THE IMPORTANCE OF A UNIFIED DATA SPACE IN DATA PROCESSING AND METHODS FOR ITS DEVELOPMENT

Ma'mura Gayratovna Ergasheva
Pedagogue, Department of general technical sciences
Asia International University

Annotation: This article discusses the importance of a Unified Data Space in modern data processing and explains the main methods for its development. A Unified Data Space helps organizations collect, store, and manage data from different sources in one system. It improves data accuracy, accessibility, and security, which are essential for effective decision-making. The article also highlights how unified data systems reduce data duplication and increase the speed of information exchange between departments. In addition, it describes key methods for developing such systems, including data integration technologies, cloud computing solutions, and standardization of data formats. The study shows that implementing a Unified Data Space helps organizations work more efficiently, supports digital transformation, and improves the quality of data analysis. Overall, the article emphasizes that unified data management is an important part of modern information technologies.

Key words: Unified Data Space, data processing, data integration, data management, cloud computing, digital transformation, data standardization, information systems, data security, data analysis

Introduction. In the era of digital transformation, the rapid growth of data across various domains has created significant challenges in data management, integration, and analysis. Organizations increasingly rely on heterogeneous data sources, distributed systems, and diverse software platforms, which often operate independently and lack interoperability. This fragmentation reduces efficiency, complicates data processing and limits the potential for comprehensive analysis and informed decision-making.

A unified data space has emerged as a conceptual and technological solution to these challenges. It provides a structured and standardized environment in which data from multiple sources can be integrated, harmonized, and accessed consistently[3]. By ensuring compatibility, semantic consistency, and centralized or logically connected data governance, a unified data space enhances data quality, reduces redundancy, and supports advanced analytical processes.

The development of such an environment requires carefully designed architectural frameworks, standardized data models, interoperability protocols, and robust security mechanisms. Modern technologies, including cloud computing, distributed databases, and data virtualization, play a critical role in establishing scalable and secure unified data infrastructures.

Data and a unified space are closely connected. A unified space is created to organize, manage, and integrate data from different sources into one common environment. Without data, a unified space has no purpose, and without a unified space, data can become disorganized and difficult to use.

In many systems, data is stored in separate databases or platforms. This can lead to inconsistency, duplication, and errors. A unified space connects these different data sources and

ensures that they follow common standards and formats. As a result, the data becomes more accurate, structured, and easier to analyze.

The unified space also improves data accessibility. Users can access needed information from one integrated system instead of searching in multiple independent systems. This saves time and increases efficiency.

In fact, What is a Unified Space? A unified space is an integrated environment where different types of data, systems, and technologies are connected and organized in a single structured framework. It allows information from various sources to be collected, stored, processed, and accessed in a consistent and coordinated way.

In many organizations, data is stored in separate systems that do not easily communicate with each other. This can cause duplication, errors, and difficulties in analysis. A unified space solves this problem by bringing all data into one common environment with shared standards and rules[4].

The main purpose of a unified space is to improve data management, increase efficiency, and support better decision-making. It ensures that information is accurate, accessible, and secure. Modern technologies such as cloud computing, databases, and integration platforms are often used to create a unified data space.

Imagine a hospital that stores patient information in three different systems:

1. Electronic Health Records (EHR) – stores medical history
2. Laboratory System – stores lab test results
3. Pharmacy System – stores prescribed medications

A unified space would combine data from all three systems into a single structured database. This allows doctors to access all patient information in one place without switching between systems.

```
import pandas as pd
ehr_data = pd.DataFrame({
    'PatientID': [1, 2, 3],
    'Name': ['Alice', 'Bob', 'Charlie'],
    'Age': [30, 45, 28],
    'Diagnosis': ['Flu', 'Diabetes', 'Asthma']
})
lab_data = pd.DataFrame({
    'PatientID': [1, 2, 3],
    'Test': ['Blood Test', 'Urine Test', 'X-Ray'],
    'Result': ['Normal', 'High Sugar', 'Clear']
})

pharmacy_data = pd.DataFrame({
    'PatientID': [1, 2, 3],
    'Medication': ['Paracetamol', 'Insulin', 'Inhaler'],
```

1357

```
        'Dose': ['500mg', '10 units', '2 puffs']
})
```

```
unified_space    =    ehr_data.merge(lab_data,    on='PatientID').merge(pharmacy_data, on='PatientID')
print("Unified Space:")
print(unified_space)
```

RESULT:
Unified Space:

PatientID	Name	Age	Diagnosis	Test	Result	Medication	Dose
1	Alice	30	Flu	Blood	Normal	Paracetamol	500mg
2	Bob	40	Diabetes	Urine	High Sugar	Insulin	10 units
3	Charlie	28	Asthma	X-Ray	Clear	Inhaler	2 puffs

Explain

- PatientID is the key that links all systems.
- The merge function combines data into a single unified space.
- Now, all patient information is organized, structured, and easy to access.

This is a simple demonstration of a unified space in Python. In real applications, unified spaces use databases, cloud systems and APIs[5] to integrate much larger datasets. Each system's data can be represented as a table, which mathematically is a set of tuples.

Here, each tuple (PatientID,Name,Age,Diagnosis) (PatientID, Name, Age, Diagnosis)(PatientID,Name,Age,Diagnosis) is an element of the set. Other systems, like Lab and Pharmacy, can be represented similarly:

A unified space combines all patient information into one set using the PatientID as a key: Mathematically, the operation is a join function that combines two sets based on a matching key:

Relation to the Python Code:
```
unified_space    =    ehr_data.merge(lab_data,    on='PatientID').merge(pharmacy_data, on='PatientID')
```
- The merge function performs a natural join
- on='PatientID' specifies the join key
- The result is a single unified space containing all patient information

Definatially, there are a lot of methods for creating a unified Space[9]. For example,
Data Integration –
Combine data from different systems, databases, or platforms.
Use ETL (Extract, Transform, Load) processes:
- Transform – clean, standardize, and harmonize data formats
- Load – insert the data into a unified system

1358

Relational Database Joins
- Use relational database techniques to connect tables with common keys.
- Types of joins:
  - Inner join – keep only matching records
  - Left/right join – keep all records from one table and matching from another
  - Full outer join – keep all records from all tables
- This method is often used in Python (pandas) or SQL-based unified spaces.

Data Warehousing
- Create a centralized data warehouse that stores structured data from multiple sources.
- Supports large-scale analysis and reporting.
- Often combined with OLAP (Online Analytical Processing) for multidimensional queries.

Data Virtualization
- Instead of moving all data, create a virtual unified layer that accesses data from different sources in real-time.
  - Advantages: No duplication, real-time access, flexible and scalable

Standardization and Interoperability
- Apply common data formats, codes, and protocols (e.g., HL7 or FHIR in healthcare).
- Ensures that data from different systems can be understood and processed uniformly.

Cloud-based Integration Platforms
- Use cloud services or APIs to integrate data from multiple systems into a single unified environment.
- Supports scalability, security, and remote access.

Conclusion:

A unified space plays a critical role in modern data management and decision-making. By integrating data from multiple sources into a single structured environment, it ensures consistency, accuracy, and accessibility of information. This reduces errors, eliminates data duplication, and allows organizations to analyze information more effectively.

In healthcare, a unified space enables doctors and staff to access all patient information in one place, improving the quality of care and speeding up clinical decisions. More generally, unified spaces support efficient resource management, real-time analytics, and informed strategic planning across industries.

Overall, the creation of a unified space transforms fragmented data into a coherent system, enhancing both operational efficiency and the ability to make data-driven decisions.

References:

1. Bonnefoy P.-Y., Chaize E., Mansuy R., Tazi M. The Definitive Guide to Data Integration. Birmingham: Packt Publishing, 2024.
2. Wells D. Data Interoperability: Unified Architecture Connecting All of Your Data. New York: CRC Press, 2021.
3. Jiang Sh., Yu Sh. Research on Data Integration in Dataspace. London: Scientific.net, 2020.

4. Jarke M., Quix C. Federated Data Integration in Data Spaces. Cham: Springer, 2022.
5. Auer S. Semantic Integration and Interoperability. Cham: Springer, 2022.
6. Nadal S. et al. An Integration-Oriented Ontology to Govern Evolution in Big Data Ecosystems. Berlin: Springer, 2018.
7. Lu J., et al. UDBMS: Road to Unification for Multi-model Data Management. Berlin: Springer, 2016.
8. Kotis K., Athanasakis I., Vouros G. Semantic Integration & Single-Site Opening of Multiple Governmental Data Sources. arXiv, 2014.
9. Understanding Data Spaces: A Systematic Mapping Study of Foundations, Technical Building Blocks, and Sectoral Adoption. Amsterdam: Elsevier, 2025.