

Wednesday, August 17, 2022

10:30 AM – 10:50 AM

Global Data Management through a Data Mesh

Anthony Zech

Data and AI Community of Excellence Director
ECS

Abstract:

A unified network needs to support multiple contexts for data and network connectivity to support customers from garrison to disadvantaged tactical users. This requires an approach that is decentralized and flexible while using standard toolsets that enable seamless operations from tactical edge to cloud. These requirements mean that traditional methods of establishing data lakes or warehouses will not work due to the requirements for responsive data transport and differences in information requirements between different echelons and organizations. Legacy approaches also experience further issues when teams need to evolve data to meet operational requirements. Traditional approaches generate rigid data dependencies controlled by a central team that result in accumulated technical debt when organizations attempt to modernize infrastructure elements. This rigidity is a result of the lack of organizational alignment and flexibility centralized teams have. Distributed teams encourage mission type command approaches and closely align data generation and sharing. Centralized, monolithic approaches create static organizational structures that evolve slowly due to institutional inertia and coordination requirements.

Data meshes and distributed but queryable data stores address the issues of flexibility and responsiveness while preserving capability in communication disadvantaged environments. A data mesh is a distributed approach to data storage and access that shifts data storage and access responsibilities to data producers while preserving centralized metadata repositories on data structure and identity and access management that enable organization-wide data access. This conceptualizes data as a product of every producing organization, empowering them to evolve data production while maintaining backwards compatibility. Further, it decouples data storage and production from potentially unreachable central repositories.

This approach can still leverage the advantages of cloud- in fact data meshes can more flexibly use cloud resources than legacy approaches. Data meshes can include legacy applications, but are typically built on microservices leveraging Kubernetes and containers. Using this approach for applications allows organizations to decouple applications from infrastructure. Kubernetes and container infrastructure allow applications to run locally- but also in the cloud when available.

Automatic service discovery and services built on APIs standardized via a data mesh metadata repository enable capability to gracefully and automatically expand and contract as resources become available. A data mesh paired with the decentralized data query capabilities of Elasticsearch allows access to data across the world when available. It also will gracefully degrade when communication pathways become unavailable due to operational or adversarial reasons. In this model, the data mesh implementation provides documented and discoverable ways to access data securely while Elasticsearch's distributed search capabilities provide the capability to get access to data from across the world that is transparent to users. A data mesh model with distributed search moves responsibility for providing data to the

producers of the data itself. This is a model that is distributed and is more flexible for users while also reflecting organizational structures. Crucially for the military, it is a more survivable model and one that will be far more effective in a contested information space while still preserving global visibility.