Figure 2. Ecosystems close the data–evidence loop to link the commercial and R&D domains



Source: Infosys Knowledge Institute

knowledge of technology with regulators; and taxonomies, ontologies, and models across the healthcare ecosystem (Figure 2). Subhro Mallik of Infosys said this technology is changing so fast that companies will be left behind if they don't leverage their entire ecosystem.

Although companies have collaborated on R&D since the 1990s, the scale has taken off in recent years. This is driven by incremental value from the massive quantities of data that are produced, and tech advancements that enable sharing that is safe, protects privacy, and complies with regulations.

Sometimes the goal of collaborating is to create frameworks for future collaborations. For example, the Pistoia Alliance created an ontology for interoperability and collaboration among international companies. This aids compliance with the International Organization for Standardization (ISO) standards for the Identification of Medicinal Products (IDMP), mandated by the European Medicines Agency.

By solving ambiguities in this ISO standard, it also avoids future interoperability risks when ecosystem partners work together on, for example, R&D. Eleven pharmaceutical companies, regulatory bodies, standards organizations, and nonprofits worked together to create this ontology.

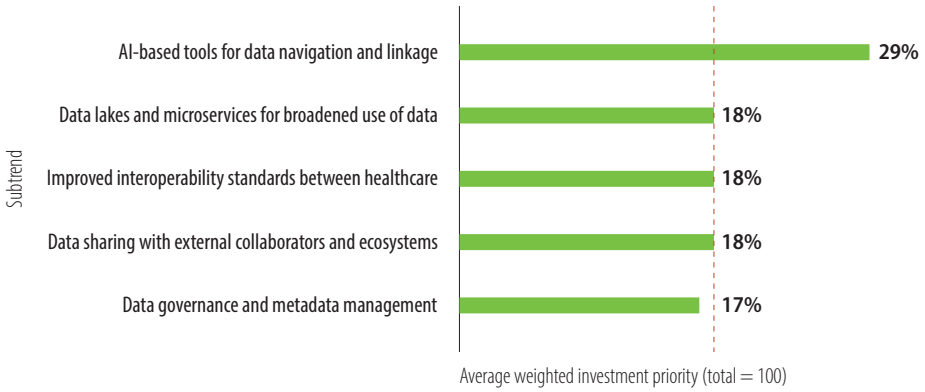
AI concierge for ecosystems

David Champeaux, lead partner, health and life sciences, EMEA, Infosys Consulting, describes how this ecosystem integration introduces an order of magnitude of complexity over integrating data within an organization. Part of this integration, he says, is navigating the systems and permissions to find the data, models, and evidence needed, like a pharmaceutical company conducting research. The researcher needs to first discover the existence of relevant datasets.

Then, they request permission to access these datasets to run algorithms on them, or even identify relevant evidence-generating methods and models to run on them.

Beyond finding and accessing data, the researchers need governance to ensure they're using the data ethically and respecting the constraints of the terms of access they've been given. Technology can enable data discovery and then navigate the permission. For example, AI tools could help

Figure 3. Life sciences companies prioritize data-related AI investments



Note. Percentages do not total to 100 due to rounding.

Source: Infosys Knowledge Institute

a researcher quickly navigate to a network of datasets and request access. The AI in effect becomes a navigation agent, a concierge.

Our survey found that life sciences executives intend to invest in AI to enhance data navigation and interoperability, and that AI tools for data navigation and linkage are the top area of investment in data integration (Figure 3).

Other technology and tools

In addition to AI, other ecosystem integration tech and tools include cloud, microservices, and interoperability standards. For example, Champeaux describes a decentralized research scenario of a pharma researcher seeking insights on population health. To develop the insight, they require access to data owned by multiple hospital groups and application of a model developed by an AI startup.

The researcher could leverage a contract to give permission to an algorithm to process that data, output the insight, pay the data and algorithm owners and enforce governance via AI-enabled permissions. This would be orchestrated under a blockchain paradigm that enables complete process traceability.

Life sciences companies need to continue their evolution toward ecosystem-wide collaboration and find the tools and tech to achieve it, even though they might find it difficult to shift their mindsets in order to pursue open collaboration, simultaneous navigate collaboration and competition, and intentionally share knowledge with rivals.

However, there are downsides to not sharing. Without collaboration and ecosystem success, non-collaborative parties will be left behind, potentially at patients' expense.

In 2020, Stephanie Reel, chief information

officer of the Johns Hopkins Health System, expressed the need to protect intellectual property while also worrying about a lack of data sharing. "I don't want us to [be] too careful and too controlling because I think there is some risk that we will not make that next big discovery or cure that dreadful form of cancer," she told Becker's Health IT.

Cybersecurity presents an additional strategic concern, as companies face two conflicting interests: openness to share data while recognizing that open AI frameworks are more attractive to cybercriminals than closely held, secure information.

The regulation landscape will play a critical role in how collaborations happen. In the US, sharing will likely be left to individual entities at large. However, the highest amount of healthcare data will be generated by India and China – two of the world's most populous countries. These countries may take a different approach, one more driven by governmental oversight and management.

At the technical level, integrating data across organizations requires normalization of datasets to a common standard so researchers can use them. For instance, electronic medical records from 10 hospitals may be structured in 10 different ways. The data might not even be structured.

Even after data normalization, data from many different sources must be linked, which represents another significant challenge. A researcher may need to link datasets containing demographic information (like social determinants of health) with patient data that is based on physical health. This is a hugely complex challenge –- the linkage must preserve patient privacy while still capturing the information in the dataset. Resolving identities across datasets, first to de-identify data and then to create the linkages, is an ideal use for deep learning.

Data in an ecosystem

AI-enabled knowledge graphs, when combined with robust metadata management, create powerful linkages, and provide deeper insights. Knowledge graphs

> "I don't want us to [be] too careful and too controlling because I think there is some risk that we will not make that next big discovery or cure that dreadful form of cancer."
>
> **Stephanie Reel**
> Chief information officer, Johns Hopkins Health System

integrate data points from various sources and reveal relationships between them.

Adding metadata enhances the data's context, consistency, and interoperability. In RWE, enriched knowledge graphs link clinical trial data, electronic health records, patient-reported outcomes, and social determinants of health, uncovering complex patterns and relationships. This combination facilitates actionable insights, improves decision-making, and leads to better patient outcomes and more effective treatments.

A company's enterprise architecture can support data integration across the ecosystem, as well as guide AI-first transformation by organizing technology to support business strategy.

Marcello Di Maulo of Infosys Consulting explains that by integrating AI, machine learning (ML), and MLOps frameworks in a comprehensive RWE data management system, a company can connect the dots between people, business, and technology at an enterprise ecosystem level. This enhances collaboration and drives transformative outcomes across the healthcare ecosystem.

While data integration is a catalyst for ecosystem value creation, success requires more than tech. Human collaboration will remain important for learning, governance, exception management, and ensuring that all parties benefit from the arrangement. As a recent article on pre-competitive collaborations describes, pharmaceutical companies need to decide which collaboration domains they will pursue. Examples include basic biology; biomarkers of prognosis, diagnosis, and treatment; drug discovery, development, and response; and knowledge and research design like hypothesis development.

They also need to define their collaboration goals, and the right forum, governance, and business models required to achieve those goals (open vs. restricted, consortium, public–private partnership, innovation center). They should also define metrics for successful collaboration and chart a path to achieve them, including understanding barriers.

Effective collaboration in the healthcare ecosystem will require organizational change, as it is a mindset shift from competitive protection to enlightened self-interest. Companies need to evaluate their skills, talent, technology, and other resources to collaborate effectively. This includes reviewing policies on data sharing, ethical use of data, compliance, and governance.

Data and evidence are the lifeblood that sustains the healthcare ecosystem, and integration is its force multiplier in the AI era. This capability will aid discovery, increase drug efficacy, and improve patient outcomes — while driving operational excellence.

In the last decade, digital transformation in life sciences provided a technological foundation. In the next decade, data and evidence integration will enable companies to thrive individually and as a cohesive industry with common interests, so the healthcare ecosystem can deliver on its promise.