



Leveraging Cloud Capabilities for Clinical Data Sharing

Sarper Guerel
Solutions Architect
AWS

Agenda



- What is Cloud?
- Global Footprint
- Cloud Usage Examples
 - Compute
 - Storage
 - Biostatistical Programming
- Data Sharing Examples
 - PPFL
 - Data Exchange
 - DataZone
 - CleanRooms

What is Cloud Computing?



- Core value proposition of Cloud
- Benefits
- Agility and Innovation
- Security and Compliance

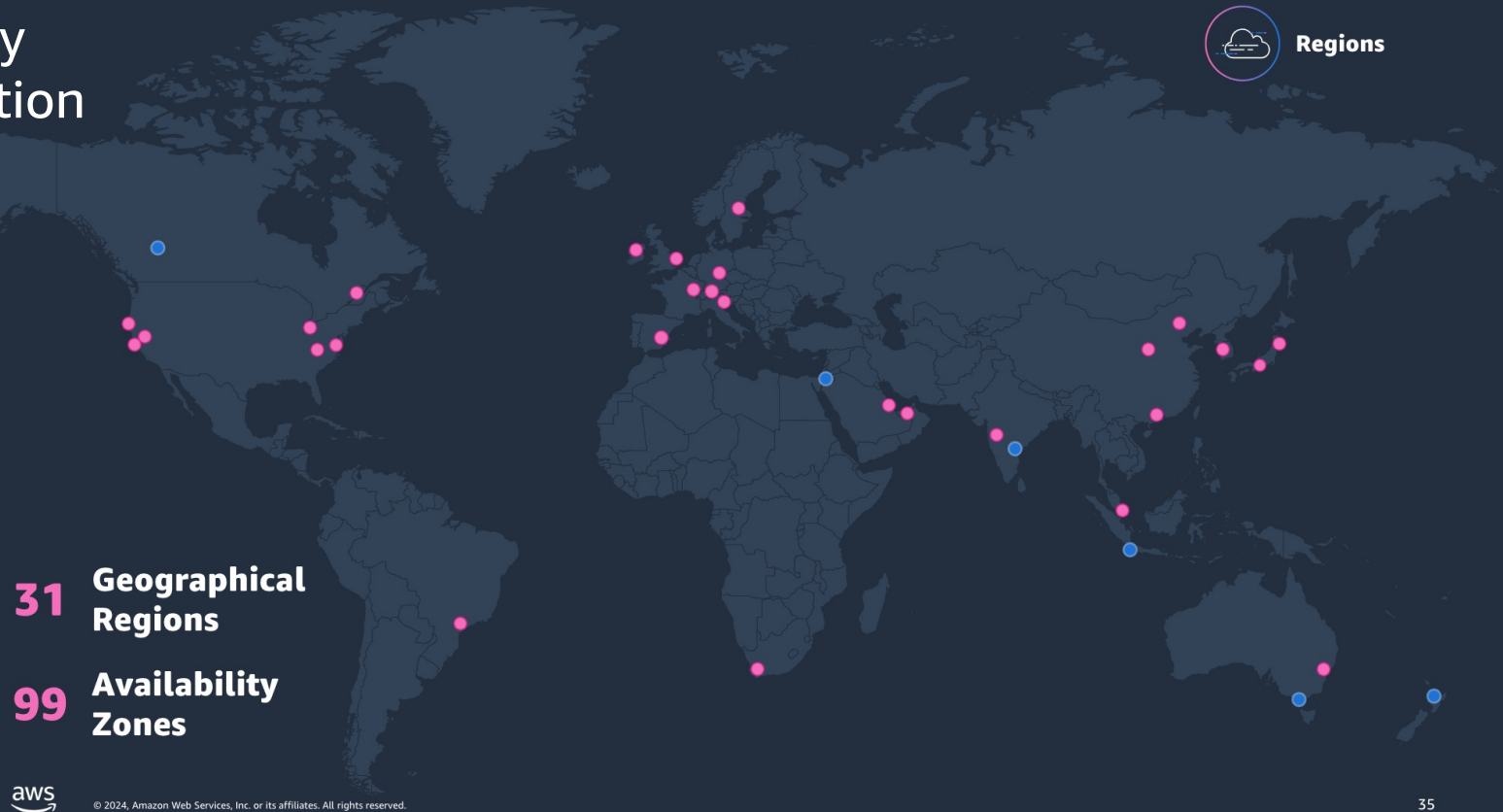


<https://youtu.be/JLkkc5t0tQA>

Cloud: Global Footprint



- Choice to build for
 - Latency
 - Cost
 - Data Sovereignty
 - Team Collaboration
 - More...



• Cloud Usage Examples: Compute



- HPC (e.g. supercomputer)
- Containers (e.g. Kubernetes)
- Purpose-Built Silicon (e.g. Trainium, Graviton, Inferentia, GPUs)
- Serverless Compute (e.g. Lambda)
- Memory Intensive Computing (e.g. 32 TB RAM compute instances)



• Cloud Usage Examples: Storage



- Block
 - Choice of cost/performance (e.g. SSD vs HDD)
- File
 - Choice of features/performance(e.g. FSX Lustre, FSX OpenZFS)
- Object
 - Virtually unlimited storage (e.g S3)
 - Immutable Long-Term Archiving (e.g. Glacier Deep Archive)



• Cloud Usage Examples: Biostatistical Programming



- Managed Posit **R-Studio** environments
- Managed **SAS** environments (Grid, Viya etc.)
- Scalability and Customization
- Connected to Data Sharing capabilities



• Data Sharing leveraging Cloud: Aspects



- Share In-Place vs. Controlled/Uncontrolled Copy
- Lineage and Provenance
- Identity and Access Management
- Automation and Self-Service
- Monitoring and Observability
- Performance and Scalability
- Immutability
- Versioning



• Data Sharing leveraging Cloud Example: PPFL



- Privacy Preserving Federated Learning (PPFL)
 - E.g. AWS PPFL reference architecture

Federated learning on heterogenous data using chest CT

1. Proof-of-concept of FL on large-scale, heterogenous medical datasets
2. Targeting 3D chest CT lung evaluation across global geographic centers
3. FL performs at par with or better than other methods

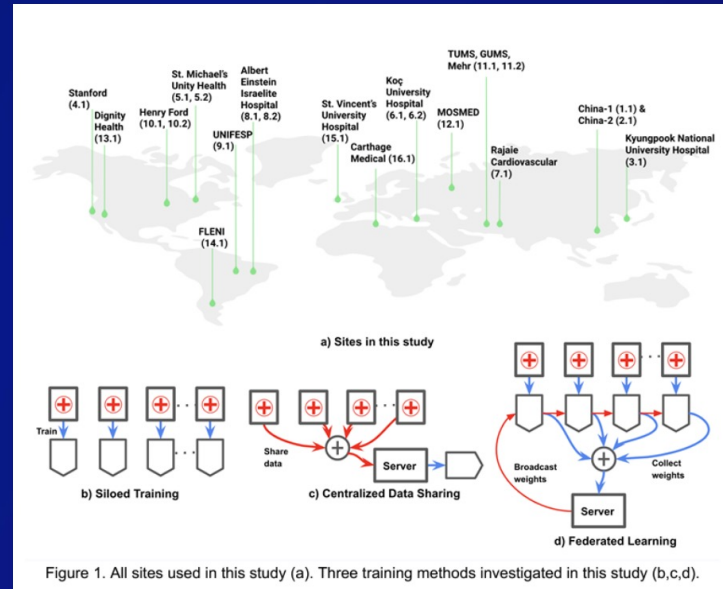


Figure 1. All sites used in this study (a). Three training methods investigated in this study (b,c,d).

16 sites (21 participating hospitals) across 5 continents, representing >10,000 patients and >1 million images

Source: Yeom, et al, 2023

• Data Sharing leveraging Cloud Example: DataZone

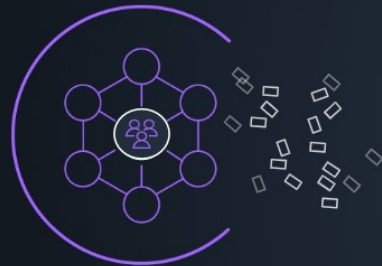


- Implementing the „Data as Product“ Concept
 - E.g. Amazon DataZone

Core components of Amazon DataZone

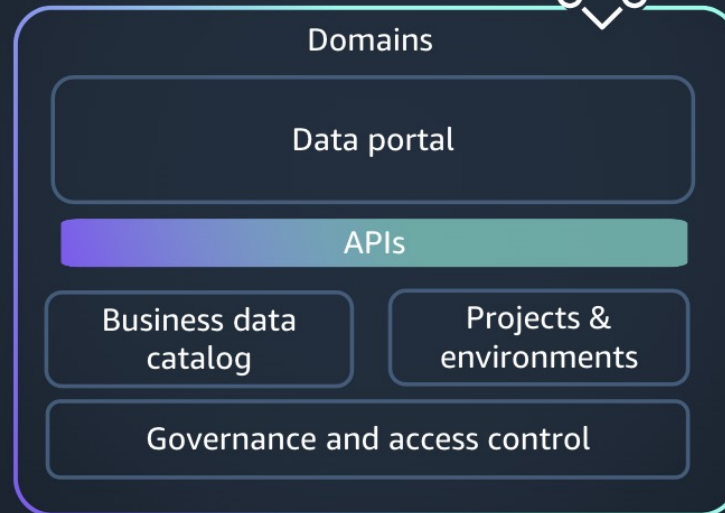
ENABLE COMPANIES TO TREAT DATA AS A PRODUCT

Amazon DataZone



Data producers

Bring data from different sources, and across accounts/regions



Data consumers

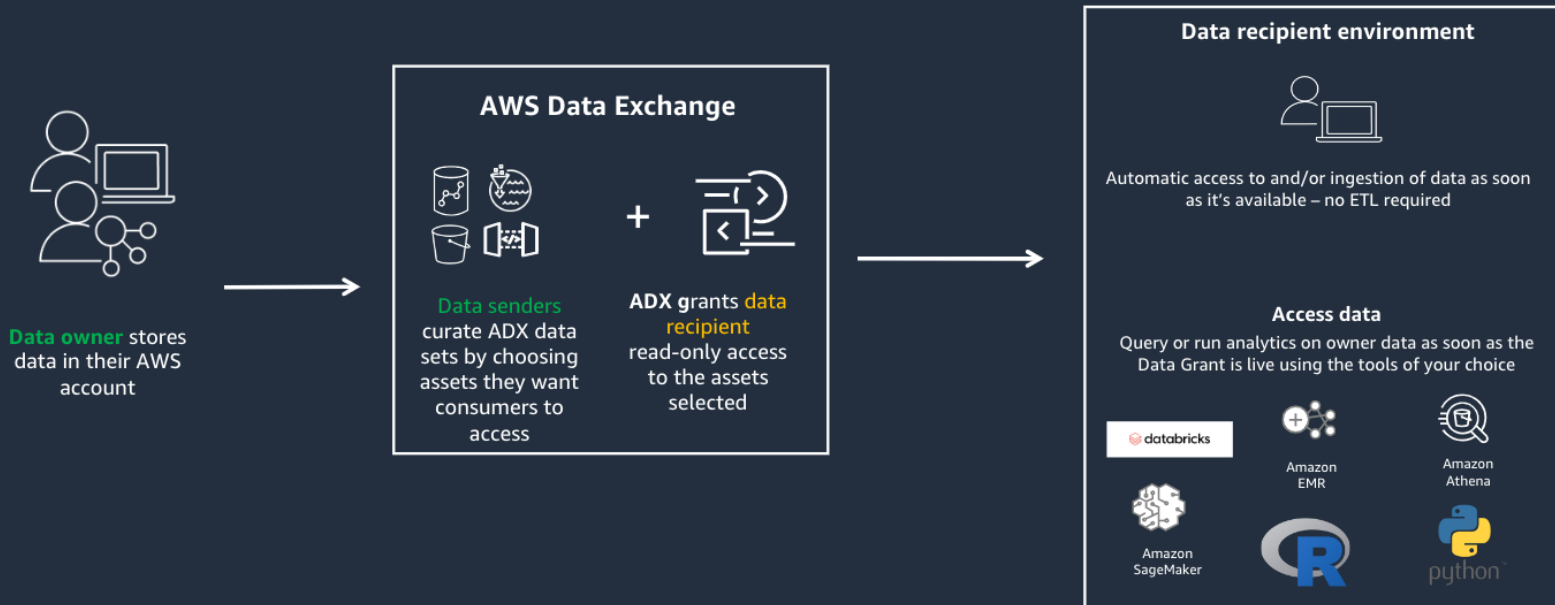
Simplify access, collaboration and consumption using different tools

© 2023, Amazon Web Services, Inc. or its affiliates. All rights reserved.

• Data Sharing leveraging Cloud Example: AWS Data Exchange



- Sharing Data across organizational boundaries
 - E.g. AWS Data Exchange



• Data Sharing leveraging Cloud Example: Clean Rooms



- Managed End-2-End Data Collaboration Capabilities
 - E.g. AWS Clean Rooms





Thank You

Connect me on LinkedIn (QR Code on right)

<https://www.linkedin.com/in/sarper-gurel-9684762>

