



## DIPLOMATERVEZÉSI FELADAT

**Klenk Botond (FTNYN1)**

Mérnök informatikus hallgató részére

### Adatterek minőségbiztosításának támogatása

Az adatmegosztást és -feldolgozást, intelligens algoritmusok ill. MI szolgáltatások felhasználását és moduláris rendszerbe illesztését az elmúlt időszakban az EU által is támogatott adatterek technológiai és rendszerszervezési megközelítésében alkalmazzák számtalan területen, gyártási adatok megosztásától kezdve energetikai alkalmazásokon át személyre szabott képzési ajánlatok elkészítéséig. Ezekben a megközelítésekben közös, hogy az adattér, mint infrastruktúra célja alapvetően az adatot megosztó és feldolgozó felek közti kapcsolat felépítése, ugyanakkor a konkrét adattranszfer a felek közt közvetlenül történik.

Ha az adatmegosztás és az arra épülő alkalmazások üzleti célt szolgálnak, kiemelten fontos az adatmegosztás sikerének biztosítása, az átadott adat és az arra adott válasz minőségének kiértékelése. Ezekre a kérdésekre ugyanakkor a legtöbb elérhető megoldás (pl. az Eclipse Data Connector vagy a különböző referencia architektúra modellek-IDSA, DSSC, Gaia-X, stb.) nem adnak választ, ill. nem kezelik a rendszer kiemelt feladataként.

A hallgató feladatának a következőkre kell kiterjednie:

- Vizsgálja meg és röviden értékelje ki, milyen kérdések merülnek fel egy adat átadásakor annak minőségével kapcsolatban, és hogyan oldható meg ezek kiértékelése adatterek (data spaces) tervezésénél és működtetésénél.
- Mutassa be, hogyan lehet egy ellenőrző komponenst az adatmegosztási folyamatba integrálni.
- Mutassa be, milyen szerepe lehet a blokklánc technológiának az adatmegosztás jellemző tulajdonságainak rögzítésében.
- Készítsen prototípus implementációt és mutassa be annak jellemzőit konkrét adatmegosztási példákon keresztül. Munkája során törekedjen megoldása általánosíthatóságára.

**Tanszéki konzulens:** Dr. Gönczy László, egyetemi docens

Budapest, 2024. március 17.

.....  
Dr. Dabóczi Tamás  
tanszékvezető  
egyetemi tanár



**Budapest University of Technology and Economics**  
Faculty of Electrical Engineering and Informatics  
Department of Measurement and Information Systems

# **Quality assurance support for dataspaces**

MASTER'S THESIS

*Author*  
Botond Klenk

*Advisor*  
dr. László Gönczy

May 11, 2024

# Contents

<b>Kivonat</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Problem Definition and Motivation . . . . .	1
1.2 Blockchain as a Solution . . . . .	1
1.3 Structure of the Document . . . . .	1
<b>2 Data Spaces</b>	<b>2</b>
2.1 Summary . . . . .	2
2.2 Initiatives . . . . .	2
2.2.1 DSSC . . . . .	2
2.2.2 Gaia-X . . . . .	2
2.2.3 IDSA . . . . .	2
2.2.4 Catena-X . . . . .	2
2.2.5 Prometheus-X . . . . .	2
<b>3 Quality Assurance</b>	<b>3</b>
3.1 Data Quality . . . . .	3
3.1.1 Volume . . . . .	3
3.1.2 Variety . . . . .	3
3.1.3 Veracity . . . . .	3
3.1.4 Value . . . . .	3
3.1.5 Velocity . . . . .	3
3.2 Data Veracity . . . . .	3
3.3 State of the Art . . . . .	3
<b>4 Architecture</b>	<b>4</b>
4.1 Demands . . . . .	4
4.1.1 Separated Component . . . . .	4
4.1.2 Not Slowing the Transfer . . . . .	4
4.1.3 Correlating Process Time . . . . .	4
4.1.4 Validate Check . . . . .	4

4.2	Questions . . . . .	4
4.2.1	How to Validate . . . . .	4
4.2.2	Who Can Access the Data . . . . .	4
4.2.3	How the Data Is Accessed . . . . .	4
4.3	Possible Solutions . . . . .	5
4.3.1	Data Space Component . . . . .	5
4.3.2	Delegates with Oraculum . . . . .	5
4.3.3	Own Checkers . . . . .	5
4.4	Summary . . . . .	5
<b>5</b>	<b>Implementation</b>	<b>6</b>
5.1	Technologies . . . . .	6
5.1.1	Express.js . . . . .	6
5.1.2	OpenAPI . . . . .	6
5.1.3	Hyperledger Fabric . . . . .	6
5.2	Implemented Solution . . . . .	6
	<b>Bibliography</b>	<b>9</b>
	<b>Acronyms</b>	<b>10</b>
	<b>Index</b>	<b>11</b>

# Kivonat

Napjaink egyre inkább adatközpontú világában, ahol az információ számtalan iparág és vállalkozás éltető eleme, az adatok értékét nem lehet eléggé hangsúlyozni. A személyes meglátásoktól a szervezeti döntéshozatalig az adatok minősége és megbízhatósága óriási jelentőséggel bír. Következésképpen a magánszemélyek és szervezetek világszerte jelentős erőforrásokat fedetnek be adataik védelmébe, gyakran magáninfrastruktúrákhoz folyamodva, hogy biztosítsák azok integritását és bizalmas jellegét.

Mivel azonban az adatok egyre szaporodnak és diverzifikálódnak, a magáninfrastruktúrák elszigetelt jellege kihívást jelent az adatok hatékony felhasználása és cseréje szempontjából. Felismerve egy egységesebb és összekapcsolt megközelítés lehetőségét, szükségessé válik egy olyan rendszer kifejlesztése, amely hatékonyan áthidalja az adatszolgáltatók és a fogyasztók közötti szakadékot. Ez a szükségszerűség hívja életre az adatterek koncepcióját - egy olyan keretrendszert, amelynek célja az adatkezelés és -csere forradalmasítása.

Az adatterek alapvetően paradigmaváltást jelentenek abban, ahogyan az adatokat érzékeljük és az adatokkal interakcióba lépünk. Ahelyett, hogy az adatokat elszigetelt tárolókba zárt statikus egységeknek tekintenénk, az adattér egy dinamikus ökoszisztémát képzel el, ahol az adatok zökkenőmentesen áramlanak a különböző források és felhasználók között. Az adatfelületek egységes platformot biztosítanak az adatintegrációhoz és interakciókhoz, és így új felismerések felszabadítását, az innováció elősegítését és a különböző területeken való együttműködés ösztönzését ígérik.

Az adatok interoperabilitására és hozzáférhetőségére való törekvés közepette az adatminőség biztosítása kiemelt fontosságú kérdésként merül fel. Az adattereken belül az adatok megbízhatóságának - jóságának - biztosítása fontos a résztvevők számára. Ez számos szempontot foglal magában, többek között az adatok pontosságát, konzisztenciáját, teljességét és származását.

Ennek fényében ez a diplomaterv az adattereken belüli adatminőségbiztosítás bonyolultságának feltárására vállalkozik, különös tekintettel az adatok jóságára. A dinamikus és összekapcsolt környezetekben az adatok integritásának és megbízhatóságának biztosításával kapcsolatos kihívások, módszerek és legjobb gyakorlatok feltárásával ez a kutatás hozzá kíván járulni az adatközpontú ökoszisztémák fejlődéséhez.

# Abstract

In today's increasingly data-centric world, where information serves as the lifeblood of countless industries and endeavors, the value of data cannot be overstated. From personal insights to organizational decision-making, the quality and reliability of data hold immense significance. Consequently, individuals and entities across the globe invest substantial resources in safeguarding their data, often resorting to private infrastructures to ensure its integrity and confidentiality.

However, as data proliferates and diversifies, the siloed nature of private infrastructures poses challenges to efficient data utilization and exchange. Recognizing the need for a more cohesive and interconnected approach, there arises an imperative to develop a system that effectively bridges the gap between data providers and consumers. This imperative gives rise to the concept of dataspace – a framework aimed at revolutionizing the landscape of data management and exchange.

At its core, dataspace represent a paradigm shift in how we perceive and interact with data. Rather than viewing data as static entities confined within isolated repositories, dataspace envision a dynamic ecosystem where data flows seamlessly between disparate sources and users. By providing a unified platform for data integration and interaction, dataspace offer the promise of unlocking new insights, fostering innovation, and driving collaboration across domains.

Crucially, amidst the pursuit of data interoperability and accessibility, the assurance of data quality emerges as a paramount concern. Within the realm of dataspace, ensuring the reliability and trustworthiness of data – often referred to as data veracity – takes precedence. This encompasses a spectrum of considerations, including data accuracy, consistency, completeness, and lineage, among others.

Against this backdrop, this thesis embarks on a journey to explore the intricacies of data quality assurance within dataspace, with a specific focus on data veracity. By delving into the challenges, methodologies, and best practices associated with ensuring the integrity and reliability of data in dynamic and interconnected environments, this research seeks to contribute to the advancement of data-centric ecosystems.

# Chapter 1

## Introduction

Intro

### 1.1 Problem Definition and Motivation

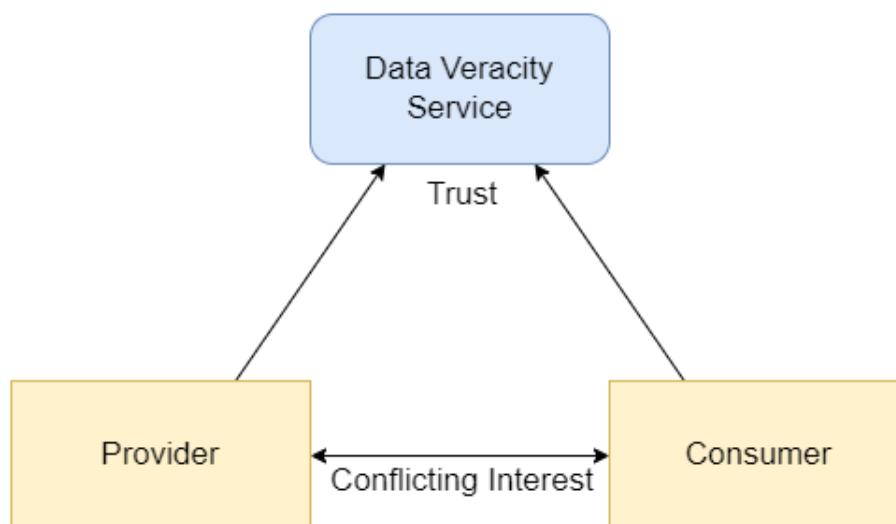


Figure 1.1: Trust of the Participants.

### 1.2 Blockchain as a Solution

blockchain

### 1.3 Structure of the Document

## **Chapter 2**

# **Data Spaces**

### **2.1 Summary**

### **2.2 Initiatives**

#### **2.2.1 DSSC**

#### **2.2.2 Gaia-X**

#### **2.2.3 IDSA**

Eclipse DSC

#### **2.2.4 Catena-X**

#### **2.2.5 Prometheus-X**



## **Chapter 3**

# **Quality Assurance**

system  
data

### **3.1 Data Quality**

#### **3.1.1 Volume**

#### **3.1.2 Variety**

#### **3.1.3 Veracity**

#### **3.1.4 Value**

#### **3.1.5 Velocity**

### **3.2 Data Veracity**

### **3.3 State of the Art**

# **Chapter 4**

## **Architecture**

### **4.1 Demands**

#### **4.1.1 Separated Component**

SoC

#### **4.1.2 Not Slowing the Transfer**

Depends on access type

#### **4.1.3 Correlating Process Time**

Non-architectural

#### **4.1.4 Validate Check**

For trust

### **4.2 Questions**

#### **4.2.1 How to Validate**

Trust central / multiple check

#### **4.2.2 Who Can Access the Data**

Delegates / transfer participants

#### **4.2.3 How the Data Is Accessed**

Direct transfer / pointer

## 4.3 Possible Solutions

From permissive to strict in the aspect of access

### 4.3.1 Data Space Component

### 4.3.2 Delegates with Oraculum

### 4.3.3 Own Checkers

## 4.4 Summary

	Component	Oraculum	Own
Validation	trust central	trust / multiple check	multiple check
Access right	delegate	delegate / participants	participants
Access method	pointer / direct	pointer	pointer / direct

Table 4.1: Caption

## **Chapter 5**

# **Implementation**

### **5.1 Technologies**

#### **5.1.1 Express.js**

#### **5.1.2 OpenAPI**

#### **5.1.3 Hyperledger Fabric**

### **5.2 Implemented Solution**

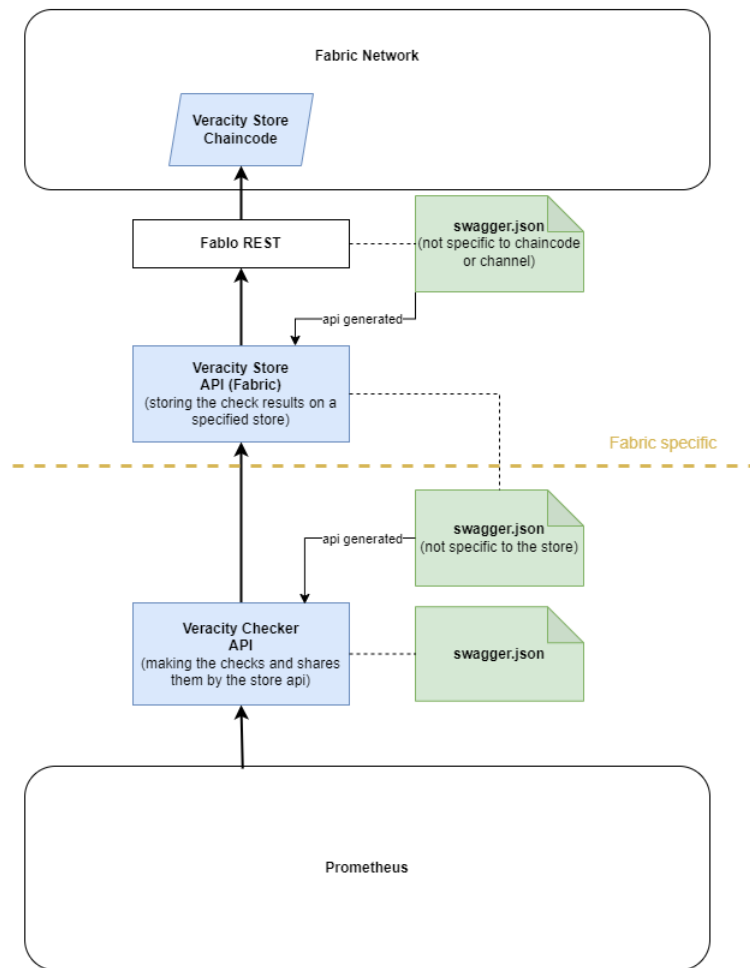


Figure 5.1: Architecture of the implemented system.

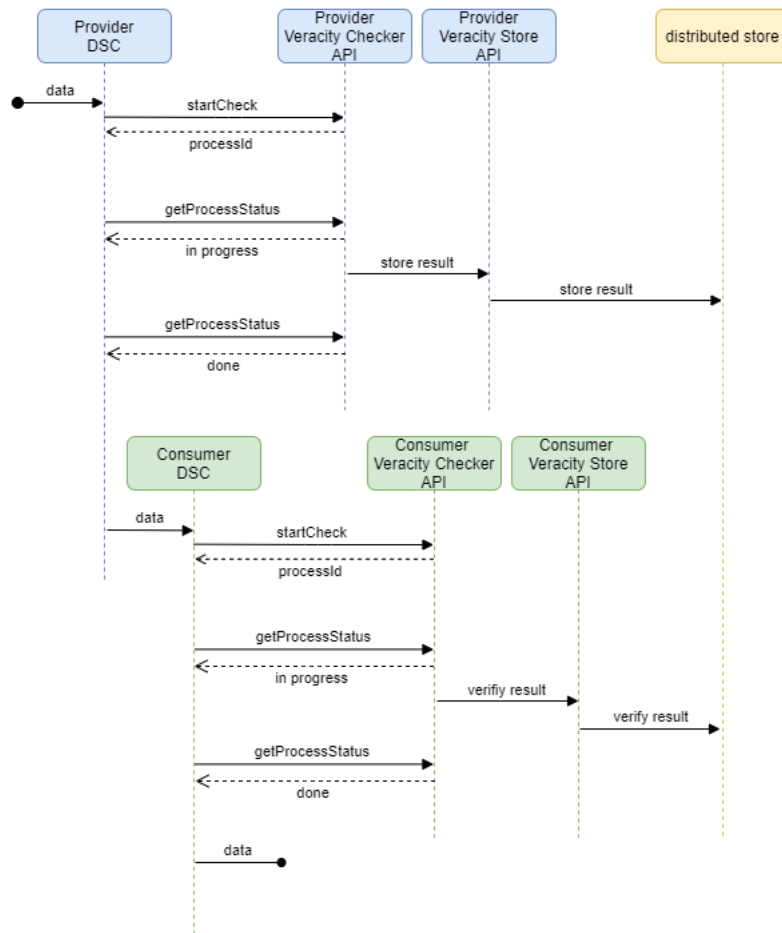


Figure 5.2: Check with verification call flow.

# **Bibliography**

# Acronyms



# Index

blockchain, 1