We are

committed to using data responsibly and transparently, and

engaging with patients and healthcare providers to help meet

patient needs. This includes managing data carefully, sharing

the results of our clinical studies, integrating patient insights

into our product development, and providing healthcare

professionals with relevant and accurate information when

they need it. – Data Protection and Legal issue(TK&U), Own initiative (B), A thorough and organised approach(B), sharing date with internal and external people(B)

n 2019, we combined our privacy training with the mandatory

Code of Conduct training. Approximately 32,000 individuals

completed our Privacy Foundation training, which includes

new hires, contingent workers, and those returning from leave

of absence. This explains our privacy principles to help them

understand how to apply them in their daily work. It also raises

awareness of why privacy matters for all those who handle

personal data. – Data Protection and Legal issue(TK&U) & Secure Environment(B)

Throughout 2019, people in key roles across

the organisation continued to undergo certification from the

International Association of Privacy Professionals (IAPP) to

increase expertise and enable us to make informed decisions

about handling personal data - – Data Protection and Legal issue(TK&U) & Secure Environment(B)

This collaboration

combines 23andMe’s genetic expertise and advanced data

science skills with GSK’s extensive scientific capabilities and

scale, to enhance the discovery and development of entirely

new medicines and potential cures. 23andMe customers

can choose to participate in research and contribute their

information to the unique and dynamic database for the purpose

of advancing scientific research. Participation is voluntary and

customers are required to consent affirmatively to their data

being used for research. Should they choose to participate,

their information is aggregated so no individual will be

identifiable to GSK. – The importance of the domain context for data analytics(TK&U) & Ability to work with external and internal people (B)

We also share anonymised patient-level data from our studies

with external researchers. We have listed 2,477 studies for data

sharing via www.vivli.org and www.clinicalstudydatarequest.com.

We launched this six years ago to facilitate innovative data-driven

research, and it is now used by multiple other study sponsors

and funders. External researchers are granted data access

based on a review of the scientific merit of their research

proposal by an independent panel. Access to GSK study

data has been approved for 157 proposals since 2013.

are strengthening our focus on patients’ needs by seeking

their insights across the business. We continue to support

several initiatives that are empowering patients to get more

involved in the development of medicines through training,

tools and dialogue – such as the European Patients’ Academy

on Therapeutic Innovation (EUPATI). – The importance of clearly defining customers requirements for data analysis (TK&U) & Ability to work with external and internal people (B)

To improve engagement with patients involved in our clinical

studies, we have developed patient engagement plans for key

assets and set up a dedicated patient panel as a key part of our

internal governance process. This allows patients to input into

the development of our research protocols, to improve patient

experience during the study, and we keep them informed about

the results after the study is completed. – The importance of clearly defining customers requirements for data analysis (TK&U) & Ability to work with external and internal people (B)

e ensure the inclusion of diverse populations in our clinical

studies so the data we generate represents as many people

as possible. By including individuals of different demographics

by age (elderly/frail and paediatric groups), sex, ethnicity and

race, we can capture potential variability in the responses

to our medicines and vaccines. This helps us to characterise

a more robust benefit-risk profile, generate greater insight for

the prescribing information and ensure the right patient gets

the right medicin

We also asked our third-party

preferred vendors to provide a plan for how they will deliver

improved recruitment and retention of diverse populations

for our full-service outsourced trials.- A thorough and Organised approach(B)

We disclose annually the individual

level of payments to HCPs when legally permitted, or otherwise

on an aggregate basis.

**Forbes**

Their conclusion was that the data at GSK R&D needed a major transformation. To lead it they brought in Mark Ramsey as the first head (and Senior Vice President) of R&D Data. He was charged with overseeing a transformation in how data and analytics were used across the organization. Vallance and his team had a vision for data within GSK R&D, which was to make it easier to access and use for exploratory analysis and decision-making about new medicines. GSK had been relatively good at making decisions with data, but the executives felt—and Ramsey quickly agreed—that the data within R&D was too siloed and fragmented to be used effectively for exploratory purposes. In particular, R&D data was kept within silos created for particular scientists, experiments, or clinical trials. Secondary analyses of it were almost impossible. – Use own initiative (B) & Ability to work independently and to take responsibility(B) & Maintain productive professional secure working Environment

integrating diverse data was clearly job one for Ramsey and his team. To guide and prioritize their activities, they identified over 20 use cases for what questions the scientists wanted to answer with R&D data, and eventually selected 10 as the focus. They were judged as having the greatest value, importance to key decisions, and role in addressing important scientific questions. More broadly, the goal of the work was to provide analytics-ready data of all kinds across R&D in a timely manner. – The data life cycle (TK&U) & The importance of clearly defining customer requirements for data analysis (TK&U), Maintain productive, professional and secure working environment(B)

GSK decided to employ the “probabilistic matching” approach used by Tamr to combine data across the organization into a single Hadoop-based data lake with three different domains. First would be “assays,” or data from experiments. Second would be clinical trial data. And third would be genetic data. The goal was to get 100% of the data into the lake within three months—an unheard-of objective using traditional data management approaches. But GSK was able to use the tools to understand the level of duplication and pull the data together in the desired timeframe. To work across the three domains, the R&D data team created an “integrated layer” on top of them with standardized ontologies; this was the only way to solve the use cases. – The processes and tools used for data integration (TK&U), the fundamentals of data structures, database system design, implementation and maintenance (TK&U)

* In the clinical trials domain, for example, Ramsey and his colleagues believed there was a massive amount of insight possible outside of the original goals for a particular trial. But combining trial data was difficult because there is a lot of variance in how they are conducted and their results recorded. But using industry standard formats, the data (originally in GSK internal formats) was ingested and mapped to the industry standard, and machine learning models learned the process. The team would feed in the source trial data, and what the target format should look like—and then let the machine go to work. Outcomes initially had 50/60% accuracy levels, and now in some domains they are at 100% accuracy. After the models were developed and refined, they could be applied to other data with relatively little human intervention—just some occasional judgments from an expert team. Logical and creative thinking skills – Logical and creative thinking skills (B) and Analytical and problem solving skills(B) the processes and tools used for data integration – (TK&U), the quality issues that can arise with data and how to avoid and/or resolve these (TK&U),
* There are other technologies in play in this implementation in addition to Hadoop and Tamr. GSK uses [Streamsets](https://streamsets.com/" \t "_blank) to move the data into the lake—a million pipelines for thousands of data sources. There is also a crawler that goes into every data source and extracts metadata and schemas from each one, and then creates loaders and pipelines to move data to the lake. There is a repository of every source data system and how often it changes, which is the mechanism for ingestion. GSK uses Apache Hive and Hbase to manage the data lake (on one of the largest Hadoop platforms in the world), and [Waterline Data](https://www.waterlinedata.com/) for cataloging. As a result of these technologies, GSK knows where all the data is, where it’s come from, and how often it’s updated. - the data life cycle, the fundamentals of data structures, database system design, implementation and maintenance, the processes and tools used for data integration (TK&U), A thorough and organised approach(B)
* There are significant reductions in times to get an answer to an *ad hoc* question. As GSK has rationalized clinical trial data, a team is focused on “clinical trial diversity” to make sure the company’s trials match the demographics of patients. Real world evidence from more than 30 sources is now rationalized to the industry standard—instead of being a catch-all category, as it is in many pharma firms. GSK is also using combined clinical trials data to reuse placebo patients where appropriate. They can simulate the control arm in some cases rather than having to give new patients placebos. - how to use and apply industry standard tools and methods for data analysis (TK&U), Logical and creative thinking skills (B)

**AI Businesses**

* GSK's automation of significant aspects of the drug discovery process is already, namely reduced cost and greatly improved speed. "A big component of what it takes to make a medicine can benefit from next-generation computing and machine learning," Karenann argues. "Approximately 1/3 of the total cost of developing a medicine (>$2.5bn) is spent during the time it takes from identifying your target (the process in your body you want to affect) to testing your molecule in humans for the first time. This process can take around five years. Our goal with artificial intelligence is to reduce the time to just one year in future." – Maintain productive, professional and secure working environment(B), how to use and apply industry standard tools and methods for data analysis(UK&U), Analytical and problem solving skills(UK&U)

The pharmaceutical giant are already developing partnerships with some of the many AI firms working on innovative healthcare solutions. "Our partnership with **[Exscientia](https://www.exscientia.co.uk)** - one of a growing number of partnerships - will combine their AI-enabled platform with our expertise in order to discover novel and selective small molecules for up to 10 disease-related targets – Ability to work with a range of internal and external people(B), the importance of the domain context for data analytics(UK&U)