EVIDEN

Innovation in Life Sciences: bring the future to life

From molecule to medical breakthroughs: powering insights to redefine care pathways

> Get Started

an atos business

EVIDEN

Contents

Foreword

Today's life sciences organizations are forging the path to precision medicine, generating new insights from the human genome, discovering therapeutic targets and developing innovative treatments.

This journey from drug discovery and development to the clinical frontline involves an evolving ecosystem of stakeholders. Scientific experts, basic and translational researchers, and clinicians from different organizations and backgrounds work with leading-edge technologists to innovate quickly.

Seamless data-centric collaboration is required between multiple players, from biotech, pharma, medical device and clinical research organizations, to tech firms, healthcare providers, academia and regulators. Stakeholders need to turn growing volumes of 'big data' into actionable, reliable and quantifiable insights that can be shared and applied to amplify best practice and foster innovation. Meeting these challenges requires scalable computing resources and digital capabilities including high performance computing, automation and Artificial Intelligence (AI).

This paper explores how advanced computing and digital tools are helping organizations to harness and extract maximum value from the plethora of biodata they work with – whether that's through powering ultra-fast genome sequencing, modelling biological processes, simulating chemical compounds or improving the outcomes and efficiency of clinical trials.

We hope you enjoy what we've shared. We look forward to continuing our collaborations with life science R&D experts working to advance and apply leading-edge technologies to help revolutionize what is possible for healthcare.

The life sciences data challenge

Today's life sciences organizations face the challenge of turning vast volumes of data into information – to help accelerate drug development or deliver a deeper understanding of the individual biology powering personalized treatments. Biodata is particularly sensitive and can be difficult to process and organize; it must therefore be robustly managed and protected.

As data-driven organizations on a quest to reduce discovery timelines, life science stakeholders are looking to supercharge how they capture, manage, utilize and share data. They need to integrate, manage, process and use huge volumes of data from a growing number of disparate sources and formats in an optimized way. This helps to accelerate innovation, improve R&D efficiency and deliver validated data and insights to partners, collaborators and regulators.

This starts with finding ways to ingress and orchestrate data in a variety of formats and structures (everything from images and text documents to patient records, videos and raw research data) and from a diversity of sources (research projects, clinical trials, high-throughput techniques, sensor and wearable devices, electronic medical records, contract research organizations, academic research partners, public health bodies and more).

To extract maximum value from these vast 'big data' resources, organizations require a data infrastructure for processing, organizing, analyzing, interpreting and disseminating data both internally and externally. A key challenge, however, is that most legacy infrastructures weren't built with today's escalating data volumes or compute requirements in mind.

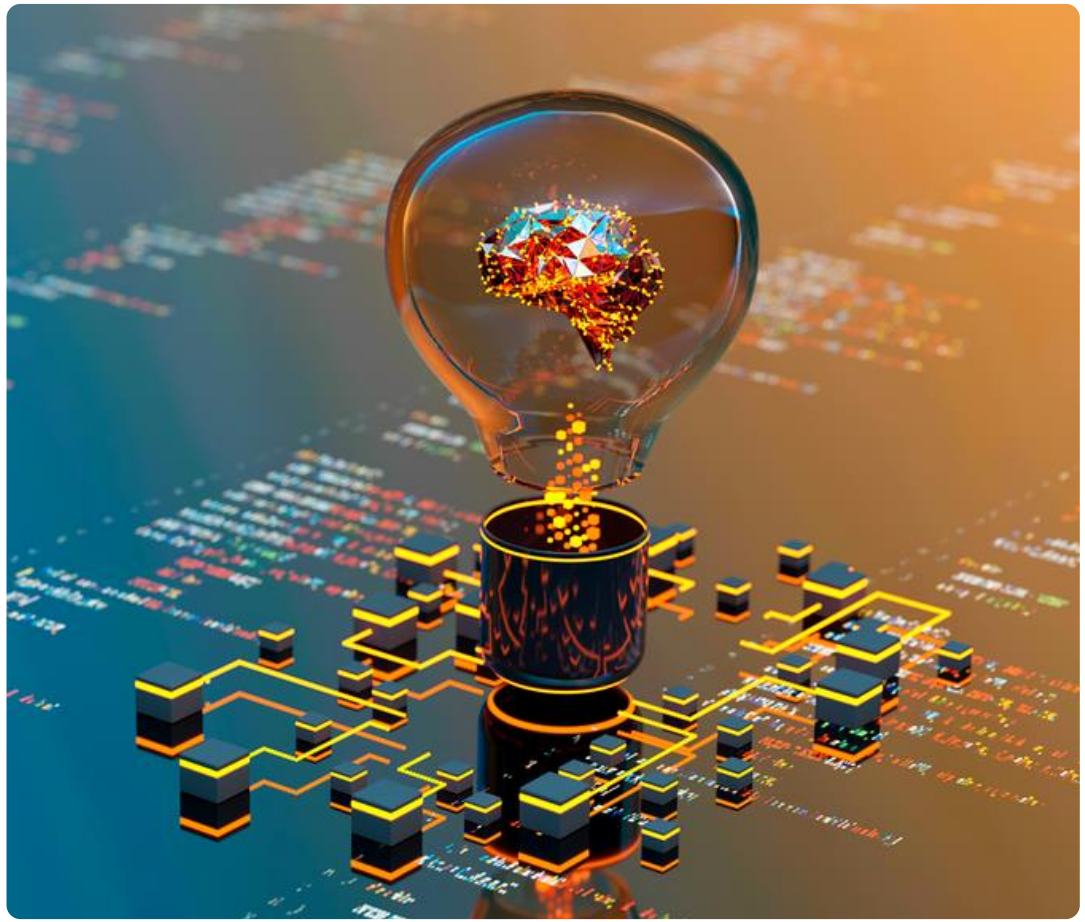
To realize the full potential of technologies such as data analytics and AI to achieve a 'faster time-to-science', good data practices that enrich the quality of data and assure data integrity are also essential. This ensures that data can easily be found, accessed, linked, reused, maintained and, importantly, traced back to its original source.

In today's ever-more interconnected world, protecting highly sensitive patient data and cutting-edge research findings means **the right cybersecurity tools and processes are essential to protect an increasingly distributed ecosystem**. These include everything from firewalls and intrusion detection mechanisms to identify verification and access management and traceability.

> The challenge with data today isn’t generating it. It’s how we manage it so that it is usable – and more than that, to generate scientific insights and drive innovation.

Turning data into actionable insights

Breakthroughs in bio-engineering – the process of bringing engineering principles of design and analysis to biological systems – combined with advances in computing, data analytics, machine learning and AI, are helping to fuel a new wave of innovation. Together with the science of bioinformatics – using software tools to analyze and understand biological entities such as DNA and proteins – these combine to transform raw data into information that will potentially lead to a better understanding of biological processes that could be translated into life-changing medical discoveries.



> Data increasingly drives everything we do and transposing that data into knowledge is the name of the game. Deep learning models and AI techniques are increasingly being applied to drug discovery across a variety of fields. These include advanced image analysis, molecular structure, and function predictions, as well as the automated generation of innovative chemical entities with bespoke properties.”

For example, the escalating volumes of data generated by genome projects around the world has resulted in the exponential growth of existing databases and the creation of new databases for use by the scientific community. Thanks to the application of computation, complex software algorithms and other bioinformatic tools, these resources can now be readily searched, interrogated and disseminated at scale.

Genomic Data Science: the fast facts

NIH National Human Genome Research Institute

According to the US National Human Genome Research Institute, **a single human genome = 200GB of data** (that's around 200 copies of the film Jaws)

Read article

NIH National Library of Medicine
National Center for Biotechnology Information

Experts predict that genomics projects will generate 40 exabytes of data worldwide by 2025. By comparison, 5 exabytes could store all the words ever spoken by human beings.

Read article

Similarly, the ability to process massive volumes of data using High Performance Computing (HPC), together with new AI computational methods, has fueled rapid advances in the prediction of protein structures (AlphaFold). This makes it easier to address R&D productivity challenges by empowering researchers to reliably interrogate vast data sets faster. As a result, they can surface meaningful and context-specific insights on everything from protein folding to protein-small molecule interactions.

By leveraging data science and HPC, lead times from discovery to clinical concepts can be significantly reduced. HPC also makes it possible to undertake the rapid data sampling, processing, and analysis that life sciences organizations require to model and run workload simulations that are accelerating scientific breakthroughs.

6 Ransomware, ENSIA Threat landscape, Jan 2019-Aoeril 2020

> Innovation is a continuous process, and the pharmaceutical industry is no stranger to this fact. It must consider how to utilize external interventions to innovate faster and more continuously”

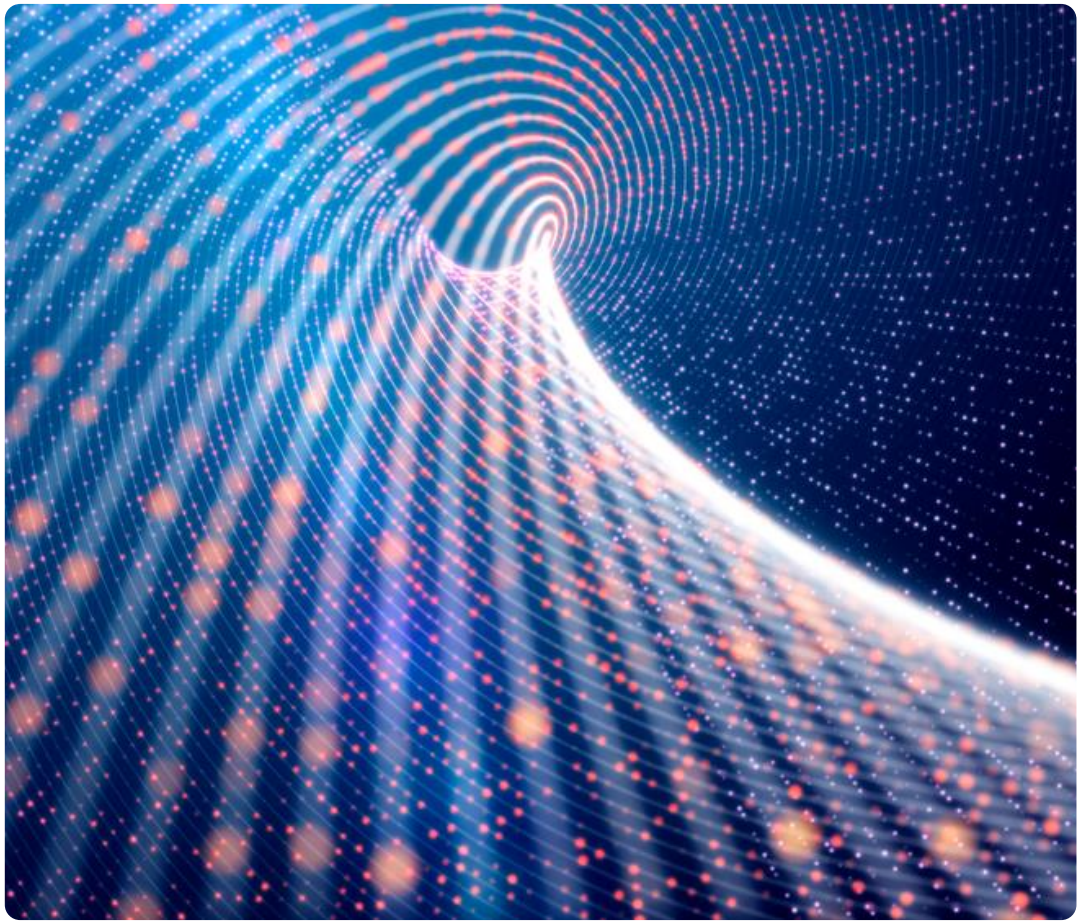
Looking ahead, the opportunities to harness the laws of quantum mechanics (through quantum computing) could be even more significant.

According to IBM, quantum will enable a range of new use cases far beyond the ability of today's classical computational techniques. These include linking genomes and outcomes to create personalized therapies, enhance the efficiency of small-molecule drug discovery and develop novel biological products based on protein folding predictions.

Read article

Using HPC to power rapid imaging analysis

The Biocomputing Unit (BCU) at the National Center for Biotechnology (CNB), part of the Spanish Research Council, used HPC and complex algorithms to process dozens of thousands of cryo-electron microscopy images into a refined 3D model of the coronavirus spike – the protein complex that has been and is still targeted to develop vaccines and treatments. By refining the 3D structure and providing unprecedented insights out of it, **the CNB was able to make an important contribution to the fight against Coronavirus**.



Accelerating disease diagnosis with AI

AI systems can now support the diagnosis of many diseases. Trained to evaluate and classify dermoscopy images and endoscopic videos. These systems can take just minutes to evaluate and highlight the most relevant image frames which are then referred to a specialist clinician for a final diagnosis. By saving precious time of specialized doctors for the more severe cases, these systems can help to transform patient care and treatment.

Through the integration of AI/machine learning and automation, life science organizations can now increase the pace and variety of their discoveries. By utilizing new computational capabilities and computer modelling techniques, they can use biological data in new and exciting ways. In addition, by connecting all stakeholders, they can use, share, and leverage data more efficiently.

Sharing and connecting data for better outcomes

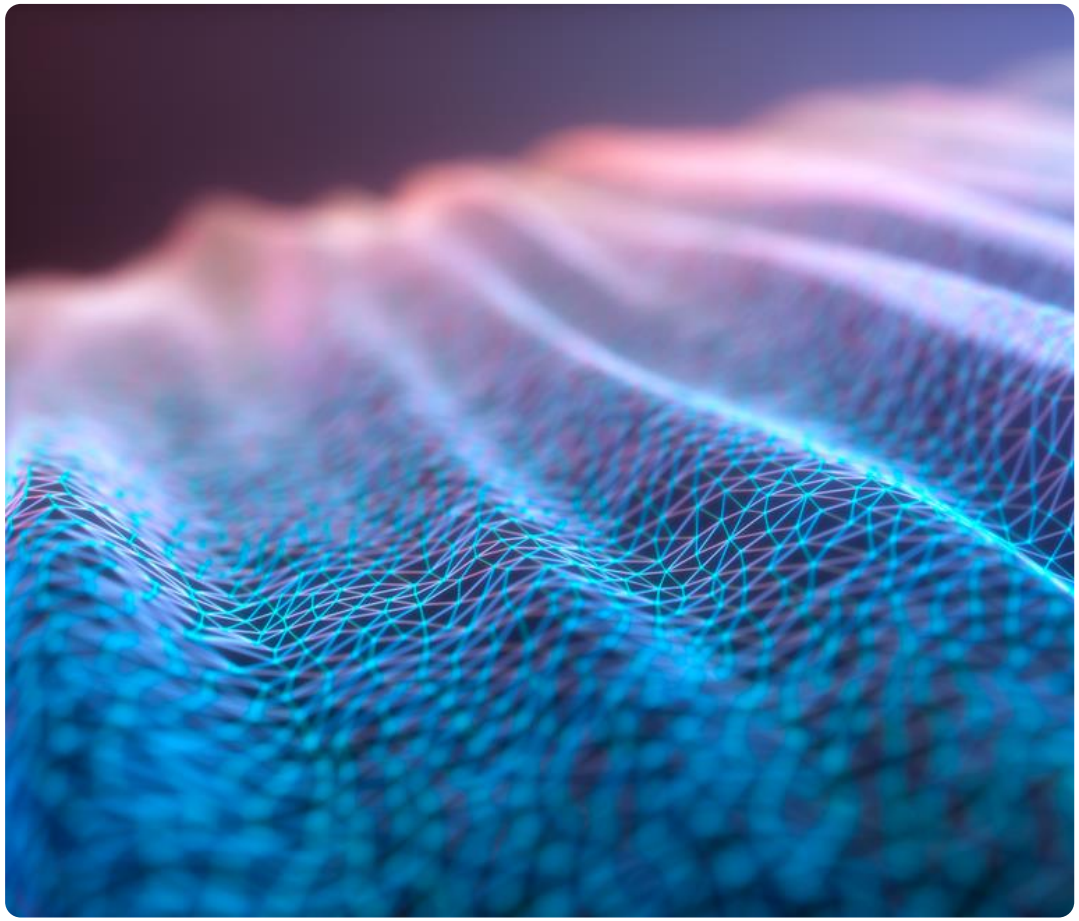
Across the life sciences industry, there is growing pressure for pharma and research companies to share data on scientific and medical discoveries.

Governments and international organizations are introducing data sharing regulations to encourage greater transparency and sharing across the sector, while eliminating some of the hurdles that can get in the way of the data sharing process.

However, data sharing can prove challenging from a variety of perspectives – cultural, ethical and technical – especially when, historically, data producers have been reluctant to share data (their IP) in the first place.

That said, initiatives for sharing research data are creating new opportunities to increase the pace of knowledge discovery and scientific progress. Indeed, the reuse and sharing of research data can generate significant benefits for all. Enabling the peer review and re-analysis of data sets creates new perspectives, as well as saving precious technical and research resources by avoiding the generation of equivalent data sets.

In the post-pandemic world, attitudes are changing as the growing number of biobanks – large scale biomedical databases and research resources – demonstrates. So much so that, wherever it makes sense, life science organizations are recognizing that connecting researchers, scientists, healthcare professionals, patients – and even citizens at large – will enable the widest possible collection and collation of data to inform basic, translational, preclinical and clinical research.



> Health records, family history, genetic disorders, lifestyle – all the social determinants of health relevant to the individual are linked to the biospecimens [and vice-versa]. This comprehensive patient profile then becomes the foundation of truly personalized precision medicine.

Dr Neeta Bhatia,

Global Portfolio Lead (Precision Medicine), Eviden Digital Health Solutions

Making this a workable reality, however, depends on establishing the right technical solutions for data sharing. These encompass everything from data storage, workflow design, data management and analysis to controlled access to data and robust data warehousing facilities. But it doesn't end there.

To satisfy real-world sustainability and usability needs, data also needs to be managed according to the FAIR Data Principles for scientific data management and stewardship. First introduced in 2016, FAIR aims to make data:

Findable
Metadata and data should be easy to find for both humans and computers.

Accessible
Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorization.

Interoperable
Data usually needs to be integrated with other data and needs to interoperate with applications or workflows for analysis, storage, and processing.

Reusable
Metadata and data should be well described so that they can be replicated and/or combined in different settings.

And, of course, to protect the privacy of citizens, mechanisms that anonymize and 'de-identify' individual patient data will also need to be in place.`

Using data to fuel collaboration

When data is complete, accurate, consistent, and sharable, organizations can enhance internal and external collaboration among all research and development partners:

- The latest scientific breakthroughs can be shared with academic collaborators and external partners such as clinical research organizations to stimulate the screening/discovery of new potential compounds.
- Collaborative open space initiatives that enable experts to address specific questions or share insights can be initiated.
- Vast amounts of new data can be aggregated and integrated for computational analysis and high throughput screening as well as enhanced modeling.

Better data management for enhanced R&D performance

Getting a new drug to market can take up to 10 years or more, with R&D and clinical trial stages accounting for a significant portion of this timeframe.

With a success rate of less than 10%, pharma organizations are looking at 'failing faster' to build the agility they need to focus on the more promising targets or drug candidates. This, pharma organizations can achieve a faster time-to-market while significantly reducing their costs. **Today's digital technologies and next-generation data management solutions are helping to accelerate these key processes – making them more precise, more adaptive, and more efficient.**

Digital twins are making it possible for R&D teams to design, plan and run multiple test simulations, utilizing cutting-edge digital and automation technologies to experiment and conduct research in an augmented, coherent and less wasteful manner.

Released from the resource limitations and cost constraints of physical lab work, teams are building digital twins of cells, tissues, organs and processes to better understand disease behavior, predict more accurately how certain target molecules and drugs will behave in the real world, and dramatically reduce drug discovery and validation timelines.

Digital twins: exploring the possibilities

Combining data and algorithms with traditional models and simulations together, and using modern technologies like smart sensors, edge computing, cloud, data science and AI, today's researchers are creating virtual 3D simulations to better understand cells and disease conditions and undertake multiple digital experiments that will help bring drugs to market faster – and more effectively.

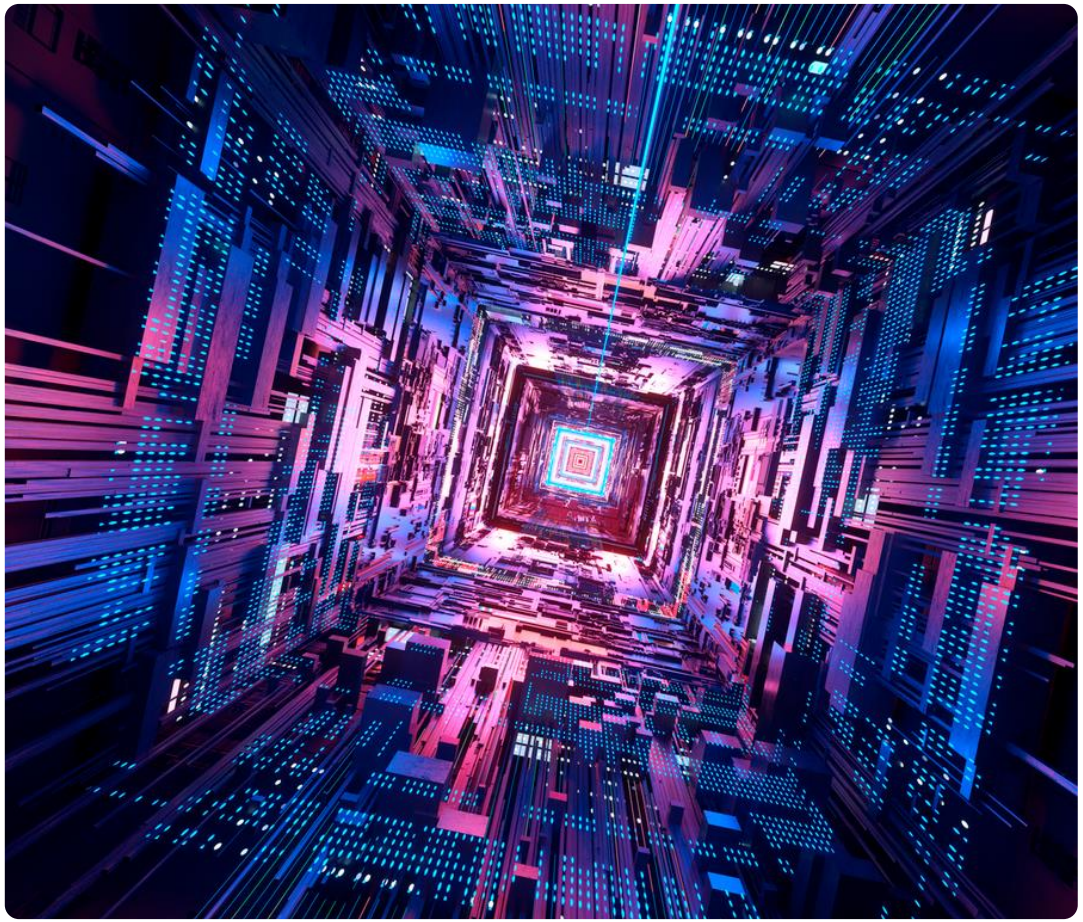
Other promising use cases for digital twins include:

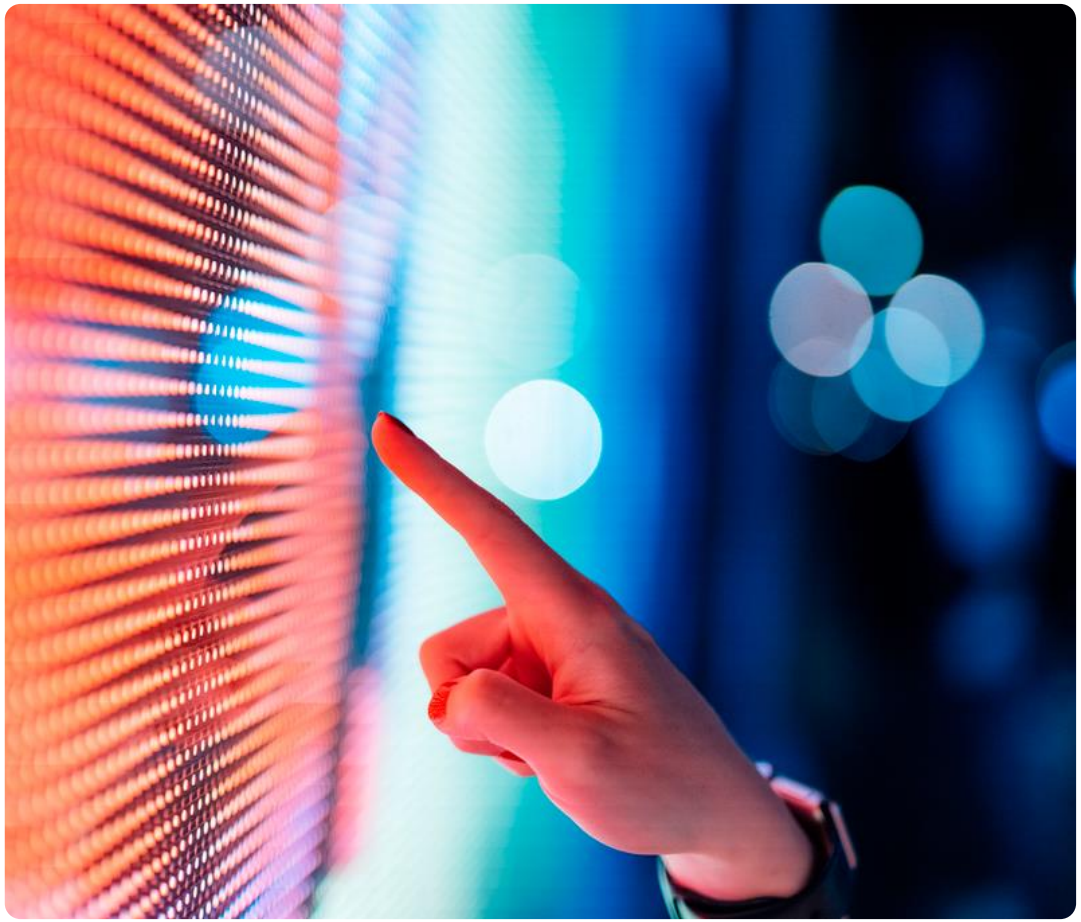
- Diagnosis and treatment – using lab results and genetic and other data to simulate a patient's digital twin to assist diagnosis and treatment
- Patient monitoring – to better understand disease progression, detect disease earlier and evaluate the effectiveness of treatment regimes
- Clinical trials – creating digital cohorts of patients to test new drugs with optimal dosage and reduce the number of 'real world' clinical trials needed
- Medical device design – the digital design of medical devices can be tested in a virtual world and safely optimized

Similarly, new digital approaches and technologies – electronic consent, remote patient monitoring and data collection and integrated point solutions – are enabling life sciences organizations to virtualize their clinical trial models and operations.

By adopting technology-enabled innovations like wearables, sensors, mobile apps and digital endpoints as well as advanced analytics, they've been able to significantly accelerate clinical trial timelines while reducing the cost associated with the secure and compliant collection and management of clinical and evidential data that's required for new therapy and device approval.

From study set-up to clinical data capture and analysis, they're using end-to-end digital platforms to support and enable smarter clinical trials.





Optimizing clinical trials with end-to-end digital support

From remote site activation to online patient recruitment and enrolment and the fast spin up of multiple trial sites in multiple locations and geographies, today's digital platforms are transforming how trials are conducted to deliver measurable productivity and performance gains.

Featuring wearable technologies and mobile sensors that make it easy for patients to self-test and capture digital biomarkers in real-time, **these end-to-end solutions are helping to dramatically reduce average study cycle times while boosting the acquisition of quality data sets**. Some of the measurable benefits reported by today's sponsors include:

25-40% reduction in clinical trial cycle times

15-25% lower handling and distribution costs

45-60% reduction in administration time and effort for the sponsor team

30% improved testing coverage

Improved compliance with regulatory requirements and CDISC standards

Reimagining a better future through data

Quantum computing has the potential to significantly accelerate, enhance and reduce the costs of data-rich R&D processes. While still an emerging technology, the long-term promise of quantum will prove transformative.

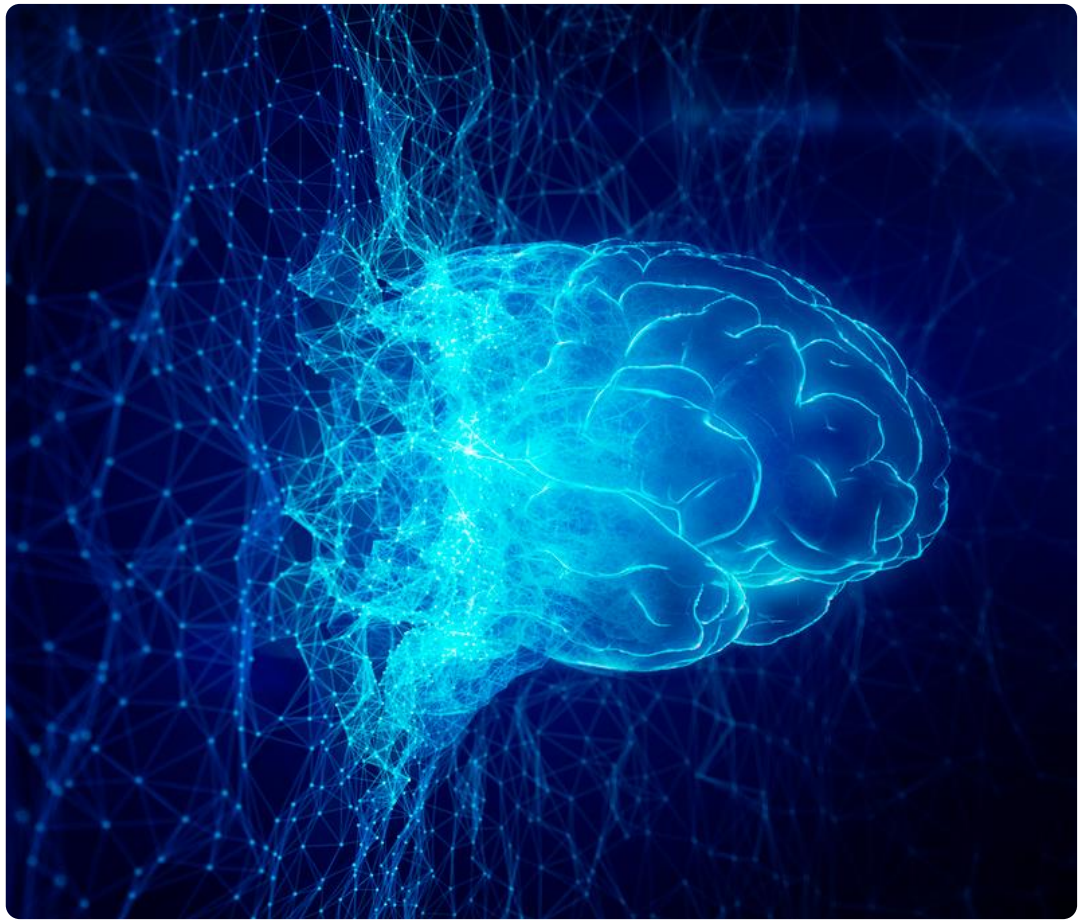
Meanwhile, today's sophisticated 3-D modelling techniques and high throughput data production technologies are already enabling a brave new world: bio-realistic simulations are set to usher in a new era of research innovations and precision healthcare.

The Virtual Human Project is already helping to reduce the need for animals in drug testing and simulating different processes of the human body. In the future, creating digital avatars of patients will enable clinicians and researchers to safely test the effects of drugs and treatments and select the most effective for each individual. Using these data powered simulations, life sciences will help realize the vision of personalized medicine and treatments at scale.

> [The virtual human is] a programme to make medicine more scientific and to know how to integrate that information in a way that's [actionable and] useful for clinical decision making."

Peter Coveney,

Professor of Physical Chemistry, Director, Centre for Computational Science at UCL,
Scientific Director of CBK Sci Con



The Virtual Human: How digital avatars are advancing healthcare

An EU-funded initiative has an ambitious vision: by simulating at scale how a person breathes, moves and functions, clinicians hope to be able to predict and prevent the likelihood of injuries or disease, designing treatment regimes that are personalized for individual patients.

The Virtual Human Project, developed and delivered by CompBioMed – a consortium led by the University College London – uses high performance computing to simulate real-time visualizations of different human body processes and makes these available to academic, clinical and industrial researchers to improve their understanding of human physiology and pathology. From this, researchers can derive predictive hypotheses and simulations to develop and test new therapies.

Offering the potential to transform medicine, the Virtual Human project is already unlocking valuable new insights. For example, making it possible to create a test bed setting that simulates the transport properties of dense cellular suspensions such as blood; or to visualize how best to deliver a drug.

It is hoped that virtual humans will make it possible to model the impacts of small changes in lifestyles and medications on health, ageing and quality of life; this will enable clinicians to predict who will need hip replacements or who is most at risk of strokes, while optimizing cancer treatments for individuals.

Data is helping life sciences organizations to reduce their eco-impact. Use of digital modelling and analytics, for example, reduces the need for wet lab experiments that are heavily reliant on the use of single plastics and refrigeration. The end-to-end digitalization of clinical trials offers huge potential for reducing the need for patients to constantly travel to hospitals or test sites. All of which adds up to a more sustainable R&D future.

Data analytics is also playing a central role in population health management, from helping to identify ‘at risk’ populations in need of care within communities, to enhancing how care is delivered by ensuring that outreach care is targeted and reaches the right people, at the right time and in the right way. All of which is helping to engineer a shift away from a ‘one size fits all’ care mentality and instituting ‘value-based care’, where the focus is on wellness and prevention that cuts healthcare costs and improves patient outcomes.

Bringing it together on the frontline

Today's life sciences organizations are already making use of cutting-edge technological innovations to unleash new ways of working – including agile co-creation partnerships – and achieve quicker and more accurate drug discovery. That includes gaining a deeper understanding of the interaction of drugs and their targets.

Enhanced data-driven collaboration between research scientists, clinicians and other stakeholders is already leveraging genomic and clinical data, combined with information on everything from environmental exposures to the lifestyle factors that can affect human health. This is creating a step-change in the predictive modeling of disease and treatment outcomes for individual patients while paving the way to ever more personalized, predictive and preventative healthcare and wellbeing for all.

About Eviden[1]

Eviden is a next-gen technology leader in data-driven, trusted and sustainable digital transformation with a strong portfolio of patented technologies. With worldwide leading positions in advanced computing, security, AI, cloud and digital platforms, it provides deep expertise for all industries in more than 53 countries. Bringing together 57,000 world-class talents, Eviden expands the possibilities of data and technology across the digital continuum, now and for generations to come. Eviden is an Atos Group company with an annual revenue of c. € 5 billion.

[1] Eviden business is operated through the following brands: Alia Consulting, AppCentrica, ATHEA, Atos Syntel, Bull, Cloudamize, Cloudreach, Cryptovision, DataSentics, digital.security, Eagle Creek, EcoAct, Edifixio, Energy4U, Engage ESM, Forensik, IDEAL GRP, IDnomic, In Fidem, Ipsotek, Maven Wave, Miner & Kasch, Motiv, Nimbix, Processia, Profit4SF, science+computing, SEC Consult, Visual BI, Worldgrid, X-Perion, zData