SummerSoC
Service Oriented Computing

From Data Mesh to Intermesh: A Platform-Driven Approach to Govern Inter-Organizational Data Sharing

19th Int'l Symposium on Service-Oriented Computing

June 17th, Hersonissos, Greece

Arif Wider & Sebastian Werner

/thoughtworks

Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences

Agenda

- What makes Data Mesh successful within organizations?
- Spotlight: Federated Computational Data Governance
- What is different with inter-organizational data sharing?
- How can we apply the principles of Data Mesh to inter-organizational data sharing?

Arif Wider

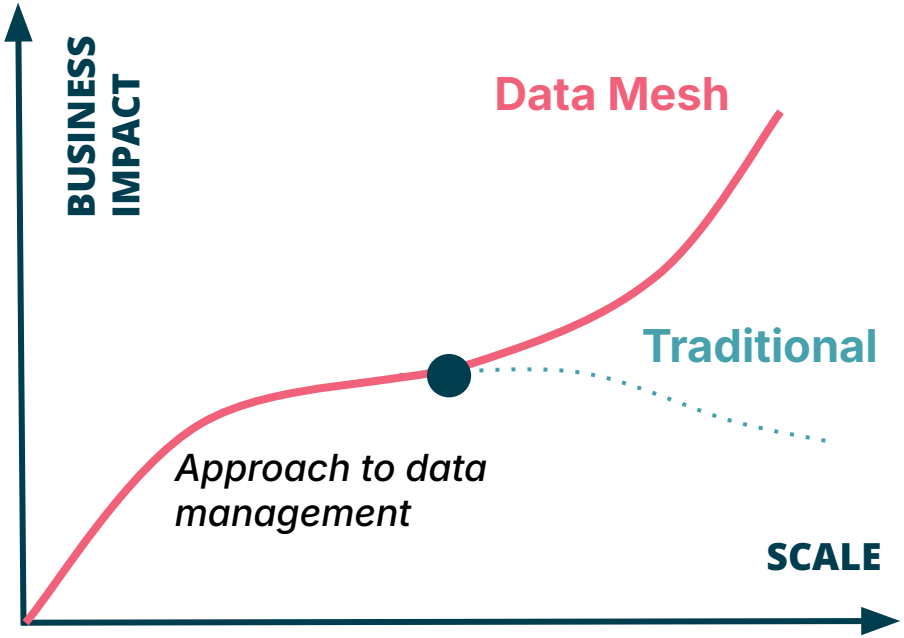
- Former Head of Data & AI at Thoughtworks Germany
- Many years hands-on consulting, often as a lead engineer in Thoughtworks' client's data teams
- Now tenured professor of software engineering at HTW Berlin, Thoughtworks fellow and independent consultant on the side
- Data Mesh expert, both in terms of research and strategic consulting

/thoughtworks

Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences

What problem is Data Mesh solving?

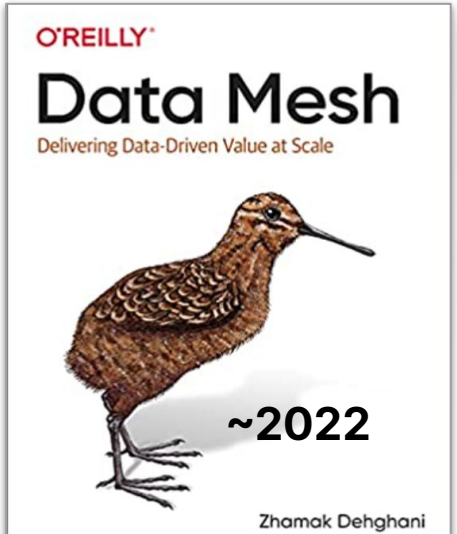
Value from data Agility, Time to market Resilience to change



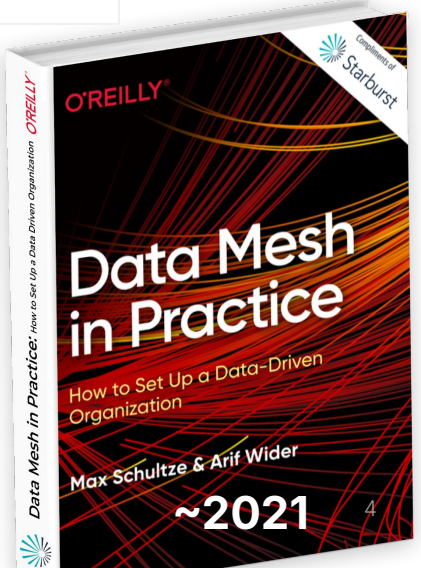
Business complexity
Data source expansion
Data use case proliferation

~2018

~2019



~2022



~2021

What makes Data Mesh successful in organizations?

- Bottlenecks and inefficiencies are reduced by moving ownership to where the expertise is.
  - Domain experts who understand the data take care of shape and quality of the data.
  - Data engineers focus on building domain-agnostic self-service tools that enable the domain experts.

➔ Responsible decentralization by a “Shift Left” of responsibility aided by platform-provided automation

Shift Left on Data Governance

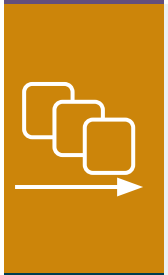
Move Responsibilities Upstream

A large part of the bottleneck in centralized paradigms is that responsibility is removed from source teams. In the data mesh concept, quality, catalog integration, access control, testing, etc. are shifted left.

We achieve the leftward shift through automation and federation, hence **Federated Computational Governance.**

Platform fitness functions for global functions are used to create an Evolutionary Architecture to ensure that data teams don't create a "Data Mess"

Implement, monitor, and enforce SLOs for data products with sensible and common definitions, e.g for availability, accuracy, latency, etc.



Data and metadata modeling follows standards defined by the governance team, implemented by the producing teams using platform tools

Domain Data Product Owners assume responsibility for the data

Federated computational governance

Pre data mesh		Data mesh
Centralized team		Federated team
Responsible for data quality		Responsible for defining how to model what constitutes quality
Responsible for data security		Responsible for defining aspects of data security *i.e. data sensitivity levels for the platform to build in and monitor automatically* …

Federated computational governance

Decentralized Data Governance as Part of a Data Mesh Platform: Concepts and Approaches

Arif Wider[1,2], Sumedha Verma[1], and Atif Akhtar[1]
[1]*Thoughtworks Germany*
Caffamacherreihe 7, 20355 Hamburg, Germany

***Abstract*—Data**
decentralized analy
decentralization effi
provided by a self-s
aspect of this pla
governance. Because
lack of coherence in
the industry, and alr
facilitates governanc
key data mesh conc
drive governance
presented are drawn
a fully-functional d
reference on how to a

***Keywords*—data**
infrastructure, data e

Data mesh is
management that
global IT consultar
2019 [1]. It has si
book on the topic
specifically addressi
that try to create bus

Towards Automating Federated Data Governance

Arif Wider[1,2], Katharine Jarmul[1,3], and Atif Akhtar[1]
[1]*Thoughtworks Germany*
Caffamacherreihe 7, 20355 Hamburg, Germany
{awider,kjarmul,syedatif}@thoughtworks.com
[2]*Hochschule für Technik und Wirtschaft Berlin*
Treskowallee 8, 10318 Berlin, Germany
wider@htw-berlin.de
[3]*Probably Private,* Berlin, Germany
katharine@probablyprivate.com

***Abstract*—Recently several approaches to federated data governance and decentralized data operations have been proposed. Data mesh, one of the most prominent of these approaches, specifically targets how to automate and scale governance at large organizations. Governance tasks, such as ensuring data protection compliance, are notoriously challenging to automate. This paper proposes new ways to verify data declassification via contract-based methods. The design benefits from compositional characteristics of data mesh and builds upon existing contract-testing solutions from the microservices domain.**

***Keywords*—federated governance, data protection, data mesh, data governance, decentralized data architecture, contract testing**

I. Introduction

One of the key challenges observed today when dealing with large analytical data systems is applying data protection

cumbersome, sluggish and prone to error. To o
challenges several decentralized data arch
federated approaches to data governance
emerged. *Data mesh* [2] is one such approach
decentralized data ownership and federated da
with advanced automation. This automation is
self-serve data infrastructure platform, specifica
meet the scalability demands of large busines
complex data landscapes [3]. Since data mesh is
approach, many solutions to governance challe
to be developed [4].

In this paper we focus on a subset of the
namely the automation of data protection po
challenge, we believe that data mesh'
characteristics lend itself to a compositional ap
governance automation. Our hypothesis is tha

AI-Assisted Data Governance with Data Mesh Manager

Arif Wider
Hochschule für Technik und Wirtschaft
Berlin, Germany
wider@htw-berlin.de

Simon Harrer
innoQ Deutschland GmbH
Monheim am Rhein, Germany
simon.harrer@innoq.com

Linus W. Dietz
King's College London
London, United Kingdom
linus.dietz@kcl.ac.uk

***ct*—The exchange of data across systems, both within**
een organizations, is governed by company policies and
ection regulations. As policies and data flows evolve over
uring continuous compliance of data exchange remains a
challenge. In federated data architectures, the validation

mesh is the *data product*. A data product is not merely a dataset managed with a product-focused mindset; it is also a data service designed for composability. Data products interact through output ports, which are clearly defined interfaces that specify the schema of the shared data. Any terms associated

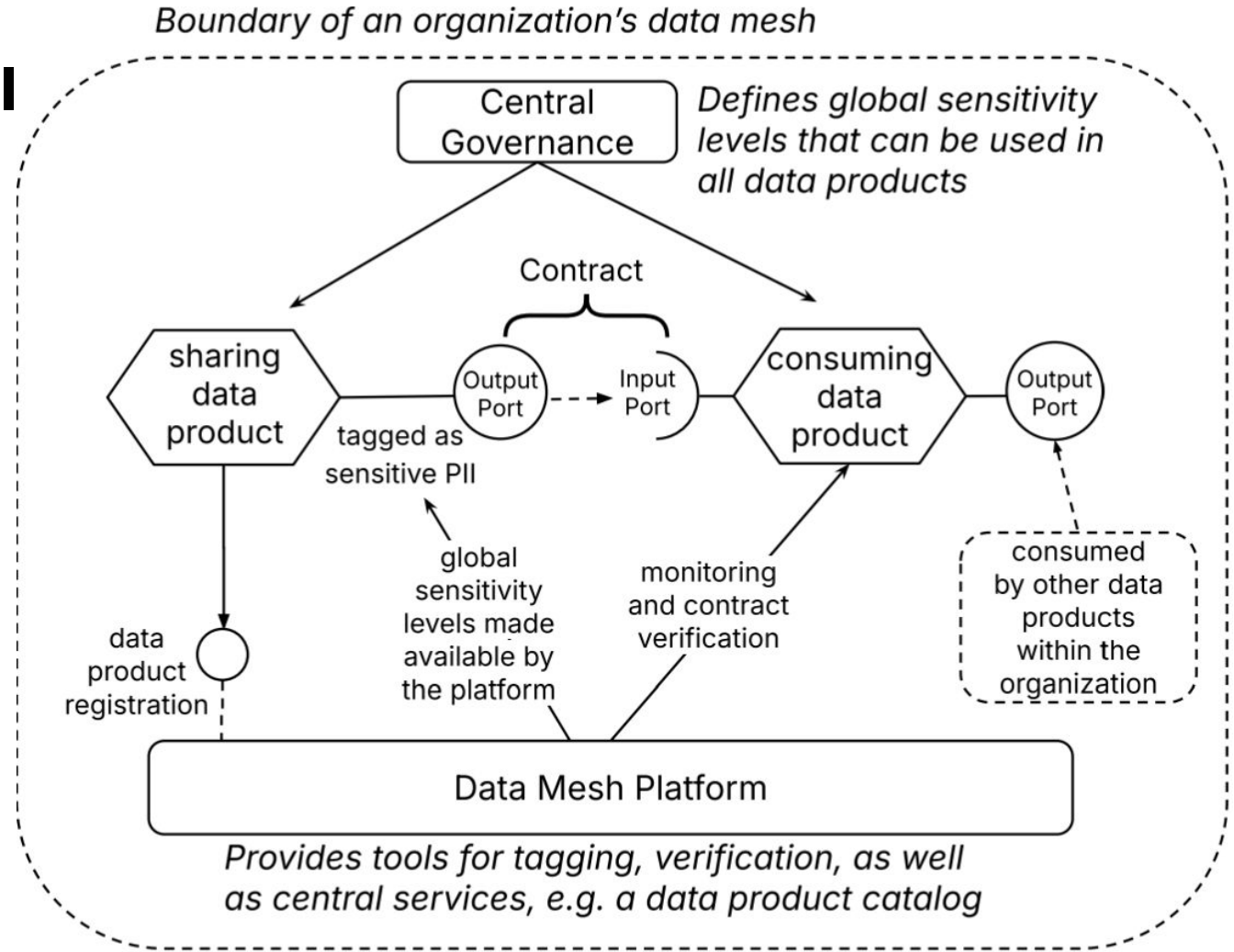
From Data Mesh to Intermesh: A Platform-Driven Approach to Govern Inter-Organizational Data Sharing

Arif Wider[1,2][0009-0006-4327-7626] and Sebastian Werner[2][0000-0003-2789-1926]

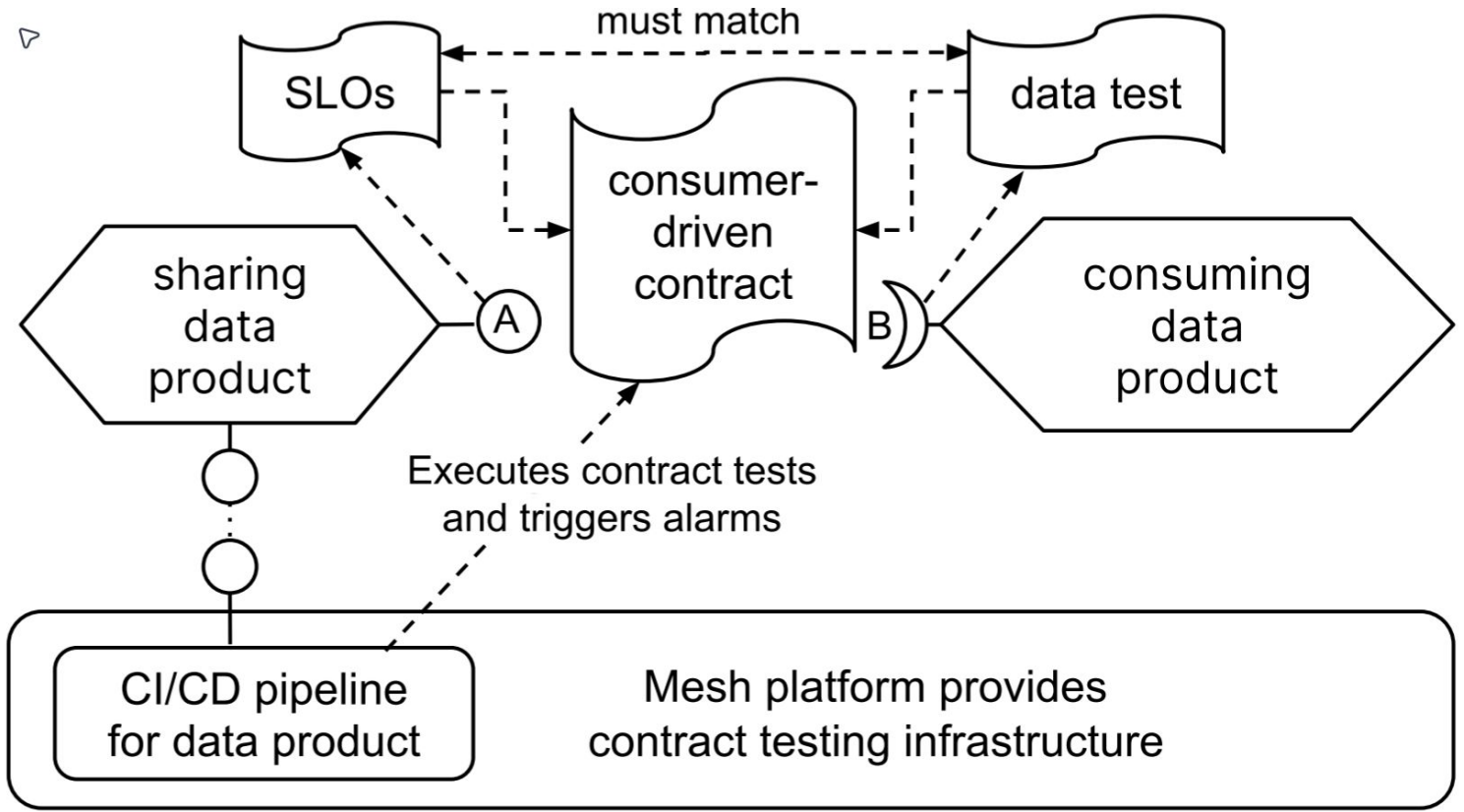
[1] Hochschule für Technik und Wirtschaft Berlin, Treskowallee 8, 10318 Berlin, Germany
wider@htw-berlin.de
[2] Thoughtworks Germany, Caffamacherreihe 7, 20355 Hamburg, Germany
{awider,swerner}@thoughtworks.com

Abstract. Analytical data is becoming an increasingly important asset for all kinds of organizations, which comes with a growing demand for efficient solutions to also share such data between organizations. Initiatives such as GAIA-X address this by promoting infrastructure standardization and have

Computational Governance Example



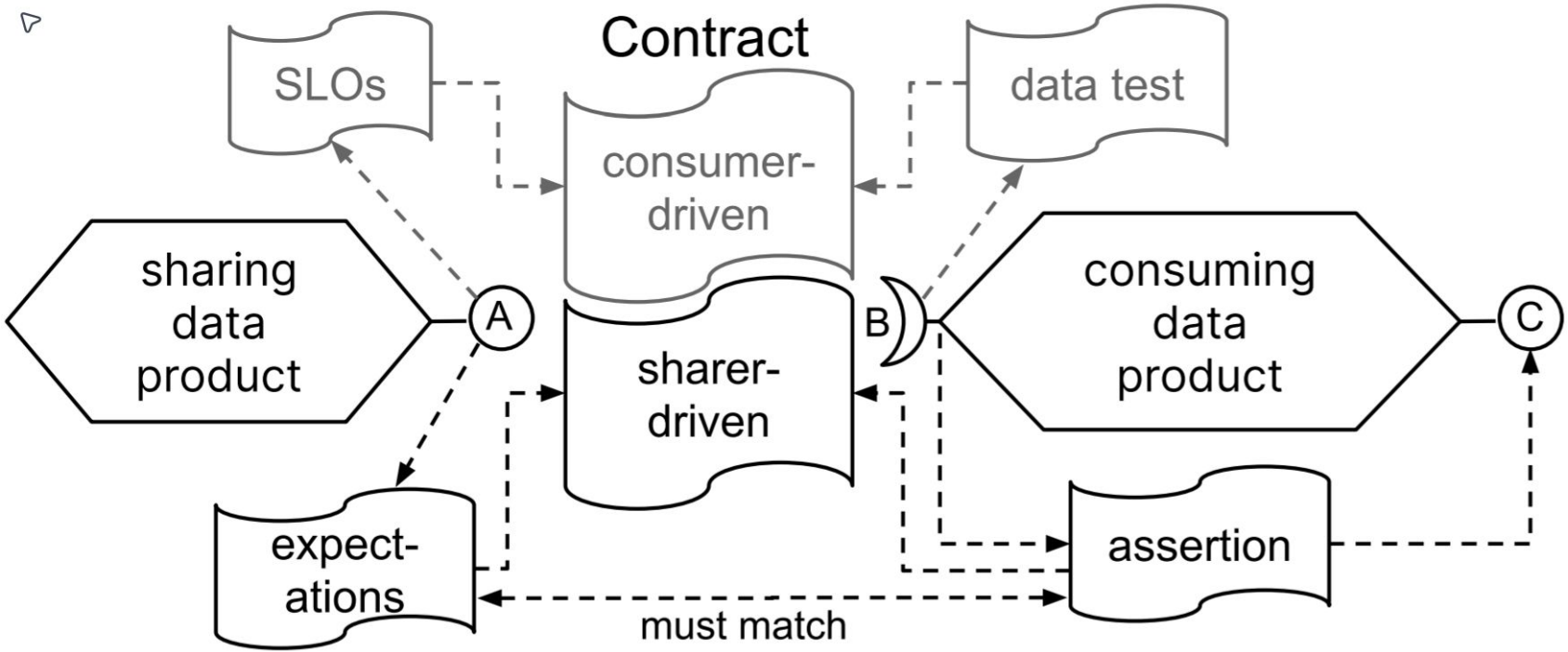
Automating Data Contract Testing



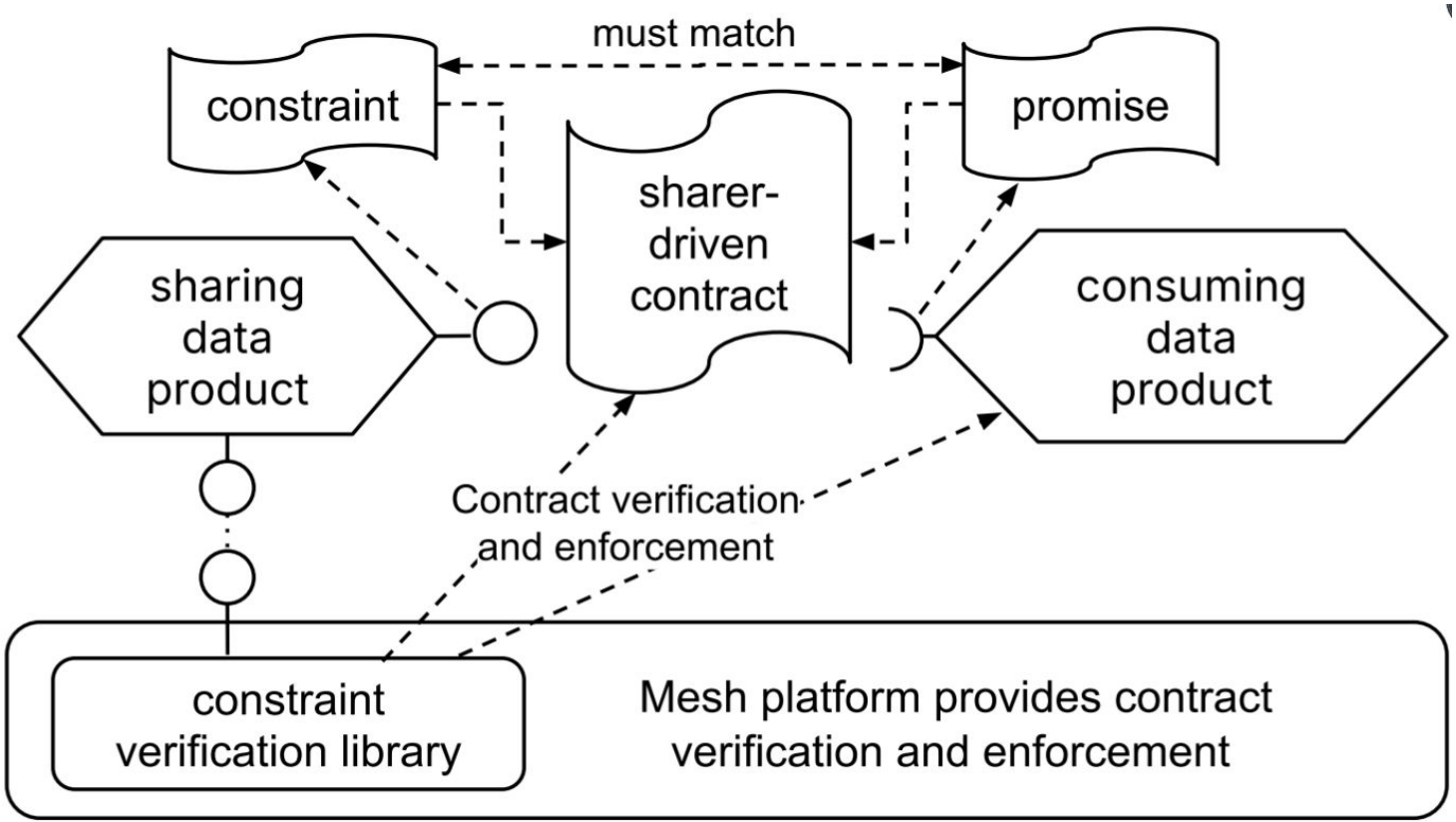
What is different with inter-org data sharing?

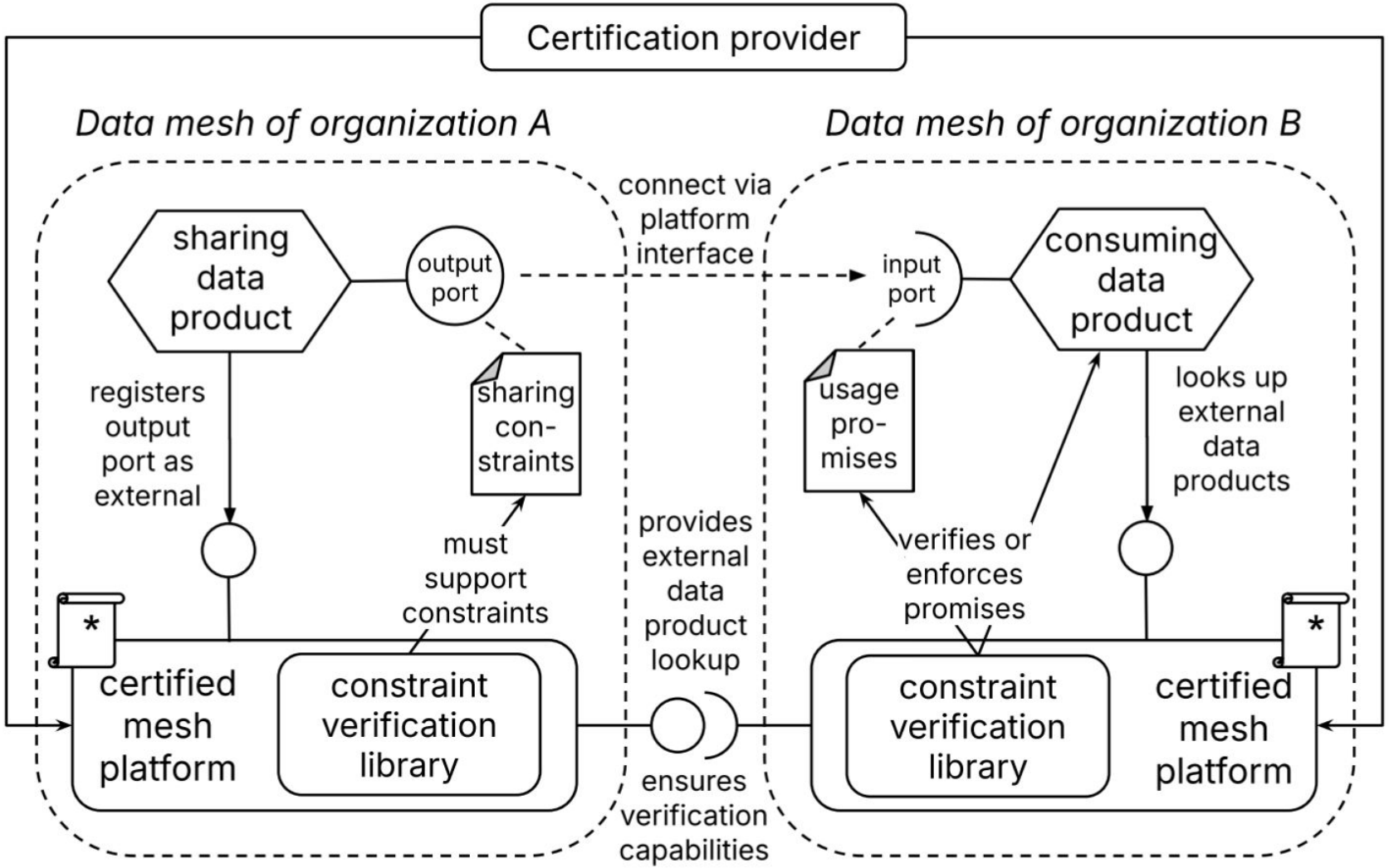
- Different data product implementations
  - External data products must be shared in a standardized way
- No implicit trust
  - Consumer-driven contracts are not enough
- No shared data infrastructure platform
  - How to check guarantees across platforms?

From consumer- to sharer-driven expectations



From consumer- to sharer-driven expectations




Certification provider
Data mesh of organization A
Data mesh of organization B
sharing data product
output port
connect via platform interface
input port
consuming data product
registers output port as external
sharing con-straints
usage pro-mises
looks up external data products
must support constraints
provides external data product lookup
verifies or enforces promises
*
*
certified mesh platform
constraint verification library
constraint verification library
certified mesh platform
ensures verification capabilities

Summary

- Within organizations, Data Mesh improves efficiency through decentralization and automation
- When applying Data Mesh’s approach to inter-organizational data sharing, the lack of trust and shared infrastructure calls for cross-organizational standardization.
- However, instead of focusing on data product standardization, we propose to focus on standardizing data mesh platform capabilities

SummerSoC
Service Oriented Computing

Thank you!

wider@htw-berlin.de

awider@thoughtworks.com

mail@arifwider.com

/thoughtworks

htw. Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences