

# Edge Data Discovery for COIN

draft-mcbride-edge-data-discovery-overview-02

Mike McBride (Futurewei)  
Dirk Kutscher (Emden University)  
**Eve M. Schooler (Intel)**  
Carlos J Bernardos (UC3M)

*April 2020*

# What's the Problem?

- Increasing #s of devices and sensors generate a torrent of data
  - at the network Edge that flows upstream
- Sometimes that data must be processed/transformed → new data!
  - e.g., transcoded, subsampled, compressed, analyzed, annotated, combined, aggregated, et cetera
- In addition, (transformed) data may be cached/stored at multiple locations in the network on route to its final destination
- **As more distributed data is created, processed and stored, it becomes increasingly dispersed**
  - Throughout the network
- There needs to be a standard way – or ways - to find it!
  - New and existing protocols may need to be identified/developed/enhanced for distributed data discovery at the network edge ...and beyond

# How does this relate to COIN?

COIN requires data input and often results in data output:

- From where does COIN expect the data to come? To where does it expect it to be cached or to flow afterwards?
- How should the availability of data be exposed, where appropriate, while at the same time its privacy preserved?
- How to ensure COIN protocols comprehend the Edge context where data may not be movable (because of its abundance)?

# Table of Contents

1. Introduction .....	2
1.1. Edge Data .....	3
1.2. Background .....	3
1.3. Requirements Language .....	
1.4. Terminology .....	
2. The Edge Data Discovery Problem Scope .....	
2.1. A Cloud-Edge Continuum .....	
2.2. Types of Discovery Edge Data .....	6
3. <del>Protocols</del> Scenarios for Discovering Edge Data Resources. . .	8
4. Edge Data Discovery .....	8
<del>4. Protocols for Discovering Functions</del>	
4.1. Types of Discovery .....	
4.2. Naming the Data Discovery .....	
5. Use Cases of edge data discovery .....	10
6. IANA Considerations .....	10
7. Security Considerations .....	
8. Acknowledgement .....	10
9. Normative References .....	11

Clarified definitions:

- Edge computing
- Information Centric Networking
- Edge data locations

*e.g., streaming sensor or measurement data, streaming media, meta-data, control data, functions/services, bags of bits...*

Focused on Data Discovery  
(vs general discovery problem)

Service Function Chaining

# Main Feedback

- ✓ Need a crisper definition of Edge data discovery problem
- ✓ Convert NDN discussion into an ICN discussion
- ✓ Better integrate the section on service function chaining (meta) data
  
- Examine the broader Edge data management problem
  - Data life-cycle: Discover, Search, Access, **Compute/Transform**, Pin/Place, Migrate, Expire, Secure, Preserve-privacy, Support mobility,...
  - Tease apart: caching, storage, replication
- Need a more thoughtful Security section
  - Treat devices as part of the Cloud-to-Edge continuum? Or separately because of security, privacy, mobility?
    - Devices are authoritative re the data they have
  - What about discovery in a P2P manner?

## Additional feedback

- ✓ Articulate IoT Data vs any Data vs Endpoint Data
- ✓ Clarify if Video analytics is a vertical segment? Used Video surveillance instead.
- ✓ Remove blurb on business case for data economy
- ✓ Need COIN for data noise elimination & obfuscation, too
- ✓ Ack: Not all Clouds silo data. Not all Edges expose data.
- ✓ Add that discovery is needed for Edge data placement, too
- ✓ Clarify why Data is needed by analytics for “local and/or low-latency decisions” - for predictive maintenance? emergency services?

# Next Steps – adopt as Research Group I-D?

- Address unanswered feedback in -04
  - Recent work on automation of data discovery to be included in -04
- **Decide scope: Edge data discovery vs data life-cycle management**
  - From COIN perspective
- Scrutinize existing discovery protocols & approaches for suitability
  - CoAP RD, DNS SD, W3C Thing Directory ...for named data?
  - Containers, IPFS, DHTs, MobilityFirst... for data management?
- Identify the gaps
  - Build out Use cases and requirements
- Recruit someone to help with Security considerations
  - Welcome others interested in the general problem space
- Near- vs Medium- vs Long-term solutions?

BACKUP

# Scenarios Requiring Discovery of Edge Data

1. A set of data resources appears (e.g., a mobile node joins the net)
  - want to be discoverable by an existing but possibly virtualized and/or ephemeral data directory infrastructure.
2. A device wants to discover data resources near its current location
  - because some resources may be mobile, asleep, or only intermittently connected, the available set of edge data may vary over time.
3. A device wants to discover where best in the edge infrastructure to opportunistically upload/migrate its data
  - if a mobile device wants to offload its data to the infrastructure (for greater data availability, battery savings, safe keeping, etc.).
  - a network element is running out of space