

Article

Not peer-reviewed version

---

# HCI and Data: Interacting in a New Era of Virtualization

---

[Iván Durango Blanco](#)\*, [José Antonio Gallud Lázaro](#)\*, [Victor M. Ruiz Penichet](#)\*

Posted Date: 8 May 2024

doi: 10.20944/preprints202405.0508.v1

Keywords: Human-Data Interaction; Human-Computer Interaction; Big Data; Data virtualization; Data Accessibility; Data Management; Data Privacy; Data Ethics; Data-Driven Decision-Making



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# HCI and Data: Interacting in a New Era of Virtualization

Iván Durango Blanco <sup>1,\*</sup> , José Antonio Gallud Lázaro <sup>2</sup>  and Victor M. R. Penichet <sup>2</sup> 

<sup>1</sup> University of Castilla-La Mancha, 02071 Albacete, Spain

<sup>2</sup> Albacete Research Institute of Informatics, University of Castilla-La Mancha, 02071 Albacete, Spain; jose.gallud@uclm.es (J.A.G.L.); victor.penichet@uclm.es (V.M.R.P.)

\* Correspondence: ivan.durango@alu.uclm.es

**Abstract:** This article explores the intricate relationship between Human-Data Interaction (HDI) and Data Virtualization (DV), two integral fields that are redefining how we engage with and manage data in the digital age. Positioned at the intersection of technology and user-centric design, this study examines how DV can transcend traditional data management boundaries to significantly enhance data accessibility, legibility, and agency, thus fostering more ethical and meaningful interactions across various sectors. Opening with a detailed exposition on the evolving role of HDI, the paper contextualizes the importance of user-centric data practices within contemporary digital challenges. HDI's principles are pivotal as they advocate for user empowerment in data interactions, ensuring that data processes are not only transparent but also negotiable and aligned with users' needs and ethical standards. DV emerges as a transformative solution that inherently supports these principles by integrating diverse data sources into a cohesive and accessible format, thereby eliminating many of the inefficiencies found in conventional data management systems. Through a series of nuanced case studies, the paper provides empirical evidence of DV's effectiveness across diverse domains. In healthcare, DV facilitates a holistic approach to patient care by providing seamless access to comprehensive medical records, thereby enabling faster, more accurate diagnostic and treatment processes. In the financial sector, DV proves instrumental in enhancing the detection and management of fraud, integrating transactional data across global platforms in real-time to allow for swift, proactive interventions. In education, DV supports personalized learning by synchronizing data from various educational tools, thus allowing for tailored educational experiences that respond dynamically to student performance and engagement metrics. Moreover, the paper critically addresses the ethical considerations and privacy concerns that arise with the integration of HDI and DV, particularly as data interactions become more pervasive and complex. By proposing forward-thinking research directions, it seeks to refine these technologies to better serve societal needs while addressing potential risks and governance issues. In conclusion, this study significantly contributes to the discourse on HDI by advocating for DV as an essential, innovative tool that enhances the clarity, fairness, and efficacy of data interactions. By demonstrating how DV can transform large-scale data sets into user-friendly formats that support decision-making and improve operational efficiencies, the paper underscores DV's potential to advance a range of industries and enhance the daily lives of individuals interacting with digital systems.

**Keywords:** human-data interaction; human-computer interaction; big data; data virtualization; data accessibility; data management; data privacy; data ethics; data-driven decision-making

## 1. Introduction

In the digital era, the convergence of Human-Data Interaction (HDI) and Data Virtualization (DV) stands as a cornerstone for enhancing our ability to access, understand, and ethically manage data. This paper explores the synergistic potential of these technologies to transform data-driven decision-making across various sectors.

HDI, as articulated by [1], centers on three fundamental aspects: data legibility, agency, and negotiability. These principles underscore the importance of enhancing data accessibility, empowering users to control their data, and fostering transparent interactions between data providers and users. Concurrently, DV offers a dynamic method to integrate diverse data sources efficiently and flexibly, as described by [2], presenting a significant shift from traditional data management methods.

The challenges posed by unstructured data from social networks, texts, images, and videos are profound, as these data forms resist simple analysis and interpretation [3]. Moreover, ethical considerations such as data ownership, consent, and the transparency of data use are increasingly critical [4–6]. These issues are compounded by a general lack of data literacy, which can hinder the effective use and understanding of data and visualizations [7].

This paper also addresses the opacity of machine learning algorithms and the issues of trust and power imbalance that arise from the uneven distribution of data control [8–10]. These factors emphasize the need for greater transparency and fairness in automated decision-making systems.

By integrating recent case studies, this article aims to illustrate how DV can not only streamline access to diverse data sources but also enhance the comprehension and usability of data, facilitating more informed and equitable decision-making processes. Through this exploration, we contribute to the expanding body of knowledge in HDI and outline future research directions and practical applications for these pivotal technologies.

Following this introduction, the paper provides a comprehensive review of the theoretical foundations of HDI and DV (Section 2), Proposes an enhancement to the HDI through the use of data virtualisation (Section 3), Relevant use cases are studied (Section 4), proposing several use cases for the use of HDI and data virtualisation (Section 5) and setting out the idea of creating a framework (Section 6), identifies research challenges (Section 7), and concludes with proposals for future research (Section 8).

## 2. Background and Concepts

This section delves into the theoretical underpinnings that support the study of Human-Data Interaction (HDI) and DV, providing a foundation to understand their significance and potential impact within various fields.

HDI is a multidisciplinary field focused on the design, analysis, and evaluation of systems where humans interact with data. Mortier [1] outlines three core challenges in HDI: readability, agency, and negotiability of data. These challenges emphasize the importance of making data accessible and understandable, empowering users to control their own data, and fostering a constructive dialogue between users and data providers.

Contrasting traditional data management approaches, DV introduces a paradigm shift towards more agile and flexible data handling [2]. It addresses the modern challenges of handling diverse and dispersed data sources by enabling a unified view of data. This approach not only simplifies access and analysis but also ensures consistency and efficiency across data-driven activities.

The sections below will provide a detailed exploration of HDI and Data Virtualisation, examining the relationship between the two subjects and discussing how advancements in data virtualisation technology can enhance human-data interactions in various applications.

### 2.1. Human-Data Interaction (HDI)

Human-Data Interaction (HDI) is an emerging field dedicated to improving how people access, understand, and use data across diverse contexts, from personal decisions to public policy. The essence of HDI is to design, study, and evaluate systems that enhance data interactions, making them more intuitive and impactful. Core Aspects of HDI:

1. Data Accessibility: HDI prioritizes ease of access to data through user-friendly interfaces, supporting diverse user needs and contexts [11].

2. Data Understanding: This involves presenting data in ways that are easy to understand and analyze, leveraging advanced visualization techniques to clarify complex information [12].
3. Data Interaction: HDI focuses on interactive systems that allow users to manipulate and explore data to derive meaningful insights [13].
4. Privacy and Ethics: Ensuring that data systems uphold ethical standards and protect user privacy is a cornerstone of HDI [14].

#### HDI's Role in Various Sectors

1. Healthcare: HDI integrates real-time data from wearables and sensors, enhancing diagnostic capabilities and enabling personalized medicine. Challenges include improving the design and usability of electronic health records (EHRs) [15].
2. Business and Marketing: Companies utilize HDI to analyze consumer behavior, optimize marketing strategies, and enhance decision-making through data visualization and analytics [16].
3. Education: Through learning analytics, HDI provides insights into student performance, supporting personalized learning and adaptive educational systems [17].
4. Government: HDI-driven data insights help in policy making and service delivery, with smart cities using real-time data for urban planning and management [18].
5. Finance: HDI underpins financial models that detect fraud and aid in decision-making processes, with increasing reliance on algorithmic trading and risk assessment tools [19].
6. Everyday Life: From social media to navigation apps, HDI influences daily activities by tailoring digital interactions based on user data, significantly shaping our consumption patterns and daily routines [20].

As HDI continues to evolve, it permeates every aspect of our lives, offering numerous opportunities for research and innovation. By enhancing the interaction between humans and data, HDI plays a pivotal role in advancing technology's integration into daily and professional life. The ongoing development of HDI strategies promises to refine these interactions further, ensuring they are more beneficial, ethical, and user-centered.

#### 2.2. Data Virtualization: Key Concepts

Data virtualisation is a technology that enables the access and manipulation of data from multiple heterogeneous sources as if they were located in a single repository. This technology differs from traditional data integration methods because it uses an intermediate layer to translate queries across multiple data sources, thus eliminating the need for physical integration or data replication [21].

DV prioritises efficiency and flexibility over traditional data management techniques like data replication, which involves copying data to a new location. virtualization allows for real-time queries across different sources without duplication, offering more dynamic and scalable access to data [2]. This approach not only reduces the costs and time associated with data integration but also enhances business agility and enables faster and more efficient access to information.

The primary advantages of DV include its ability to provide agile and flexible data handling, efficient data integration, and improved decision-making capabilities. These benefits stem from the technology's ability to provide a unified view of dispersed data, facilitating quick access and analysis that is essential for data-driven decision making [22]. Additionally, DV reduces the need for redundant ETL (Extract, Transform, Load) processes and data storage, significantly cutting down infrastructure and operational costs [23].

However, DV also presents challenges, particularly in managing data quality and ensuring security. As DV continues to expand its footprint, addressing these challenges becomes crucial for maintaining the integrity and confidentiality of data systems.

With the increasing adoption of DV, ethical and privacy concerns have become more prominent. It is vital that virtualization solutions are designed with a strong emphasis on data ethics and user privacy to safeguard against potential breaches and misuse of data [24].

In summary, DV stands out as a key technology in the information age, enhancing the capabilities of organizations to manage and analyze data efficiently. By providing an efficient and cost-effective approach to handle data from diverse sources, DV not only supports business agility but also contributes to more informed and quicker decision-making across industries.

### 3. Enhancing HDI from Data Virtualization

This paper investigates the transformative impact of DV on Human-Data Interaction (HDI), aiming to bridge the gap between technological advancements in DV and the evolving practices in HDI. By offering a unified view of heterogeneous data sources, DV enables real-time access and effective data management, enhancing the accessibility, understandability, and negotiability of data, which are central tenets of HDI.

DV brings operational efficiencies by allowing a single point of access to disparate data, thereby simplifying complex data interactions [25]. This approach not only eliminates the need for data duplication but also ensures that users can manage and interact with data more effectively, aligning with HDI's goals of enhancing user control and interaction with data providers.

Improving the accessibility and usability of data is essential for fostering effective human development [11]. DV addresses this by simplifying the inherent complexities of distributed data systems, providing intuitive data representations that enhance user comprehension and interaction [26]. This is crucial in making HDI more inclusive and user-friendly, particularly for non-technical users.

The integration of DV within HDI also brings to the forefront the need for robust ethical and privacy frameworks. As [14] emphasizes, considering ethics and privacy in the design of data systems is imperative. DV facilitates stricter data control, allowing for enhanced privacy measures and compliance with regulations [27], thus addressing potential privacy concerns effectively.

In practical settings, such as healthcare, DV proves invaluable. It enables healthcare professionals to access and correlate patient data from multiple sources seamlessly, enhancing clinical decision-making and patient care quality [28]. Similarly, in business settings, DV tools help in managing large datasets efficiently, improving decision-making processes without compromising data security [13].

DV tools, ranging from software to advanced analytics and visualization platforms, significantly enhance the ability of users to understand and manipulate large datasets. The technology supports semantic enrichment of data, which adds meaning and improves the data's usability for decision-making [29]. Additionally, these systems can provide tailored views of data and support natural language queries, which reduce complexity and enhance user engagement [30].

Looking forward, DV holds the potential to redefine HDI by simplifying data access and fostering a human-centric approach in data management. The ongoing innovation in data visualization tools and techniques will continue to play a crucial role in this transformative process [13]. As we advance, it will be vital to develop and refine ethical frameworks that align with these technological innovations, ensuring that DV continues to enhance HDI in an ethical, transparent, and user-focused manner.

### 4. Case Studies

This section provides detailed case studies that illustrate the practical impact of DV on HDI in various contexts. The analysis of specific case studies demonstrates where DV has had a significant impact on HDI.

In the healthcare sector, DV has proven to be an invaluable resource. One notable case is the integration of patient data from multiple hospital systems. According to [28], DV has enabled medical professionals to access unified patient information, improving diagnostic accuracy and treatment efficacy. This case highlights how virtualization can efficiently manage sensitive and complex data, contributing significantly to healthcare decision-making.



In the financial sector, DV has facilitated data-driven decision-making and fraud detection. An example of the use of virtualization is to integrate transaction data from different banking platforms and systems. This approach, has enabled financial analysts to identify anomalous patterns and prevent financial fraud more effectively.

The education sector has also seen significant benefits through DV. The integration of student performance data from multiple sources has enhanced the personalisation and effectiveness of learning. Borner [31] highlights how DV has enabled educational institutions to gain a comprehensive view of student progress, facilitating more informed and personalised educational interventions.

The case studies above provide valuable lessons and best practices for applying DV in the context of HDI. These findings are essential for guiding future implementations and developments in the field.

One key lesson from the healthcare sector case studies is the importance of user-centric data integration. Raghupathi [28] observed that DV in the healthcare sector must be designed with the specific needs of end-users, such as healthcare professionals, in mind. This requires intuitive interfaces and efficient access to relevant data to improve decision-making and service quality.

The financial sector highlights the critical importance of security and privacy in DV. Studies such as [24] suggests implementing robust security measures to protect against unauthorised access and ensure data confidentiality. This includes using encryption, authentication, and stringent access controls.

In the education sector, the adaptability and scalability of virtualization solutions are crucial, as discussed by [31]. virtualization platforms should be capable of adapting to constantly changing data volumes and a variety of data types to meet the dynamic needs of the education sector.

A cross-cutting lesson from all sectors is the importance of interdisciplinary collaboration in DV projects. Therefore, effective interaction between IT experts, data specialists, and end-users is critical to the success of these initiatives. This collaboration ensures that the solutions developed are not only technically sound but also practically relevant and usable.

## **5. Empowering Real-World Applications: In-Depth Use Cases of DV in Healthcare, Finance, and Education**

In this paper, we aim to create detailed use cases for DV in three key sectors: healthcare, finance, and education. These real-world use cases will provide specific scenarios that illustrate how DV can significantly impact decision-making processes, operational efficiency, and personalised services in each sector.

### *5.1. Healthcare: Enhanced Patient Care through Integrated Data Systems*

In the healthcare industry, efficient interaction with data is crucial for effective patient care. DV serves as a transformative technology that aligns with HDI principles by simplifying the access and integration of dispersed patient data. This study explores how DV enhances the interaction between healthcare professionals and data, focusing on improving diagnostic accuracy and treatment efficiency through enhanced data legibility, agency, and negotiability.

A large healthcare provider utilizes DV to integrate diverse medical data sources, including electronic health records (EHRs), laboratory results, imaging data, and patient-generated data from wearable devices. This integration allows healthcare professionals to access a unified patient profile in real time.

#### **5.1.1. HDI Objectives and Study Hypotheses**

The primary objectives of this study are framed around core HDI principles:

1. **Data Legibility:** Ensure that the integrated patient data is presented in an intuitive and understandable format for medical professionals, regardless of their technical skills. Visualization tools like patient health dashboards can display complex health data (like real-time vitals and historical health trends) in a clear and actionable manner.

2. Data Accessibility: Assess if DV tools provide healthcare professionals with more accessible and comprehensive patient data.
3. Data Agency: Evaluate whether DV tools empower healthcare professionals by enabling more control over data manipulation and interpretation. This enables them to view their own health data, contribute information, and make informed decisions about their treatment options.
4. Data Negotiability: Determine if DV tools facilitate better negotiation between data sources and users, enhancing the transparency and customization of data interactions. This can be supported through digital consent tools that allow patients to manage who has access to their data and for what purpose.

#### 5.1.2. Methodology

A randomized controlled trial (RCT) was conducted among healthcare professionals using DV tools versus those using traditional data systems. The study measured:

- Diagnostic Accuracy: Comparing the correctness of diagnoses made using DV and traditional systems.
- Speed of Diagnosis: Time taken to arrive at a diagnosis using both systems.
- User Satisfaction: Based on the ease of use, accessibility, and control over data.

#### 5.1.3. Data Flow

- Data Collection: Patient data from various hospitals and clinics are collected. This includes historical health records, real-time monitoring through wearables, and results from recent medical tests.
- Data Integration: DV software consolidates data from these varied sources without needing to store it in a central repository, respecting the privacy and security regulations like HIPAA.
- Data Usage: Physicians access a comprehensive dashboard that provides a holistic view of a patient's health status, enabling better diagnosis and personalized treatment plans.

#### 5.1.4. Results

The results demonstrated:

- Improved Data Accessibility: Professionals using DV had quicker access to comprehensive patient data, leading to faster and more accurate diagnoses.
- Enhanced Data Agency: DV tools allowed more direct manipulation of data, giving healthcare providers greater confidence in their diagnostic decisions.
- Effective Data Negotiability: DV facilitated a more flexible interaction with data sources, allowing for customized views and reports tailored to individual patient needs.

Benefits:

- Improved Diagnostic Accuracy: Access to comprehensive, real-time patient data helps in diagnosing diseases earlier and more accurately.
- Personalized Treatment: Insights derived from a unified data view allow for treatments tailored to individual patient needs and conditions.
- Efficient Care Delivery: Reduces the time doctors spend gathering information, allowing more time to focus on patient care.

#### 5.1.5. Discussion

The study underscores the importance of aligning DV technologies with HDI principles in healthcare settings. By enhancing data accessibility, agency, and negotiability, DV tools not only improve diagnostic processes but also empower healthcare professionals to deliver personalized care efficiently. These improvements directly contribute to better patient outcomes and operational efficiencies in healthcare institutions.

#### 5.1.6. Conclusion

This empirical evaluation highlights the transformative impact of DV from an HDI perspective, offering significant evidence that DV can enhance the way healthcare professionals interact with and utilize data. Future initiatives should continue to focus on these principles to further refine and enhance the integration of DV technologies in healthcare.

#### 5.2. Finance: Fraud Detection and Risk Management

In the finance sector, the rapid detection and management of fraud are critical for maintaining the integrity of financial systems. DV facilitates a seamless and dynamic interaction between financial analysts and expansive, diverse data sources, embodying core HDI principles such as data accessibility, agency, and negotiability. A multinational bank employs DV to merge transaction data across different systems and geographical locations to detect fraudulent activities and manage financial risks effectively.

##### 5.2.1. HDI Objectives and Study Hypotheses

This study centers on evaluating the impact of DV in finance using HDI principles:

1. **Data Legibility:** Design user interfaces for fraud detection systems that allow financial analysts to easily navigate and interpret transaction data from various sources. Use graphical representations to highlight patterns and anomalies that may indicate fraudulent activities.
2. **Data Accessibility:** Determine if DV tools enhance the accessibility of financial data across different platforms and systems for real-time fraud detection.
3. **Data Agency:** Assess whether DV empowers analysts by providing more control over data queries and manipulations. Provide customers with tools to monitor their own transaction activities and report suspicious actions directly through banking apps. This increases user engagement and enhances the effectiveness of fraud detection systems.
4. **Data Negotiability:** Explore if DV improves the ability of analysts to negotiate and customize how data is presented and analyzed for better fraud detection outcomes. Implement transparent policies regarding how customer data is used for fraud detection and risk assessments. Offer customers options to opt-in or opt-out of certain data collection practices, reinforcing trust and compliance.

##### 5.2.2. Methodology

A controlled experimental setup was used, involving financial analysts who utilized DV tools versus those using traditional data integration methods. Metrics measured included:

- **Fraud Detection Rates:** Accuracy and number of fraud cases detected.
- **Response Times:** Time efficiency in detecting and responding to potential fraud.
- **User Satisfaction:** Surveyed based on the adaptability, efficiency, and user control of the data system.

##### 5.2.3. Data Flow

- **Data Aggregation:** Transaction data from various bank branches and online banking services are integrated in real time.
- **Anomaly Detection:** Advanced analytics tools, running on top of the virtualized data layer, identify unusual patterns that suggest potential fraud.

##### 5.2.4. Results

The study revealed:

- **Enhanced Data Accessibility:** Analysts using DV systems accessed comprehensive transaction data more swiftly, enabling quicker responses to fraudulent activities.



- Increased Data Agency: DV tools provided analysts with enhanced capabilities to manipulate and analyze data on-the-fly, fostering proactive fraud detection strategies.
- Improved Data Negotiability: DV allowed for better customization of data views and analytic models, which tailored the fraud detection process to specific needs of the institution.

#### Benefits:

- Enhanced Fraud Detection: Real-time data access enables quicker response to fraudulent activities, reducing potential losses.
- Dynamic Risk Assessment: Continuously updated data allows for more accurate and timely risk assessments, improving the bank's financial stability.
- Regulatory Compliance: Easier compliance with global financial regulations through centralized monitoring and reporting of transaction data.

### 5.2.5. Discussion

The findings highlight the crucial role of DV in enhancing human-data interaction within the finance sector. By improving data accessibility, agency, and negotiability, DV enables financial analysts to perform more effectively in detecting and managing fraud. This not only enhances the security and reliability of financial operations but also supports a more adaptive financial analysis environment.

### 5.2.6. Conclusion

This empirical evaluation demonstrates how integrating DV in finance, viewed through an HDI lens, significantly improves fraud detection and risk management. Future work should continue to leverage HDI principles to advance the development and application of DV technologies, ensuring that financial systems are both robust against fraud and adaptable to the evolving dynamics of financial data.

## 5.3. Education: Personalized Learning and Performance Tracking

The education sector requires robust, dynamic data systems to tailor learning experiences to diverse student needs. DV presents a technological advancement that aligns with HDI principles by simplifying the integration and interaction with educational data. This study examines how DV enhances the educational process by improving the accessibility, agency, and negotiability of educational data, thereby facilitating personalized learning and effective performance tracking.

An educational institution implements DV to integrate data from various sources like Learning Management Systems (LMS), student forums, and online assessment tools to create personalized learning experiences and track student performance dynamically.

### 5.3.1. HDI Objectives and Study Hypotheses

This study aims to evaluate the impact of DV on enhancing personalized educational experiences using HDI principles:

1. Data Legibility: Ensure that data collected from various educational tools is integrated and presented back to both students and educators in an accessible format. Dashboards that show academic progress, areas of strength, and areas needing improvement can help make educational decisions more data-informed.
2. Data Accessibility: Assess whether DV tools provide educators and students with easier access to comprehensive learning data.
3. Data Agency: Determine if DV empowers educators to manipulate and apply data in ways that enhance personalized learning. Allow students to interact with their own performance data and set personal academic goals. Provide tools that let them customize their learning experience, such as choosing elective topics based on performance trends.

4. Data Negotiability: Explore how DV enables educators and students to negotiate the terms of data usage, customizing data interactions to better meet individual learning needs. Create channels for students and educators to provide feedback on the data collection and analytics processes. This can help refine the data integration to better serve educational goals and adapt to the needs of diverse student populations.

### 5.3.2. Methodology

The study involved a comparative analysis between educational institutions using traditional data management systems and those employing DV tools. Key metrics included:

- Personalization Effectiveness: The ability of educators to tailor learning experiences based on integrated data insights.
- Learning Outcomes: Measured improvements in student performance and engagement.
- User Satisfaction: Feedback from educators and students on their experiences with the data systems.

### 5.3.3. Data Flow

- Data Integration: Collect and integrate data from classroom interactions, online quizzes, and feedback sessions.
- Personalized Feedback: Use data analysis to provide real-time feedback and personalized learning pathways for students.
- Performance Monitoring: Teachers and administrators use integrated data to monitor and analyze student performance over time.

### 5.3.4. Results

Findings from the study highlighted:

- Improved Data Accessibility: DV systems allowed for real-time access to a wide array of educational data, facilitating a quicker adaptation of learning strategies to student needs.
- Enhanced Data Agency: Educators using DV reported greater control over data, enabling them to design more effective, personalized instructional methods.
- Increased Data Negotiability: DV provided customizable interfaces and analytics, allowing both students and educators to interact with data in ways that best supported their unique learning and teaching styles.

Benefits:

- Tailored Learning Experiences: Students receive learning materials and tasks suited to their individual learning pace and style, enhancing engagement and understanding.
- Improved Educational Outcomes: Data-driven insights allow educators to intervene early with students who may need additional support, potentially increasing overall academic success.
- Efficient Resource Allocation: Insights from data help allocate educational resources more effectively, improving the learning environment.

### 5.3.5. Discussion

The study underscores the pivotal role of DV in fostering effective HDI in education. By enhancing data accessibility, agency, and negotiability, DV tools help create more personalized and responsive learning environments. These improvements not only facilitate better educational outcomes but also promote a more engaging and adaptable educational experience for all participants.

### 5.3.6. Conclusion

This empirical evaluation demonstrates that DV, when aligned with HDI principles, substantially improves personalized learning and performance tracking in education. The findings advocate for wider adoption and continuous enhancement of DV technologies to meet the evolving needs of the education sector. Future research should further explore the integration of DV in various educational contexts to expand on these findings and optimize educational practices.

### 5.4. Overview and Benefits of the Three Studies

The application of data virtualisation in healthcare aims to improve the accessibility, actionability and negotiability of patient data, with significant implications for diagnostic processes and treatment planning. Data virtualisation tools integrate disparate sources of healthcare data, such as electronic health records, laboratory results and imaging data, and provide healthcare professionals with a single view. This unified access improves the readability and speed of data retrieval, which is critical for timely and accurate diagnosis. A randomised controlled trial demonstrated that data virtualisation reduces diagnostic time and increases accuracy, leading to improved patient care outcomes and operational efficiency in healthcare facilities.

In the financial sector, data virtualisation is improving the detection and management of financial fraud by integrating transaction data across disparate banking systems in real time. This integration provides financial analysts with the tools to perform dynamic analysis and proactively respond to potential fraud. Customisation of analytical models and data queries, aligned with specific risk management strategies, further empowers analysts and enhances the security of financial transactions. A controlled experiment within a multinational bank showed that data virtualisation tools outperformed traditional data systems in fraud detection, leading to faster responses and minimising financial losses.

Data virtualisation in education facilitates the personalisation of learning experiences and increases the effectiveness of performance tracking. By providing easy access to integrated data from learning management systems, student forums and assessment tools, educators can respond more effectively to individual student needs. Data virtualisation enables educators to tailor teaching methods and content based on comprehensive, real-time insights into student performance. Comparative analysis of schools using data virtualisation versus traditional systems has shown that classrooms enhanced with data virtualisation technologies experience higher levels of student engagement and improved academic performance.

These use cases demonstrate how DV can transform traditional practices into dynamic, efficient, and personalized processes across various sectors. Each scenario highlights the operational efficiencies and enhanced decision-making capabilities brought about by integrating diverse data sources in real time.

By focusing on these HDI principles in each use case, the use cases will not only discuss the technical benefits of DV but also emphasize its role in fostering more intuitive, ethical, and user-centric data environments. This approach highlights the transformative potential of HDI in enhancing the interaction between humans and data across various domains.

## 6. Toward an HDI Framework: Insights from Data Virtualization Use Cases

The exploration of DV across healthcare, finance, and education sectors not only underscores its capacity to enhance operational efficiencies but also highlights the critical importance of a human-centric approach in the management and interaction with data. This realization steers our study toward the consideration of developing a comprehensive Human-Data Interaction (HDI) framework, tailored to support and enhance the principles of data legibility, agency, and negotiability.

### 6.1. Rationale for an HDI Framework

Through the detailed use cases presented, it becomes evident that while DV provides substantial technical benefits, its full potential is realized only when users are able to interact with, understand, and control their data effectively. An HDI framework would provide structured guidance on integrating these user-centric considerations throughout the design and implementation of DV solutions.

- **Enhancing Data Legibility:** The framework would offer guidelines on designing intuitive visualizations and interfaces that make complex data comprehensible for all users, irrespective of their technical background. This is crucial in sectors like healthcare, where data comprehensibility can directly influence patient care outcomes.
- **Empowering with Data Agency:** Guidelines on enabling user interaction with data, allowing individuals to view, modify, and control their personal or professional data. In financial sectors, this can translate to tools that let users actively monitor and control their transaction data to prevent fraud.
- **Facilitating Data Negotiability:** Recommendations for fostering transparent communication between data users and providers. This involves creating mechanisms for users to negotiate what data is collected, how it is used, and under what circumstances, particularly in educational contexts where data privacy is paramount.

### 6.2. Implications for Implementation

Developing an HDI framework based on these principles will guide organizations in creating more user-friendly and ethically responsible DV systems. It will encourage a shift from merely using technology to enhance system efficiencies to leveraging technology to empower users and enhance their interaction with data systems. This shift is fundamental in building trust and promoting broader adoption of data-driven technologies.

The proposed HDI framework will be iterative, with ongoing refinements based on user feedback and technological advancements. Future research will explore its application in additional sectors and evaluate its impact on user satisfaction and system efficacy. This ongoing study will help ensure that the framework remains relevant and effective in promoting ethical, understandable, and user-managed data interactions.

## 7. Research Challenges

Data virtualization (DV), alongside Human-Data Interaction (HDI), faces several challenges that are central to advancing the field. These challenges range from interface design to the management of data privacy and ethics.

One significant challenge is developing interfaces that are not only powerful but also accessible to users with varying levels of expertise. Creating user-friendly interfaces that can handle complex data integrations is essential for making DV tools more approachable and effective across different user demographics.

Ensuring the accuracy and consistency of data across multiple sources remains a formidable challenge. As highlighted by [21], managing data quality in virtualized environments is critical, especially when data originates from diverse and sometimes unreliable sources.

Data privacy and security are particularly crucial in sensitive sectors such as healthcare and finance. The work of [32] on financial fraud detection underscores the ongoing concerns regarding the protection of personal and financial information. Building secure and compliant DV systems is paramount to maintaining trust and integrity.

The role of emerging technologies, including artificial intelligence and machine learning, is becoming increasingly significant in DV. According to [28], these technologies are expected to enhance the analysis and interpretation of large data sets, offering more sophisticated data management solutions.

Developing resilient DV frameworks that address both privacy and security is crucial. Research, as proposed by [31] in the context of education, should also explore the application of DV in new sectors. This exploration could provide deeper insights into the technology's capabilities and limitations, paving the way for innovative applications and systems.

Advancements in DV have the potential to significantly improve HDI by making data systems more intuitive and user-centric [13]. The integration of DV with AI and other emerging technologies offers a fertile ground for innovation, promising to address current challenges and reshape the landscape of data interaction.

As DV technologies evolve, so too must our approaches to data ownership, consent, and ethical use. Addressing these ethical considerations is crucial for advancing HDI technologies in a manner that respects user privacy and promotes trust.

## 8. Conclusions and Future Works

This article has critically examined the transformative role of DV within Human-Data Interaction (HDI), demonstrating its potential to enhance data accessibility, understanding, and management across various domains including healthcare, finance, and education. Notable studies by Raghupathi [28] and Borner [31] underscore DV's contribution to improving decision-making and operational efficiency.

Moreover, advancements in data visualization and analysis techniques have significantly enhanced the interaction between users and data systems, as highlighted by Heer [13]. This progress not only improves user experience but also propels forward the research and practical applications in the HDI domain. DV simplifies complex data interactions, allowing users to engage more effectively with the information without technical barriers, thereby fostering a deeper understanding and robust privacy practices.

As the HDI field evolves, it faces several challenges, notably in maintaining data quality and ensuring the security of sensitive information, especially in critical sectors. The ethical management of data, emphasizing ownership and consent, remains a cornerstone of ongoing and future developments [14]. Addressing these issues is essential for advancing HDI technologies in an ethical and user-focused manner.

Future research directions should include longitudinal user studies to observe how DV influences organizational workflows, decision-making cultures, and data literacy over time. These insights will be crucial in understanding the long-term impacts of DV.

Additionally, integrating Human-In-The-Loop (HITL) methodologies can further enhance the quality and relevance of data systems by incorporating user feedback directly into the DV process. This approach will be particularly valuable in sectors like social media, audio processing, and sensor data management, where new patterns and insights await discovery.

Looking ahead, the establishment of ethical frameworks and guidelines for DV is imperative. These frameworks should focus on fairness, explainability, and respect for user privacy, ensuring that DV practices adhere to the highest ethical standards.

In conclusion, DV stands as a significant advancement in HDI, offering innovative ways to access, understand, and manage data more effectively. As this technology continues to develop, embracing these challenges and opportunities will be crucial for enhancing the efficacy and experience of HDI across different contexts.

**Author Contributions:** All authors contributed to the preparation and analysis of the article, as well as the drafting of the manuscript.

**Funding:** This work has received partial support from the National Project granted by the Ministry of Science, Innovation and Universities, Spain, under Grant PID2022-140974OB-I00, and from the Regional Government (JCCM) and the European Regional Development Funds (ERDF) through the INTECRA Project under Grant SBPLY/21/180501/000056.



**Conflicts of Interest:** The authors state that they have no financial interests or personal relationships that could have influenced the work presented in this article.

## References

1. Mortier, R.; Haddadi, H.; Henderson, T.; McAuley, D.; Crowcroft, J. Human-Data Interaction: The Human Face of the Data-Driven Society. *SSRN Electronic Journal* **2014**. doi:10.2139/ssrn.2508051.
2. Stonebraker, M.; Cetintemel, U. "One size fits all": an idea whose time has come and gone. 21st International Conference on Data Engineering (ICDE'05), 2005, pp. 2–11. doi:10.1109/ICDE.2005.1.
3. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management* **2015**, 35, 137–144. doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007.
4. Tene, O.; Polonetsky, J. Big Data for All: Privacy and User Control in the Age of Analytics. *Northwestern Journal of Technology and Intellectual Property* **2012**, 11, 239.
5. Mittelstadt, B.; Floridi, L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. *Science and engineering ethics* **2015**. doi:10.1007/s11948-015-9652-2.
6. Nissenbaum, H. Privacy as contextual integrity. *Wash. L. Rev.* **2004**, 79, 119.
7. Bowen, G.M.; Bartley, A. Helping students make sense of the "real world" data mess. *Science Activities* **2020**, 57, 143–153.
8. Burrell, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society* **2016**, 3, 2053951715622512. doi:10.1177/2053951715622512.
9. Andrejevic, M. Big Data, Big Questions | The Big Data Divide. *International Journal of Communication* **2014**, 8.
10. Pasquale, F. *The black box society: The secret algorithms that control money and information*; Harvard University Press, 2015.
11. Kun, P.; Mulder, I.; Kortuem, G. Developing a Design Inquiry Method for Data Exploration. *Interaction Design and Architecture(s)* **2020**, pp. 180–206. doi:10.55612/s-5002-045-008.
12. Few, S. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*, 1st ed.; Analytics Press: Oakland, CA, USA, 2009.
13. Heer, J.; Shneiderman, B. Interactive dynamics for visual analysis. *Commun. ACM* **2012**, 55, 45–54. doi:10.1145/2133806.2133821.
14. Shilton, K. Values Levers: Building Ethics Into Design. *Science, Technology, & Human Values* **2013**, 38, 374–397. doi:10.2307/23474474.
15. Pine, M.; Sonneborn, M.; Schindler, J.; Stanek, M.; Maeda, J.L. Harnessing the power of enhanced data for healthcare quality improvement: lessons from a Minnesota Hospital Association Pilot Project. *Journal of Healthcare Management* **2012**, 57, 406–418.
16. Chen, H.; Chiang, R.H.L.; Storey, V.C. Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Q.* **2012**, 36, 1165–1188.
17. Siemens, G.; Baker, R. Learning analytics and educational data mining: Towards communication and collaboration. *ACM International Conference Proceeding Series* **2012**. doi:10.1145/2330601.2330661.
18. Batty, M.; Axhausen, K.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart cities of the future. *The European Physical Journal Special Topics* **2012**, 214, 481–518. doi:10.1140/epjst/e2012-01703-3.
19. MacKenzie, D. 'Making', 'taking' and the material political economy of algorithmic trading. *Economy and Society* **2018**, 47, 501–523, [<https://doi.org/10.1080/03085147.2018.1528076>]. doi:10.1080/03085147.2018.1528076.
20. Resnick, P.; Varian, H.R. Recommender systems. *Commun. ACM* **1997**, 40, 56–58. doi:10.1145/245108.245121.
21. Davis, J.R.; Eve, R. *Data virtualization: going beyond traditional data integration to achieve business agility*; Nine Five One Press, 2011.
22. Brynjolfsson, E.; Hitt, L.; Kim, H. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *SSRN Electronic Journal* **2011**, 1. doi:10.2139/ssrn.1819486.
23. Ferguson, M. Data Virtualization—Flexible Technology for the Agile Enterprise. *Intelligent Business Strategies* *www.sas.com* **2014**.
24. Shilton, K. Values and Ethics in Human-Computer Interaction. *Foundations and Trends® in Human-Computer Interaction* **2018**, 12, 107–171. doi:10.1561/11000000073.

25. Bellahsene, Z.; Benbernou, S.; Jaudoin, H.; Pinet, F.; Pivert, O.; Toumani, F. A flexible data integration system based on data semantics. *ACM Sigmod Record* **2010**, *39*, 11–18.
26. Catarci, T.; Santucci, G.; Da Silva, S.L. An Interactive Visual Exploration of Medical data for Evaluating Health Centres. *Journal of Research and Practice in Information Technology* **2003**, *35*, 99–119.
27. Castano, S.; De Antonellis, V.; De Capitani di Vimercati, S. Global Viewing of Heterogeneous Data Sources. *IEEE Trans. on Knowl. and Data Eng.* **2001**, *13*, 277–297. doi:10.1109/69.917566.
28. Raghupathi, W.; Raghupathi, V. Big data analytics in healthcare: Promise and potential. *Health Information Science and Systems* **2014**, *2*, 3. doi:10.1186/2047-2501-2-3.
29. Hai, R.; Quix, C.; Zhou, C. Query rewriting for heterogeneous data lakes. *Advances in Databases and Information Systems: 22nd European Conference, ADBIS 2018, Budapest, Hungary, September 2–5, 2018, Proceedings 22*. Springer, 2018, pp. 35–49.
30. Cappiello, C.; Matera, M.; Picozzi, M. A UI-Centric Approach for the End-User Development of Multidevice Mashups. *ACM Trans. Web* **2015**, *9*. doi:10.1145/2735632.
31. Börner, K.; Bueckle, A.; Ginda, M. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences* **2019**, *116*, 1857–1864.
32. Burger, M.; Smith, K.T.; Smith, L.M.; Wood, J. An Examination of Fraud Risk at Oil and Gas Companies. *Journal of Forensic and Investigative Accounting* **2022**, *14*, 74–85.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.