Q&A

AWS Interoperability and Precision Medicine

An interview with Pat Combes

TO DELIVER patient-centered care, organizations in the heavily-regulated healthcare industry – from providers and payors to healthtech – need to increase the pace of innovation and unlock the potential of data, all while keeping health information secure and private. AWS empowers health organizations to improve patient outcomes, enable faster development of new therapeutics and treatment paths, and accelerate the digitization and utilization of their data with the broadest and deepest portfolio of cloud services and purpose-built partner solutions.

AWS introduced AWS for Health, an offering of curated AWS services, such as Amazon HealthLake, and AWS Partner Network solutions used by thousands of healthcare and life sciences customers globally to help accelerate transformation. AWS for Health provides proven and easily accessible capabilities that help organizations increase the pace of innovation, unlock the potential of health data, and develop more personalized approaches to therapeutic development and care. AWS for Health simplifies the process for healthcare and life sciences enterprises and innovative startups to identify industry-leading, cloud-based solutions across 16 critical solution areas in Healthcare, Genomics, and Biopharma.

Section 1: Precision Medicine and Personalized Medicine

The Journal recognizes a distinction between precision medicine and personalized medicine – in short, precision medicine concerns the design, development, and validation of a biomarker or drug whereas personalized medicine concerns the relationships between a patient and healthcare professional or consumer. Let's start with how

AWS for Health views precision medicine and personalized medicine in the context of its mission.

Q. Could you please provide your perspective on how AWS for Health views precision medicine and personalized medicine, respectively?

A. Covid-19 has dictated a clear direction for *precision medicine*. Guidelines for patient care and planning care have shifted from statistically significant outcomes gathered from a general population to outcomes based on the individual. Because cloud-based machine learning (ML) and artificial intelligence (AI) technologies accelerate the digitization and use of healthcare and life sciences data, clinicians can produce more precisely targeted therapies that wouldn't be possible using classical methods. They can make more accurate predictions of adverse events on the individual level. The information to do this has always existed, but human hands and eyes could never catalog it, sift through it, and find the critical connections to put it to use. Machines can.

The key to *predictive, personalized medicine is making sure that data flows to the right places securely and with patient privacy in mind*, and cloud technologies are one of the primary tools helping customers to achieve this outcome. Cloud data storage and management platforms like data lakes house structured and unstructured patient and administrative information in a centralized, secure location. AI and ML services automate data ingestion and processing and standardize that data by creating true data interoperability with syntactic interoperability (a common structure) and semantic interoperability (a common language). Achieving data interoperability will enable data to be shared with, and leveraged by, every entity in the value chain. At the same time, it will help anonymize individual patient data to uphold privacy, so organizations meet high security and compliance standards. Practitioners will be able to combine a patient's own data, at new levels of completeness and detail, with anonymized large-population patient data, including genomics. AI/ML-based learning models can use that large, detailed view to predict specific, individual health threats with unprecedented accuracy.

Making predictions is only part of the challenge. The same data interoperability and ML models that deliver the predictions where and when they're needed can also develop evidence-based treatment plans that apply precisely to the patient and condition at hand. This is a frontier of care that can't work if the flow of data is interrupted or disjointed anywhere in the process. As healthcare organizations take the necessary steps toward syntactic and semantic interoperability, the industry will be able to use data to place a renewed focus on personalizing the consumer health journey. Interoperability of healthcare data is key to being able to identify the unique needs of each individual, which is essential to creating a frictionless and more personalized patient experience.

Genomics and healthcare organizations strive to bring personalized medicine to the point-of-care by combining clinical information with genomic data. For instance, the Melbourne Genomics Health Alliance, a collaboration of 10 leading hospitals and research organizations, is working to make genomics a part of everyday medicine in the Australian state of Victoria. To avoid data dispersion and needless duplicative efforts in systems building, the Alliance leveraged cloud technology to build GenoVic, a shared clinical system that facilitates the management and use of genomic sequencing data by numerous laboratories, clinicians, patients, and researchers across multiple, independent organizations.

> *"Our goal is to support organizations undergoing transformation from bench-top to the point of care – helping them to reinvent how they collaborate, make data-driven clinical and operational decisions, enable precision medicine, accelerate therapy development, and decrease the cost of care."*

Q. What are AWS's goals in the next five years with this build-out?

A. AWS will continue to work with our healthcare and life sciences customers to innovate, transform, and tackle important challenges – 90 percent of what we build is driven by what customers tell us matters, and the other 10 percent are things we hear from customers, so we try to read between the lines and invent on their behalf.

Our goal is to support organizations undergoing transformation from bench-top to the point of care – helping them to reinvent how they collaborate, make data-driven clinical and operational decisions, enable precision medicine, accelerate therapy development, and decrease the cost of care.

Section 2: Interoperability is Key to Functionality and Precision Medicine

As noted in the introduction, AWS has put a premium on interoperability. To that end, a Fast Healthcare Interoperability Resources (FHIR) interface has been adopted by AWS (as have other companies working in the healthcare space). More specifically, *FHIR Works on AWS* was introduced as a new AWS Solutions Implementation with an open-source software toolkit to create an interface for existing healthcare applications and data.

Q. Can you describe the efforts to design the interface for more compatible data structures – e.g., common glossary for terms, units for measurements, front-end inputs?

A. Healthcare organizations are creating vast volumes of patient information every day, and the majority of this data is unstructured, such as clinical notes, lab reports, insurance claims, medical images, recorded conversations, and graphs. This data must be aggregated, structured, and normalized before the data can provide customers with valuable insights, and that is often a time-consuming and error-prone process.

As I mentioned above, achieving true healthcare interoperability requires understanding, evaluating, and solving issues in the underlying syntactic and semantic characteristics of the data. Syntactic interoperability requires a common structure so that data can be exchanged and interpreted between health information technology (IT) systems, while semantic interoperability requires a common language so the meaning of data can be transferred along with the data itself.

The industry has made meaningful progress on this front. The FHIR open standard has emerged from the nonprofit HL7 organization to act like a lingua franca, providing a universal adaptor for sharing clinically relevant data easily and securely from any EHR or clinical system and allowing software developers to build high-quality applications. FHIR enables healthcare solutions providers to build secure, compliant, and scalable solutions for the delivery and exchange of medical information across the healthcare industry.

It's worth noting that beyond data structure, data stores have evolved to enhance and accelerate data management and analysis. New HIPAA-eligible data lake services optimized for healthcare and life sciences allow organizations to ingest, store, query, and analyze their health data at scale. These services use machine learning to understand and extract meaningful medical information from unstructured data, and then organizes, indexes, and stores that information in chronological order. The result provides a holistic view of patient health.

Amazon HealthLake is a HIPAA-eligible service for healthcare and life sciences organizations that aggregates an organization's complete data across various silos and disparate formats into a centralized AWS data lake and automatically normalizes this

information using machine learning. The service identifies each piece of clinical information, tags, and indexes events in a timeline view with standardized labels so it can be easily searched. It also structures all of the data into the FHIR industry standard format for a complete view of the health of individual patients and entire populations. As a result, Amazon HealthLake makes it easier for customers to query, perform analytics, and run machine learning to derive meaningful value from the newly normalized data.

For example, AWS is supporting the build-out of Realm IDx's, LATTICE, a groundbreaking integrated diagnostic data platform that combines genomics, pathology, and radiology data along with other critical information to uncover new, clinically relevant biomarkers and create the next generation of diagnostic tests. Through LATTICE, Realm IDx and the PRECEDE Consortium are also able to integrate data across disparate clinical sites to improve the survival rate for pancreatic cancer patients.

Realm IDx is leveraging machine learning services, such as Amazon HealthLake, along with AWS storage, compute, database, and analytics services, to help clinicians deliver the right care, to the right patient, at the right time. Amazon HealthLake will allow Realm IDx to unlock the real power of this multi-modal approach to find novel associations and signals in the organization's data. It will provide Realm IDx's expert team of data scientists and developers the ability to integrate, label, and structure this data faster, and discover insights that our clinicians and pharmaceutical partners require to truly drive precision medicine.

Q. Notes from healthcare providers tend to be unstructured. What processes does AWS use to structure these notes to be interpretable across platforms – e.g., natural language processing?

A. We currently have two services that help increase both the ability to interpret voice-based clinical data and the ability to search, capture, and analyze that content in a contextually significant way:

Amazon Transcribe Medical is a HIPAA-eligible, machine learning automatic speech recognition (ASR) service that allows developers to add medical speech-to-text capabilities to their applications. Amazon Transcribe Medical provides accurate and affordable medical transcription, enabling healthcare providers, IT vendors, insurers, and pharmaceutical companies to build services that help physicians, nurses, researchers, and claims agents improve the efficiency of medical note-taking. With this service, customers can complete clinical documentation faster, more accurately, securely, and with less effort, whether in a clinician's office, research lab, or on the phone with an insurance claim agent.

Speech-to-text technology is evolving and improving accuracy through precision tuning, customization, and the ability to filter content. Amazon Transcribe Medical supports APIs that uniquely understand medical conversations. It can be tuned to specific audio formats, including low-fidelity phone and voice recorder audio commonly used in clinical settings, and use custom vocabularies germane to various medical practices and specialties to generate more accurate transcriptions. Transcript output can also be made ready for downstream activities such as call transcript analysis, subtitling, and content search. Content can be redacted automatically to identify and remove sensitive, personally identifiable information (PII) from transcripts, and transcription data is secured through custom key management services.

"Amazon Transcribe Medical is a HIPAA-eligible, machine learning automatic speech recognition (ASR) service that allows developers to add medical speech-to-text capabilities to their applications."

In addition to facilitating the sharing and interpretability of voice notes, we have a natural language processing (NLP) service, Amazon Comprehend Medical, a HIPAA-eligible, machine learning service that allows developers to process unstructured medical text and identify information such as patient diagnosis, treatments, dosages, symptoms and signs, and more. Comprehend Medical helps healthcare providers, insurers, researchers, and clinical trial investigators as well as healthcare IT, biotech, and pharmaceutical companies improve clinical decision support, streamline revenue cycle and clinical trials management, and better address data privacy and protected health information (PHI) requirements.

The service recognizes entities, key phrases, language, sentiments, and other common elements in a document. Clinicians can use the service to examine a large body of documents that are organized into topics or clusters based on similar keywords, creating the ability to search through volumes of anonymized patient note data to find commonalities in symptoms, diagnoses, and prescribed therapies.

For instance, one challenge that Cambia Health Solutions (Cambia) previously faced was getting the data from clinical systems and hospitals and adding it to individuals' overall health profiles. Traditionally, this data comes in fax or PDF format, which can be time consuming to integrate into the service manually. Cambia is using Amazon Comprehend Medical to develop proofs of concept to address this challenge. Previously, Cambia's artificial intelligence and data science team's capacity was consumed with managing the base infrastructure and building tools from scratch. Through using Amazon Comprehend Medical and Amazon SageMaker, Cambia was able to shift 30–40 percent of its data science resources from supporting operations to developing solutions that directly impact business and consumer outcomes.

Q. Precision medicine clearly requires the ability to access these various data sets. Can you cite an example of how interoperability has been used by AWS For Health to advance a precision medicine solution?

A. One of the best examples of how ML is enabling precision medicine is Moderna's ability to accelerate every step of the process in developing a messenger ribonucleic acid (mRNA) vaccine for COVID-19. Moderna began work on its vaccine the moment the novel coronavirus' genetic sequence was published. Within days, the company had finalized the sequence for its mRNA vaccine in partnership with the National Institutes of Health. Moderna was able to begin manufacturing the first clinical-grade batch of the vaccine within two months of completing the sequencing – a process that historically has taken up to ten years- with AWS technology.

Q. What other aspects of interoperability are critical to AWS for Health?

A. A promising area is the development of APIs for clinical exchange and administrative automation. Healthcare HIPAA-compliant interoperability APIs are helping the healthcare industry develop and use open standards, such as FHIR, for easy exchange of information, freeing providers, payers, and patients themselves from the confines of proprietary data formats and systems. The industry is developing a reference implementation for serverless implementation of FHIR APIs and access to over 130 HIPAA-eligible features and services with a wide range of certifications and attestations, covering compliance programs worldwide.

A good example of how standards and open APIs can help improve patients' health outcomes and overall experience are the FHIR-enabled storage and APIs created by Seattle-based Fred Hutchinson Cancer Research Center that enable care coordination between oncologists and primary care providers. "Fred Hutch" used the APIs to provide patients with an application to support their regimes, including appointment follow-up and engagements

with multiple providers, providing visibility into disease progression.

Beyond interoperability, ensuring data is findable, accessible, and reusable is critical. For example, the AWS Data Exchange (ADX) makes it easy to find, subscribe to, and use third-party data in the cloud. One qualified data provider using ADX includes Change Healthcare, which processes and anonymizes more than 14 billion healthcare transactions and $1 trillion in claims annually.

Furthermore, the Registry of Open Data on AWS (RODA) makes it easy for people to find datasets that are publicly available through AWS. When data is shared on AWS, anyone can analyze it without needing to download or store it, which allows users to spend more time on analysis rather than acquisition. In just 11 days, the University of British Columbia (UBC) accessed and analyzed genomic sequencing data in the NIH Sequence Read Archive (a 20 PB resource) on AWS and discover 130,000 novel viruses, including 7 new coronaviruses – that's 10 times more than were identified over the past century of virology research.

Lastly, AWS Identity and Access Management (IAM) provides fine-grained access control across all of AWS. Customers can specify who can access which services and resources, and under which conditions, as well as manage permissions to their workforce and systems to ensure least-privilege permissions.

Section 3: Internal Skill Sets and Organizational Structure at AWS for Health.

In parallel with putting in place a platform comprised of functional modules to create and support its healthcare enterprise, AWS for Health has brought in leaders and teams with skill sets relevant to drug development. Bringing medical and biotechnology (med/biotech) skills into a technology company requires blending skills and groups in the organization – e.g., those with pharmacy degrees or bioinformatics experience into a software-focused department. Success in these efforts calls for informed planning, focused recruitment, and inter-group communication, especially between the technology and med/biotech teams.

Q. Despite the array of partners, gaps can form in areas of *data coverage*. What steps does the AWS for Health team take to ensure ongoing updates and coverage meet both system and medicinal requirements?

A. Precision medicine benefits from learning from the broader population. We work hand in hand with our customers and our partners across the spectrum of genomics, imaging, clinical data, and more to understand where these gaps are, enhance the solutions we offer, and improve access to populations. Our roadmaps are driven by what our customers tell us.

We've developed and launched initiatives to address pressing health challenges facing society – like the COVID-19 pandemic and health inequity – in which technology can make a positive difference. We've worked with our customers and partners closely on these efforts to bring innovation, research, and medical breakthroughs to the forefront of these health-related challenges.

> *"By listening and working with the stakeholders, customers, and partners, we work backwards from the problem we are trying to address."*

For example, the AWS Diagnostic Development Initiative (DDI) is a two-year, $20 million commitment to accelerate research and innovation to advance the collective understanding and detection of COVID-19 and other infectious diseases to mitigate current and future outbreaks. Through DDI, DNAstack is collaborating with AWS to harmonize, process, and share viral genome sequences deposited into NCBI SRA. Sequencing data is uniformly re-processed by a secondary analysis pipeline executed through Amazon Genomics CLI and made freely available with support from the AWS Open Data Sponsorship Program. Another great example of innovation through DDI is Karius, a company that uses genomics data and artificial intelligence to advance infectious disease diagnostics. By mapping a patient's microbial landscape from a single, non-invasive blood draw, physicians can gain genomic insights to more quickly and accurately treat infectious diseases. The company relies on a host of AWS products to provide the compute infrastructure to analyze the genomic data from cell-free DNA sequencing.

We've also launched the AWS Health Equity Initiative, a $40 million commitment to harness the power of the cloud to advance health equity globally. One exciting example to highlight is the work we are doing with nonprofit Genetic Alliance, and specifically the organization's iHope Genetic Health (iGH) program, which provides clinical whole genome sequencing (cWGS) diagnostic services to the undiagnosed in low to middle income countries. One novel and critical part of the program is the requirement that the data generated be given to the individuals and their families. The platform for these and other clinical data, LunaDNA, provides the processes and mechanisms for individuals to control their own data and grant access to their data for research that matters to them. Luna also enables community engagement, recognizing that the health needs and situations of individuals and families changes over time. The iHope Genetic Health program, made possible by a $120 million contribution of reagents and sequencers from Illumina, will also provide resources to patients, clinicians, and laboratories to accelerate diagnoses and enable follow-up.

Q. What insight could you provide for how AWS for Health is balancing the optimal internal mix of technology, science, and medicine expertise?

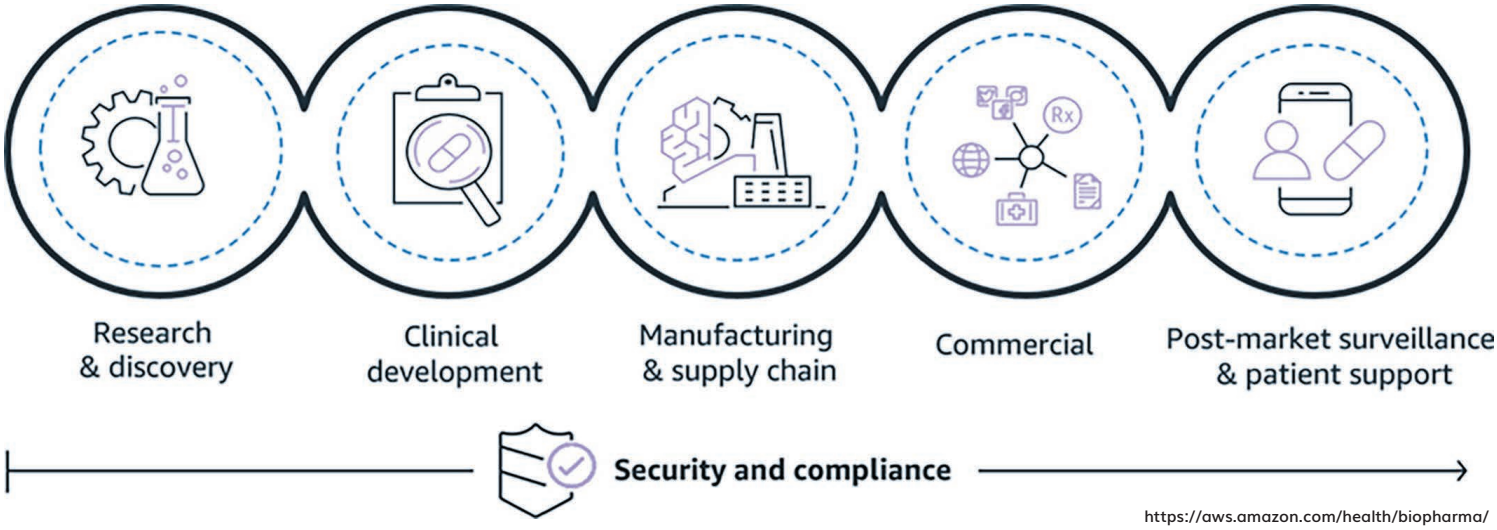
A. By listening and working with the stakeholders, customers, and partners, we work backwards from the problem we are trying to address. Technology is a tool that clinicians and scientists use to create outcomes for our patients. By working with our partners and customers, both long and short term, we are able to create a sense of familiarity and understand their problems.

Section 4: Virtual Pharma via Internal Plans and External Partners

AWS for Health has brought together external partners and internal capabilities that cover use cases from supporting discovery to post-market pharmacovigilance be viewed as a virtual drug company (https://aws.amazon.com/health/biopharma/). The accompanying figure depicts the classic linear process, but the drug development process tends to be tortuous and recursive.

We have a breadth of partners covering each area of the genomics workload from data transfer and storage, workflow automation, data aggregation and governance, and tertiary analysis including annotation and interpretation of results. Partner capabilities include the accelerated compute of secondary analysis pipelines like NVIDIA Parabricks and Illumina DRAGEN.

Often organizations work with multiple partners to meet their specific needs. To identify potential new disease therapies, find new gene targets, and better understand neurological disease biology, Biogen worked with DNAnexus for genomic sequencing and processing in addition to Databricks to design a software solution stack that harnessed AWS to analyze the data efficiently and securely. All parties operating on AWS simplified the data transfer process. Another example – Genomics England is transitioning from project to platform, leveraging AWS Partners Lifebit and Kainos to migrate its data to a secure platform for global accessibility, with the goal of making genomic healthcare a reality.



https://aws.amazon.com/health/biopharma/

To help customers migrate to the cloud, optimize operations, and expand capabilities, AWS for Genomics features dozens of vetted healthcare and life sciences technology and consulting partners. AWS Life Sciences Competency Partners, including DNAnexus, Seven Bridges, Lifebit, and Qiagen, have demonstrated technical expertise and customer success in building solutions on AWS across the genomics workflow, designed to help translate samples into actionable insights and scientific breakthroughs. Vetted consulting partners, such as Accenture, BioTeam, PWC, Slalom, and Leidos, offer services to help customers of all types and sizes accelerate their journey to the cloud and get the most out of AWS services. The AWS Marketplace offers a curated digital catalog of third-party software, services, and data that makes it easy to find, deploy, and manage genomics-specific solutions on AWS.

Q. How has AWS for Health managed communication with these disparate parties?

A. By listening to our customers and partners, as well as joining/supporting pan industry groups (e.g., HL7.org).

Q. Have any partners considered using these capabilities to develop their own drugs from start to ready for clinical trials?

A. Machine learning plays a significant role in the distillation of genomic knowledge into actionable clinical knowledge, and there are two main areas where we've begun to see ML drive change: accelerating scientific breakthroughs and driving clinical applications. We're starting to see organizations leverage ML to rapidly extract insights from large cohorts of data and identify new correlations. With ML, biopharma organizations can bring genomics into drug discovery and development to migrate from one-size-fits-all drugs to personalized treatments. And at the bedside, ML can help healthcare providers rapidly find the answers needed to provide an individualized treatment course based on the patient's genetic profile. One organization that comes to mind is the Global Alliance for Genomics & Health (GA4GH), a nonprofit alliance that brings together more than 600 leading organizations across healthcare, research, life science, and technology to create scalable frameworks and standards to enable interoperability across the global genomics community. The Alliance shares genomic data in the cloud with researchers and medical professionals across the globe, which leads to the discovery and development of drugs, flexibility of collaboration, and more.

"Machine learning plays a significant role in the distillation of genomic knowledge into actionable clinical knowledge, and there are two main areas where we've begun to see ML drive change: accelerating scientific breakthroughs and driving clinical applications."

Q. Would you like to offer any closing comments?

A. For the last decade, organizations have focused on digitizing healthcare. In the next decade, making sense of all this data will provide the biggest opportunity to transform care. However, this transformation will primarily depend on data flowing where it needs to, at the right time, and supporting this process in a way that is secure and protects patients' health data.

It comes down to interoperability. It may not be the most exciting topic, but it's by far one of the most important, and one the industry needs to prioritize. By focusing on interoperability of information and systems today, we can ensure that we end up in a better place in 10 years than where we are now. And so, everything around interoperability, around security, around identity management, differential privacy, are likely to be part of this future.

As the industry moves toward value-based care, artificial intelligence and machine learning, paired with data interoperability, have the potential to improve patient outcomes while driving operational efficiency to lower the overall cost of care. By enabling data liquidity securely, and supporting healthcare providers with predictive machine learning models, clinicians will be able to seamlessly forecast clinical events, like strokes, cancer, or heart attacks, and intervene early with personalized care and a superior patient experience. JoPM

Pat Combes

Patrick Combes joined Amazon Web Services in April 2016 and is responsible for AWS' worldwide technical strategy in Healthcare and Life Sciences (HCLS). Patrick helps develop and implement the strategic plan to engage customers and partners in the industry and leads the community of technically focused HCLS specialists within AWS. Patrick has a B.S. in Computer & Electrical Engineering from the University of Illinois at Champaign-Urbana with additional graduate work focused on programming language design. He also holds several associate and professional certifications for architecting on AWS.