AWS for Data Spaces & Medical Research

Webinar and Discussion Panel
for GÉANT Members

May 10th, 2023

Agenda

Time	Topic	Speaker
10:00-10:15	1. Introduction	Agata Jablonka, AWS
10:15-10:25	2. Data Spaces and AWS technology	Syrine Souissi, AWS
10:25-10:35	3. EHDS: high-level overview	Khrystyna Shlyakhtovska, AWS
10:35-10:45	4. Trusted Research Environment on AWS	Tim Cutts, AWS
10:45-11:00	5. Guest speaker	Professor Edward Curry, University of Galway | Director, Insight SFI Research Centre for Data Analytics
11:00 – 11:30	6. Panel discussion	Agata Jablonka, AWS

1. Introduction

2. Data Spaces and AWS Technology

AWS customers want to share data with a trusted community of organizations, but face challenges

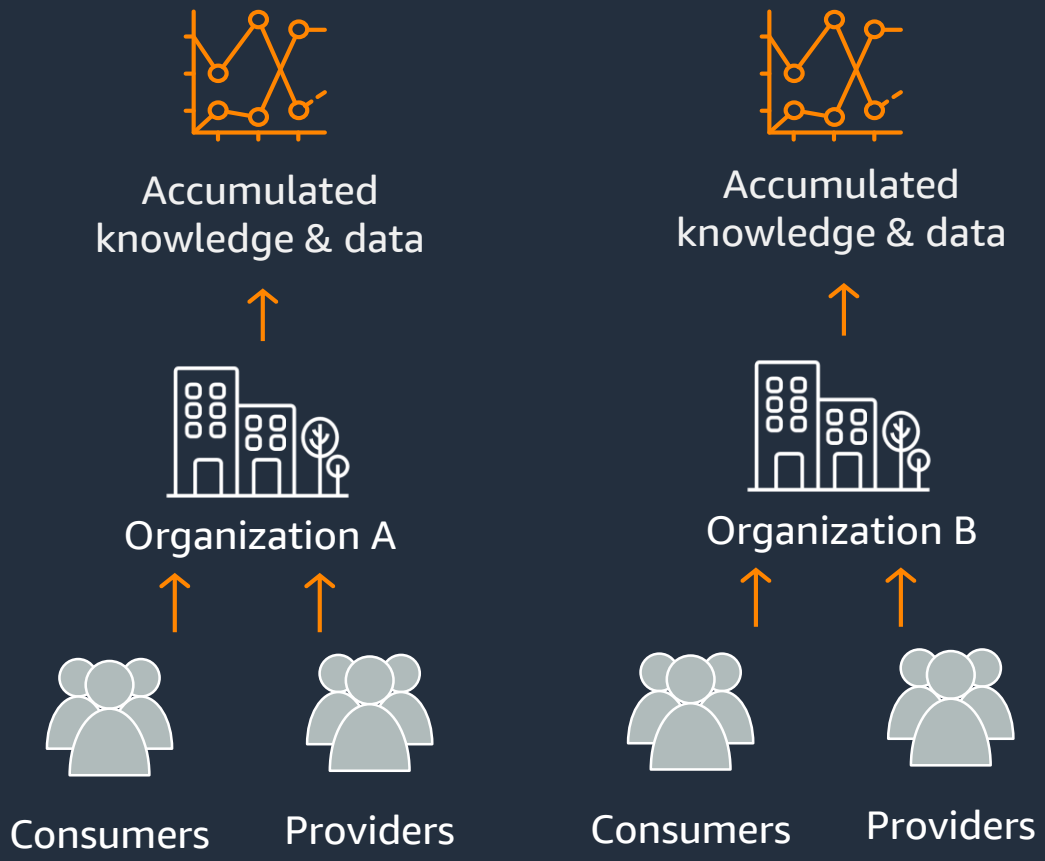
Data siloes

Lack of trust

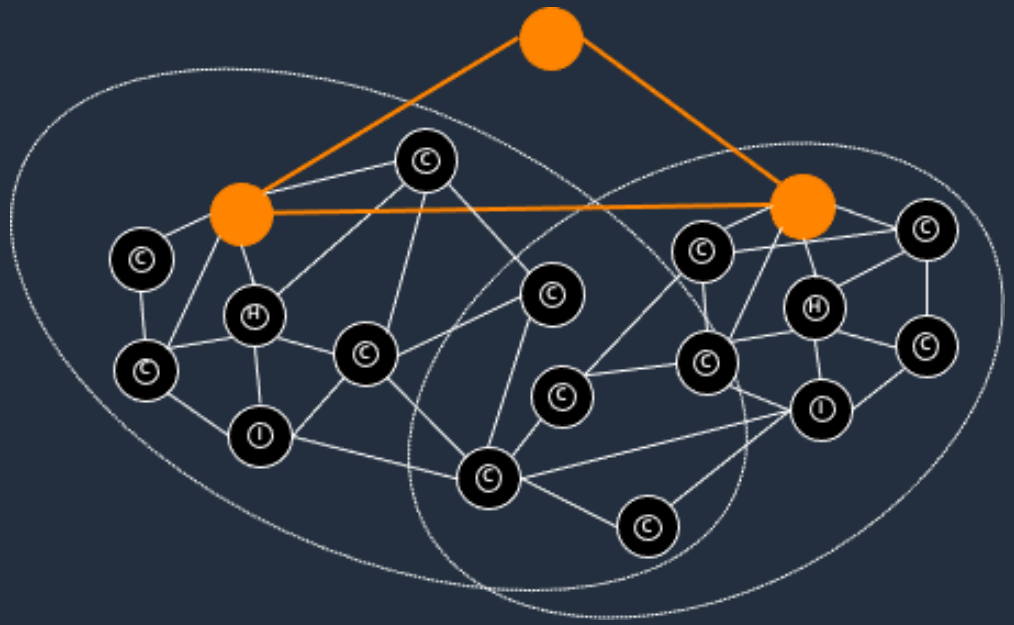
Lack of standards

Data spaces for inter-organizational data sharing

Data silos across organizations

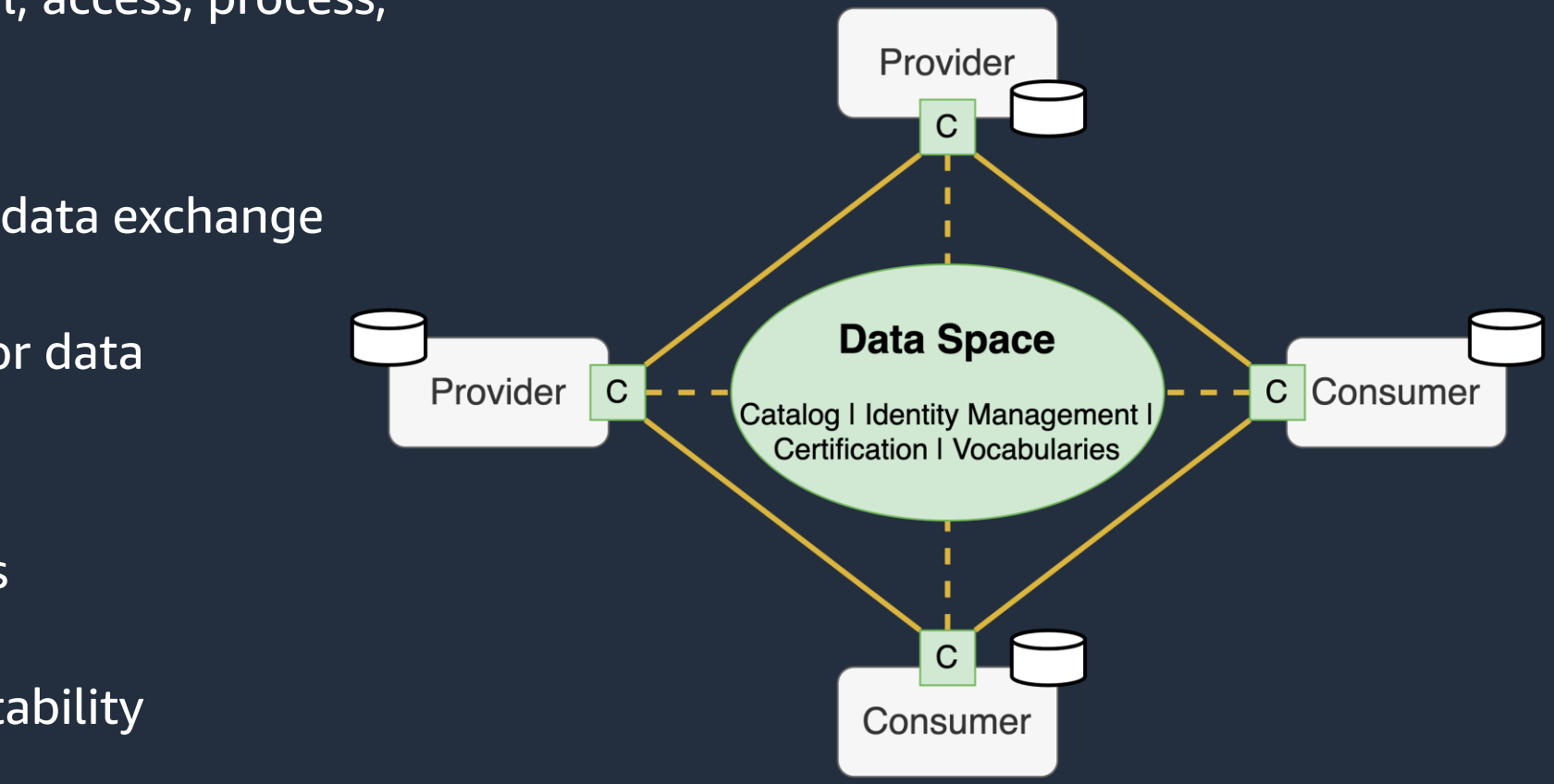


Data Space

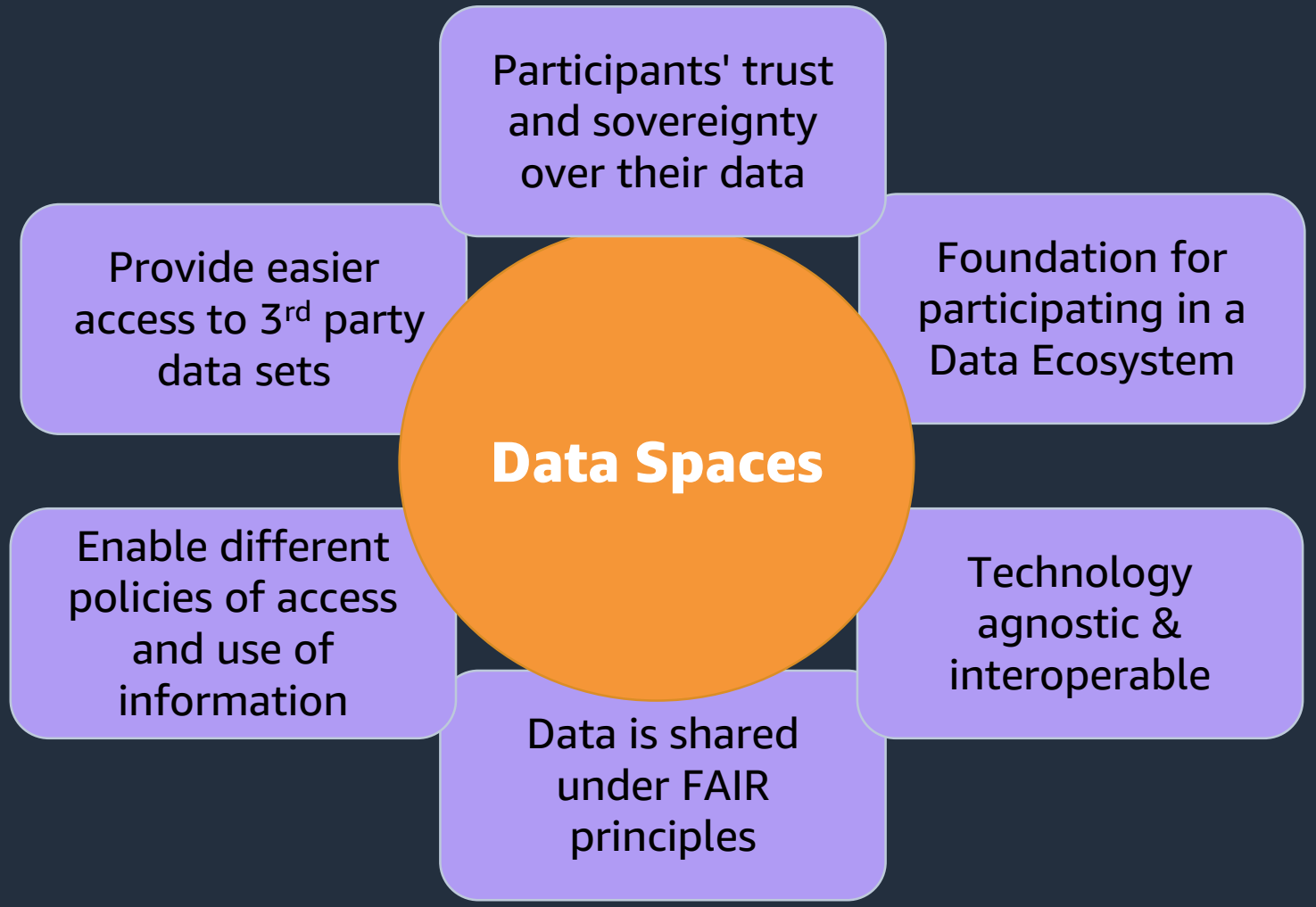


Key requirements of a data spaces

1. Participants want to pool, access, process, use and share data
2. Open and decentralized data exchange
3. Software functionality for data sovereignty and trust
4. Standard building blocks
5. Interoperability and portability

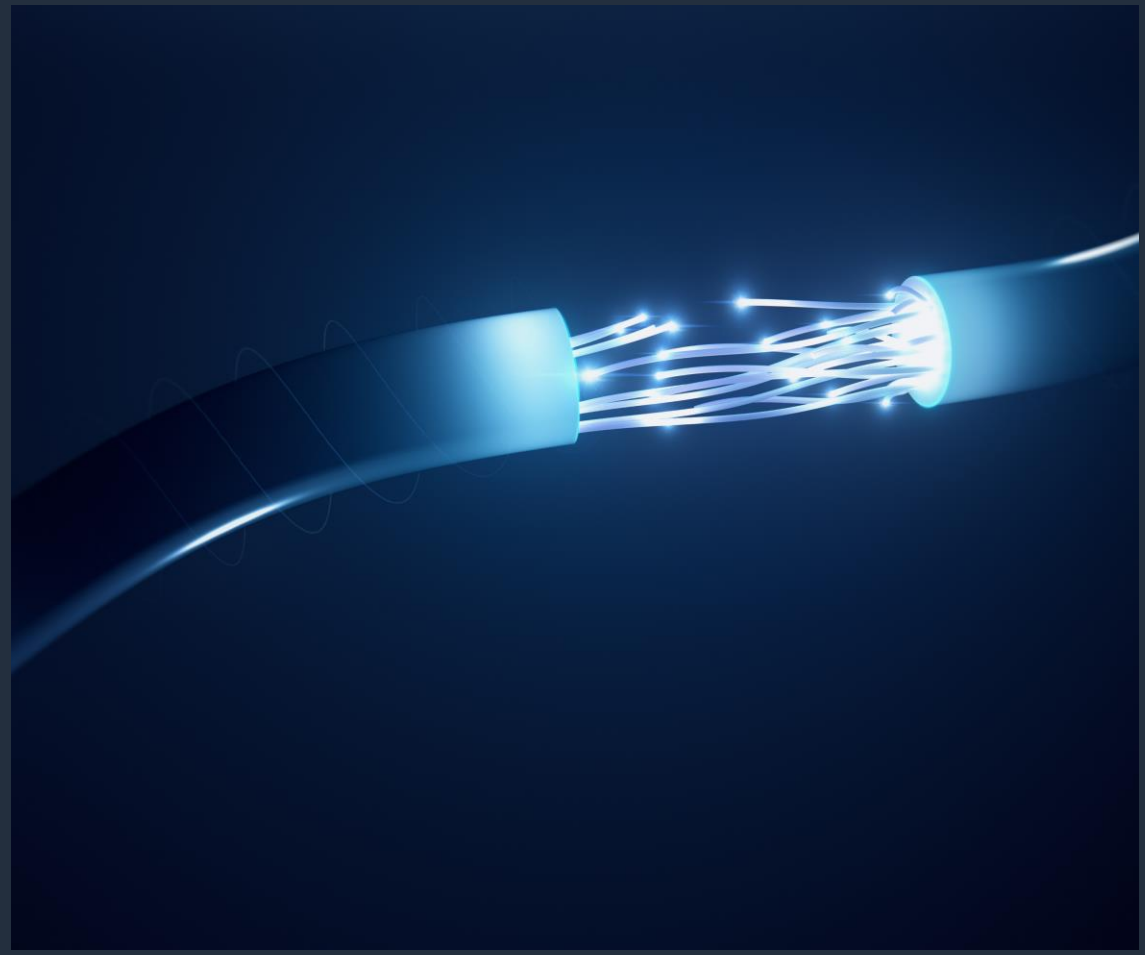


Benefits of Data Spaces



The connector component

- Share data
- Consume data offers
- Transfer data
- Maintain control over usage
- E.g.: Eclipse Dataspace Component Connector (EDC Connector)



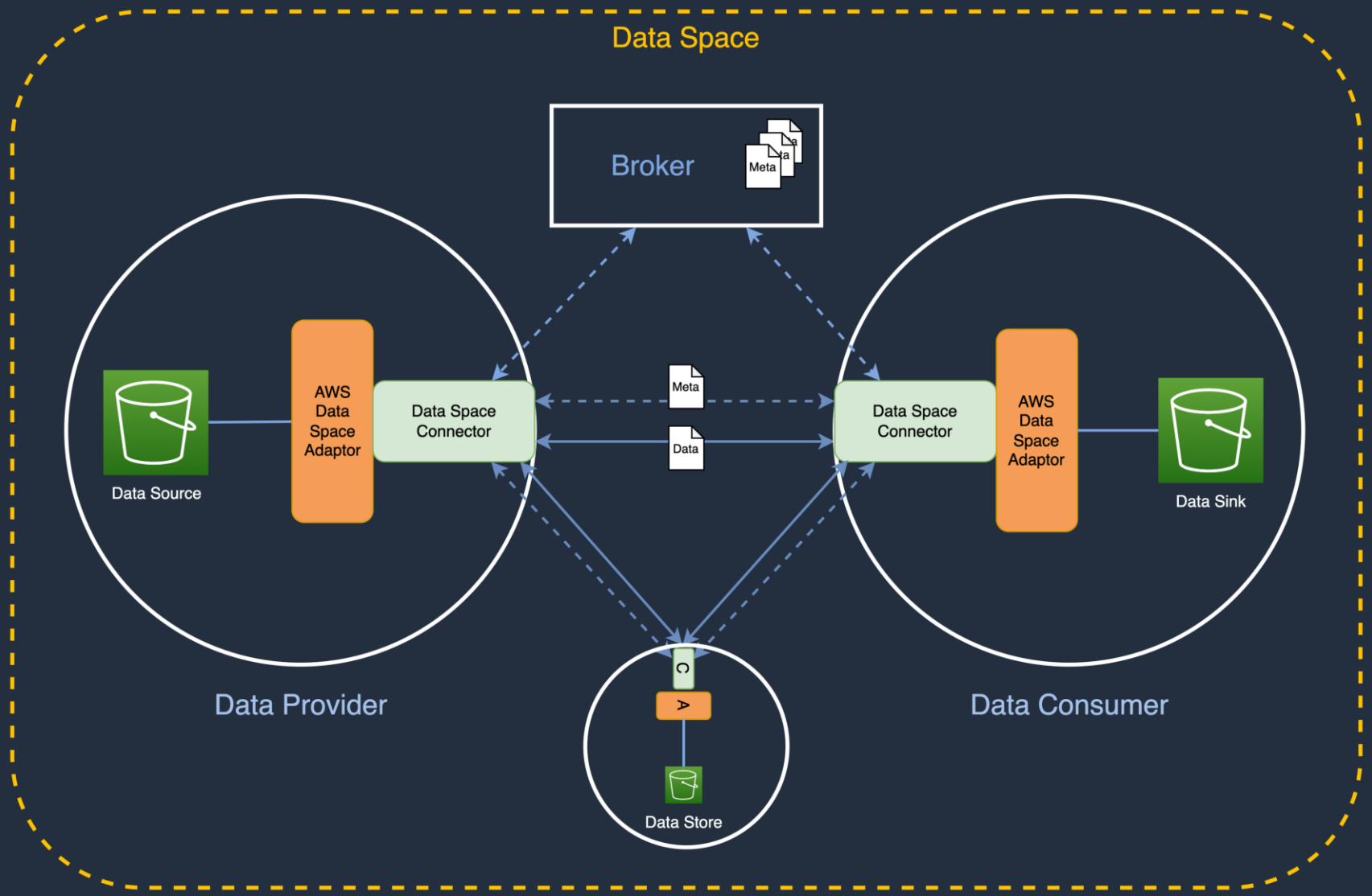
The AWS role in Data Spaces

1. Trusted infrastructure provider

2. Advisor with prescriptive guidances

3. Supplier of data services

Data spaces on AWS – high level architecture



Use cases across industries

Mobility

Preventing car accidents through local hazard tracking via car sensors and data-sharing

Energy

Creating better models to understand renewable distribution systems

Industrial & Manufacturing

Achieving supply chain transparency through an automatized exchange of data between all involved supply chain parties

Public administration

Increasing transparency and integrity through linking industry data and public procurement data across Europe

Green Deal

Creating open digital solutions for identification, tracking, mapping, integrity-check and sharing of product information.

Health

Identifying patients of high risk of disease by integrating long-term follow-up data from hospitals and health insurers

…Financial
…Agriculture
…Open Science Cloud (Research)
…Media
…Cultural Heritage

3. EHDS: High-level Overview

“I want you to work on the creation of a European Health Data Space to promote health-data exchange and support research on new preventive strategies, as well as on treatments, medicines, medical devices and outcomes.

As part of this, you should ensure citizens have control over their own personal data.”.

Ursula von der Leyen,

President-elect of the European Commission in the Mission letter to Stella Kyriakides, Commissioner-designate for Health on 10 September 2019

What is the European Health Data Space (EHDS)?

The EHDS is a “health specific data environment which comprises rules, common standards and practices, infrastructures and a governance framework for the use and reuse of electronic health data.”

Empower individuals to control (and benefit from) shared access to their personal health data

Support the free movement of data and citizens within a Pan-EU interoperable environment for the provision of healthcare **(primary use)**

Foster Pan-EU sharing of health-related information, setting legal ground for third-parties access to health data e.g. for research, innovation, policymaking and the development of personalized medicine **(secondary use)**

EHDS: Legal Framework

- Proposal for a Regulation on the EHDS
- Objective: Regulate and enable primary & secondary use of health data
- Timeline: coming into force in 2024-2025

EHDS Scale

- Funding:
  - €12 billion from Member states under Reselience & Recovery Facility
  - €810 million from European Commission
- Geography
  - Direct impact: EU
  - Indirect impact: globally

Secondary Use of Health Data: Mechanism

- Creation of HealthData@EU, a decentralized EU-infrastructure for (cross-border) use of data;
- Establishing of the national health data access bodies for monitoring rules associated to the secondary use of data.
- Definition of a set of data types that can be used for defined purposes, as well as prohibited purposes, and anonymisation requirements.
- Provisions on data altruism (providing consent to make available data – voluntarily and without reward)

Secondary Use of Health Data: Technology

- Data storage
- Dataspace connector
- Data analytics, AI & ML tools
- Tools for encryption, anonymization, pseudonymization
- Key aspect: configuration in alignment with EHDS

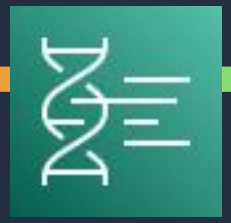
Consider: Joining a Data Space...

Does **not** replace, but only **extends** your data platform to the outside

Makes a comprehensive **data strategy** a necessity

Allows you to **derive insights and accelerate outcomes** with complementary tools

AWS Services for Healthcare and Life Sciences



Amazon Omics

Transform genomic, transcriptomic, and other omics data into insights

Amazon HealthLake

Imaging and Analytics

Provide a complete view of individual or patient population health data

Amazon Comprehend Medical

Understand medical context using natural language processing

Amazon Transcribe Medical

Automatically convert medical speech to text

Lifebit Powers Collaborative Research Environment for Genomics England on AWS

Challenge

At the onset of the COVID-19 pandemic, Genomics England turned to Lifebit to accelerate research. However, researchers relied on legacy technologies to manage genomic data, making collaboration difficult.

Solution

Lifebit used AWS to launch the first federated research environment that enables collaborative research on disparate genomic datasets worldwide, delivering virtually unlimited scalability and storage.

Benefits

- Launched a federated data analytics system in under 3 months
- Processes more than 100 PB of project data
- Maintains compliance with data privacy regulations

> "We use the whole roster of AWS computations—from general-purpose computation to graphically accelerated units—to run large production pipelines faster and more efficiently."

—Thorben Seeger, vice president of commercial, Lifebit Biotech Ltd.

Company: Lifebit Biotech Ltd.

Industry: Life Sciences

Country: United Kingdom

Website: lifebit.ai

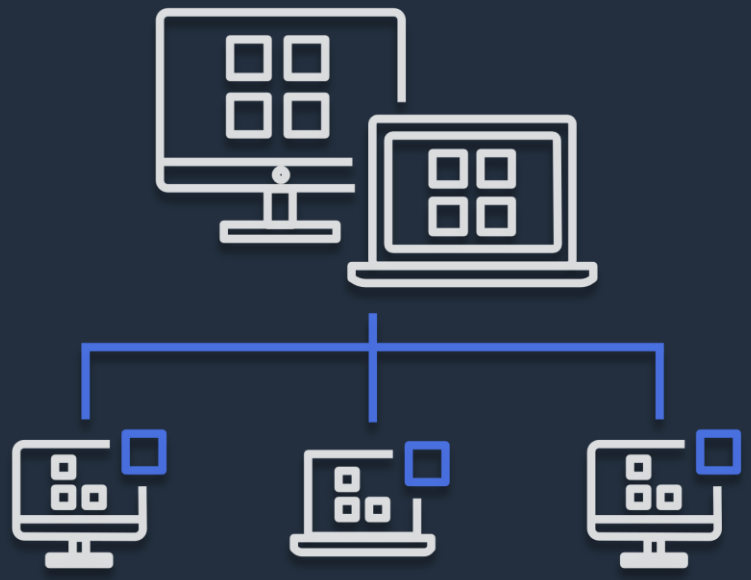
About Lifebit Biotech Ltd.

Lifebit Biotech Ltd. is a global leader in population genomics software and AI-powered drug discovery. Operating in North America, Europe, the Middle East, Africa, and the Asia-Pacific region, it powers population genomics initiatives, biobanks, research, and pharma companies.

4. Trusted Research Environment on AWS

What is Trusted Research Environment (TRE) on AWS?

A **self-service research solution** to secure and analyse **sensitive** data



Do you…?

WANT TO HOST SENSITIVE DATA?

HAVE CONCERNS WITH RESEARCH SECURITY & COMPLIANCE?

HAVE RESEARCHERS THAT USE SENSITIVE DATA?

WANT TO PROVIDE FLEXIBILITY & SCALE FOR RESEARCHERS?

Gap: Secure data analysis environments

The Five Safes Framework	
1. Safe People	Trained and accredited researchers trusted to use data appropriately
2. Safe Projects	Only used for valuable, ethical research that delivers clear public benefits
3. Safe Data	Researchers can only use data that have been de-identified
4. Safe Setting	Access to data is only possible using our secure technology systems
5. Safe Outputs	Outputs are checked to ensure they cannot identify data subjects

Office for National Statistics

UK Health Data Research Alliance

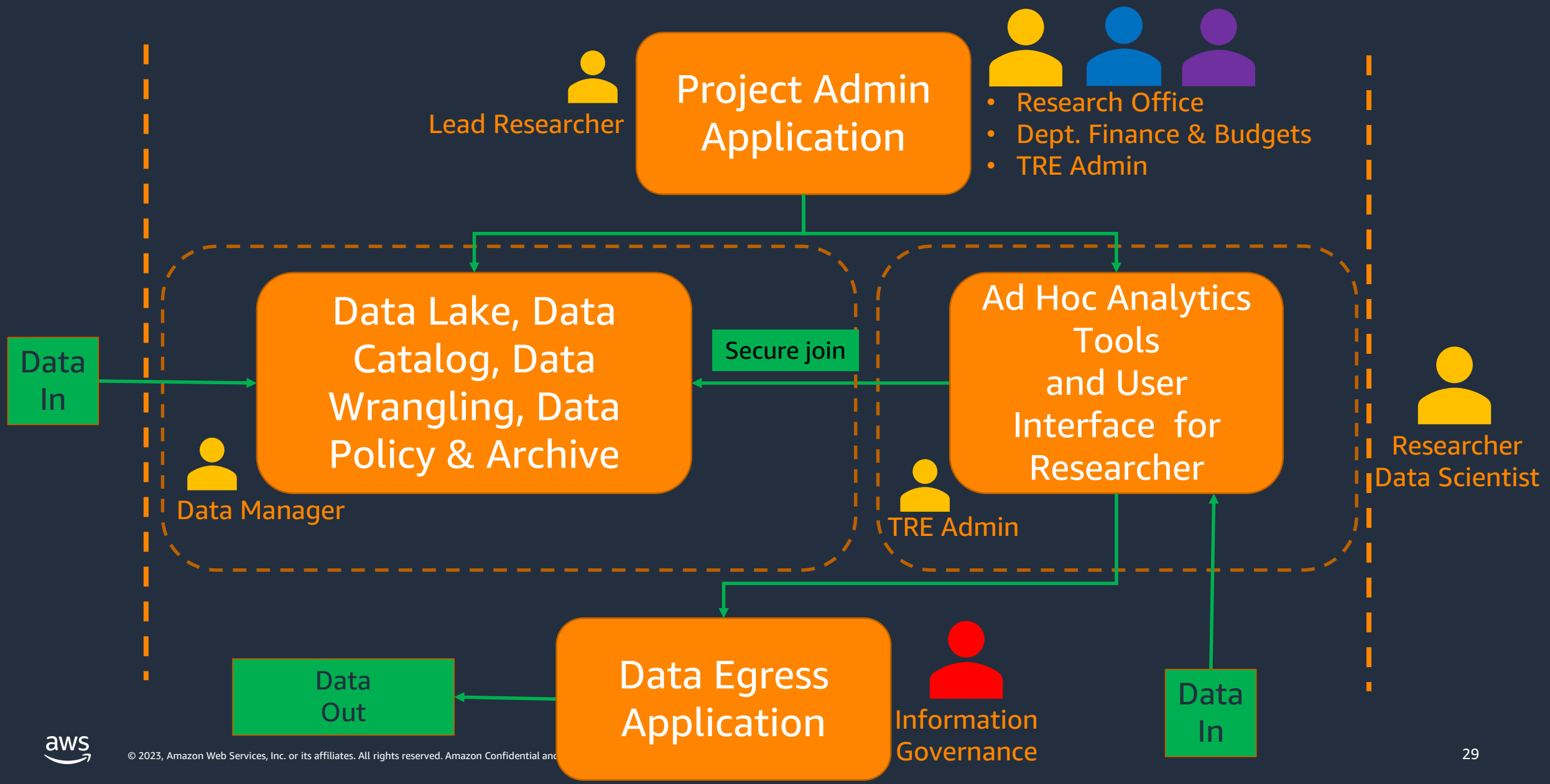
Turing Paper: Design choices for productive, secure, data-intensive research at scale in the cloud

Core Benefits of TRE on AWS

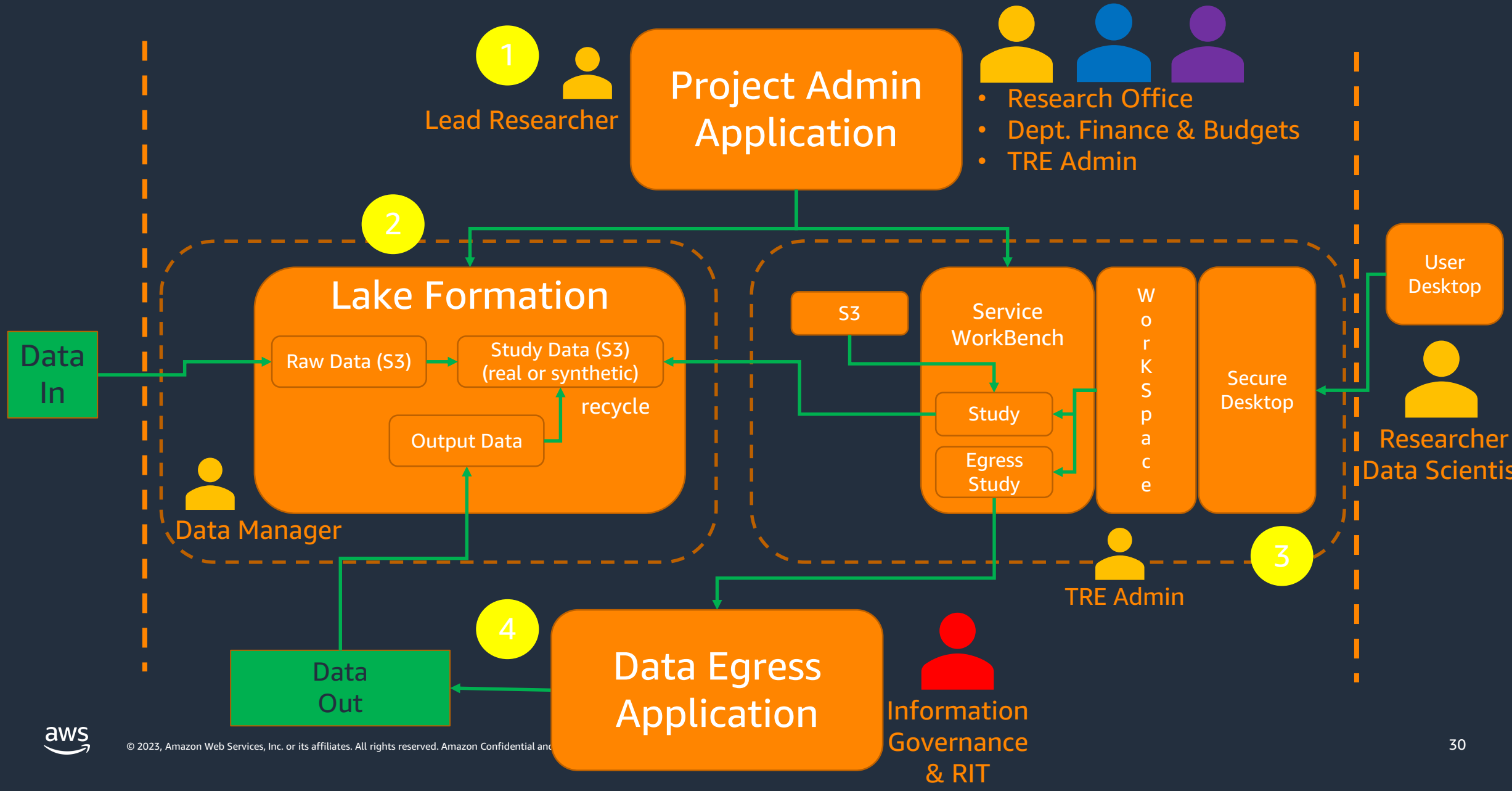
SECURITY	**Standardise researcher workspaces and manage access to AWS services with built-in security compliance, auditability, cost controls and regulatory safeguards**
EASE OF ACCESS	24x7 on demand access to the self-service portal allowing customers to spin up AWS research environments that are pre-approved by organisational IT
COLLABORATION	The ability to “bring your own account” to any project enables your team to securely collaborate with other organisations while retaining control
SCALE	Create and launch complex yet reproducible workflows leveraging the scale of the Cloud
TRANSPARENCY	Transparent view of total cost across projects, including cost centres and accounts for budget and chargeback management.
ACCELERATE RESULTS	Reduced time to results by establishing environments and access to data in less time than traditional methods allow.

TRE Architecture

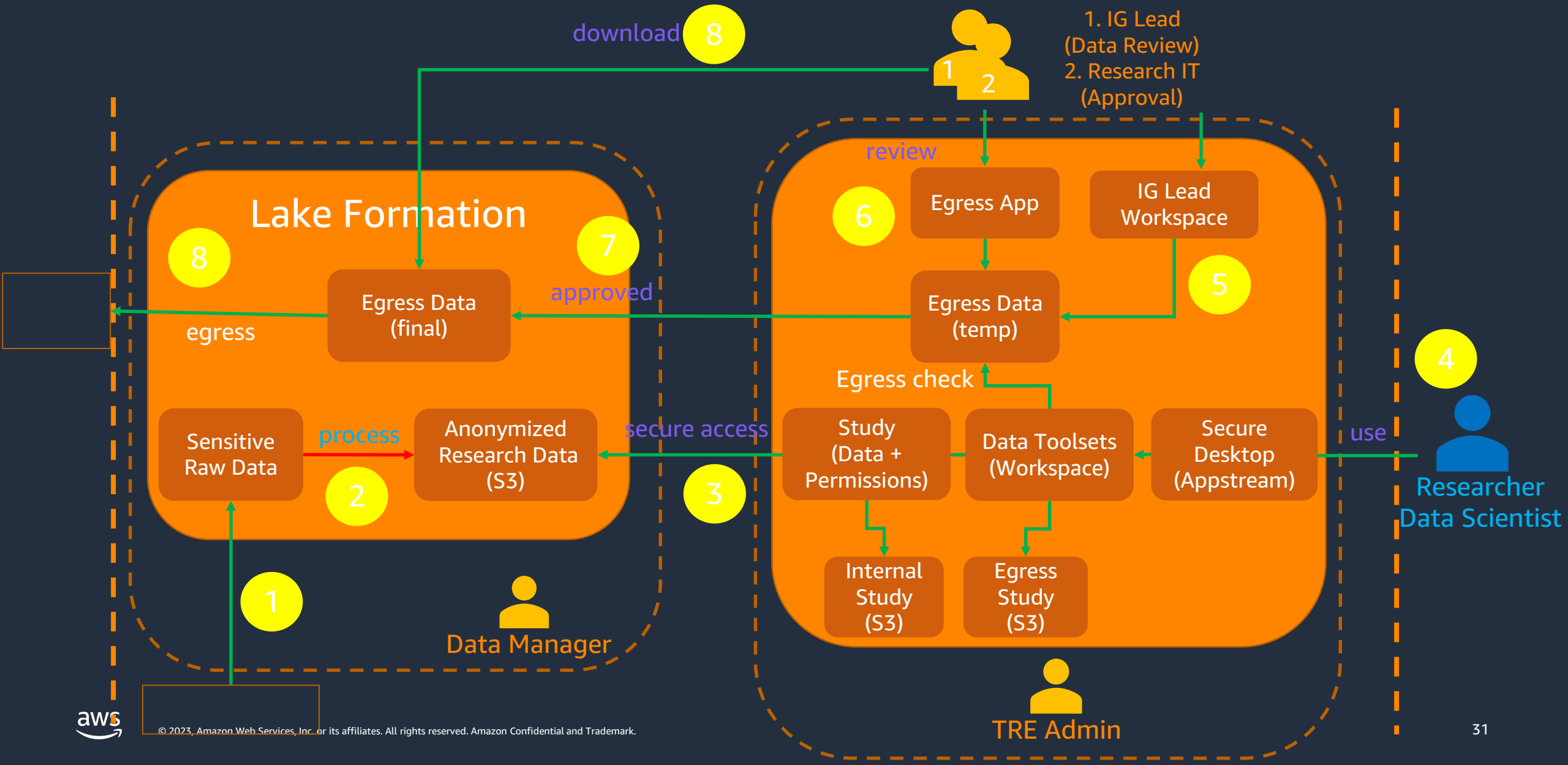
Trusted Research Environments



Trusted Research Environments (TRE)



Data Lifecycle



Next steps

BUILD WITH US

1. AWS Workshops:
   - Working Backwards
   - Data Spaces Innovation

2. Ideation and Proof of Concept

BUILD WITH PARTNERS

1. AWS Partner Network—100,000+ partners

2. AWS Marketplace and Solutions Library (ISVs)

UPSKILL YOUR TEAMS

AWS Training and Certification

5. Guest Speaker:

Professor Edward Curry,

University of Galway |
Director, Insight SFI Research Centre for Data Analytics

6. Panel Discussion

Questions to discuss

1. What is your positioning on data spaces in general and on the European Health Data Space for instance?
2. Are you interested in particular data spaces projects?
3. What opportunities/challenges or even blockers do you see?
4. What do you expect from providers like AWS?
5. Are there any other data space initiatives that you find particularly interesting?

Thank you!

Agata Jablonka,
jabagata@amazon.de

Syrine Souissi,
souiss@amazon.fr

Khrystyna Shlyakhtovska,
khrystyx@amazon.com

Tim Cutts,
tjrc@amazon.co.uk