TOWARDS EXPLOITABLE BIG DATA IN HEALTHCARE: LESSONS FROM THE HEALTH BIG DATA PROJECT

Completed Research Paper

Fabrizio Amarilli, Dublin City University, Dublin, Ireland, fabrizio.amarilli@dcu.ie

Giada Porcelli, Fondazione Politecnico di Milano, Milan, Italy, giada.porcelli@fondazione.polimi.it

Luisa Campane, Fondazione Politecnico di Milano, Milan, Italy, luisa.campane@fondazione.polimi.it

Juliana Pesenti, Fondazione Politecnico di Milano, Milan, Italy, juliana.pesenti@fondazione.polimi.it

Paolo Locatelli, Fondazione Politecnico di Milano, Milan, Italy, paolo.locatelli@polimi.it

Abstract

Big data holds significant promise for transforming healthcare, enabling enhanced medical research, preventive care, and personalized medicine. However, realizing these benefits depends on healthcare organizations' ability to produce exploitable data—data that is high-quality, standardized, interoperable, and accessible. While prior research has identified the capabilities required for big data use, little is known about how these capabilities interact in practice to make data exploitable. This study investigates how healthcare organizations develop exploitable big data by examining the dynamic interplay among governance, human, and technical capabilities. Drawing on an ethnographic study of the Health Big Data (HBD) project, we identified four patterns of capability interaction each shaping specific properties of exploitable data. Our findings contribute to dynamic capability theory by offering a process-oriented view of capability interplay and revealing the evolving role of governance, as an initiator, mediator, and outcome of interaction. We also provide guidance for healthcare organizations to develop big data capabilities, foster stakeholder engagement, and enhance big data exploitability.

Keywords: Exploitable big data, Dynamic capabilities, Healthcare Information Systems.

1 Introduction

Big data has emerged as a transformative force in the healthcare domain, promising to unlock unprecedented insights and drive improvements in patient care and medical research (Günther et al., 2017; Grover et al., 2020). Yet, despite significant investments in digital infrastructure and analytics, many healthcare organizations continue to struggle with generating data that is truly exploitable—data that is high-quality, standardized, interoperable, and accessible (Adjerid et al., 2023; Grover et al., 2020). These properties are essential for enabling the integration and use of data across clinical, research, and organizational settings, but they remain difficult to achieve in practice.

While prior studies have identified the governance, human, and technical capabilities required for big data readiness (Akter et al., 2016; Gupta & George, 2016), most research has treated these domains as separate entities. Less attention has been paid to the interplay among capabilities, investigating how they evolve, interact, and reinforce each other during the development of exploitable big data. Moreover, governance is often conceptualized as a static enabler: a predefined framework or policy structure that guides data initiatives (Kizito, 2020; Tallon et al., 2013). This framing overlooks the possibility that governance itself may shift in response to new coordination needs and capability configurations. The present study addresses these gaps by asking the following research question: How do healthcare organizations collect and produce big data that is exploitable for patient care and research?

To explore the research question, we conducted an ethnographic study within the Health Big Data (HBD) project, a large-scale, national initiative involving over 50 research hospitals organized into four clinical networks. Through long-term engagement and participant observation, we traced how organizations build and align capabilities to support data collection, standardization, and sharing across institutional boundaries.

Our findings show that governance, human, and technical capabilities interact through four recurring patterns, each shaping different aspects of data exploitability. We also find that governance plays multiple roles: it can act as an initiator of capability development, as a mediator across domains, or as an emergent outcome of stakeholder collaboration. These results contribute a deeper process-oriented understanding of big data development in healthcare, highlighting the dynamic coordination mechanisms through which organizations generate high-value data assets. The study also offers practical insights for designing capability development strategies that are adaptive, collaborative, and aligned with healthcare's evolving digital landscape.

2 Literature review and theoretical framework

2.1 The potential of big data in healthcare

Big data refers to the vast amounts of structured and unstructured data generated from various sources, including individual-level data (e.g., wearable devices), organizational systems (e.g., enterprise applications), and external environments (e.g., social media, IoT) (Grover et al., 2020). Big data is characterized by the "three Vs", volume, velocity, and variety, which describe its scale, rapid generation, and heterogeneity (George et al., 2016). Since its emergence, the concept has attracted growing interest, with extensive research exploring its characteristics, applications, and challenges.

Across domains such as retail, manufacturing, urban planning, and finance, researchers have investigated how big data supports operational optimization, informed decision-making, and product and service innovation (Abbu & Gopalakrishna, 2022; Bedeley & Iyer, 2016; De Rijck, 2022; O'Leary, 2013; Tallon et al., 2013). The integration of big data with advanced analytics, machine learning, and artificial intelligence has further extended its strategic relevance.

Healthcare represents a promising domain of application for big data. The growing digitization of healthcare records, medical imaging, genomics, and other data-intensive technologies has generated unprecedented volumes of health-related information (Bedeley & Iyer, 2024). When harnessed effectively, big data analytics can support advances in disease prevention, clinical diagnosis, treatment optimization, and research (Eschenbrenner & Brenden, 2022).

Analysis of the literature suggests that in order to be exploitable, that is, usable for clinical or research purposes, big data must exhibit key properties, including quality, standardization, consistency, interoperability, accessibility, and transparency (Adjerid et al., 2023; Günther et al., 2017; Wang & Hajli, 2017) (Table 1). While prior research has provided in-depth insights into these essential properties and the benefits they offer, there is comparatively limited attention to the processes through which healthcare organizations develop data that achieves these characteristics.

Properties of exploitable big data	**Sources**
Quality ensures the accuracy, completeness, and reliability of the data, which is critical for making informed decisions and avoiding errors in healthcare.	Brown et. al, 2014; Wang & Hajli, 2017.
Standardization and consistency ensure uniformity across different data sources by using standardized formats, vocabularies, and coding systems. It enables seamless integration and analysis of data from multiple sources, reducing errors and inconsistencies.	Adjerid et al., 2023; Günther et al., 2017.

Interoperability provides seamless integration and sharing of data across systems, applications, and organizations by using common standards, protocols, and interfaces. It facilitates the exchange of data between different systems, enabling more effective data analysis and decision-making.	Günther et al., 2017; Wang & Hajli, 2017.
Accessibility and transparency refer to data that is easily accessible and understandable to stakeholders by using clear and concise language, visualizations, and dashboards. This property enables healthcare organizations to share data with patients, clinicians, and other stakeholders, promoting transparency and accountability.	Wang & Hajli, 2017.

Table 1. Properties of exploitable big data.

2.2 Dynamic capability as a lens of investigation

The dynamic capabilities perspective offers a valuable theoretical lens for understanding how organizations respond to complex, evolving environments. Dynamic capabilities refer to an organization's ability to build, integrate, and reconfigure internal and external competencies in response to change (Teece et al, 1997). In healthcare, where regulatory complexity, data heterogeneity, and institutional inertia are common, this perspective helps explain how organizations adapt and innovate in the use of big data.

Recent studies conceptualize big data not as a purely technological asset, but as a socio-technical system composed of interdependent capabilities (Lumor et al., 2021). These include governance capabilities, which encompass coordination mechanisms, decision-making structures, and policies that steer data development and use, human skills capabilities (e.g., domain expertise, leadership, analytical skills), and technical capabilities (e.g., infrastructure, analytics platforms, integration mechanisms) (Akter et al., 2016; Marfo et al., 2017). These capabilities are dynamic in nature. They evolve through interaction and are shaped by iterative cycles of sensing, seizing, and transforming (Gupta & George, 2016; Teece et al., 1997). Marfo et al. (2017) emphasize that higher-order dynamic capabilities emerge from combinations of lower-order tangible, intangible, and human resources.

Despite this understanding, prior research has primarily treated these capability domains independently, often analyzing them in isolation or as static prerequisites for big data success. The interplay among governance, human, and technical capabilities, that is, how they trigger, shape, and reinforce each other over time, remains underexplored, particularly in the healthcare context.

Moreover, in the context of big data initiatives, governance is frequently framed as a static enabler, composed of a set of predefined structures, policies, or compliance mechanisms, that guides successful digital health initiatives (Tallon et al., 2013; Khatri & Brown, 2010). It is often described as a stable framework established to steer technology use before new systems or data strategies are implemented out (Kizito, 2020). However, emerging research suggests that governance can also be a dynamic capability that co-evolves with organizational learning and technological maturity. For example, Mikalef et al. (2020) argue that governance frameworks in data-driven environments must adapt to increasing analytics sophistication and reflexive organizational practices. Yet, how this adaptive process unfolds in practice, particularly in healthcare organizations, remains largely underexplored.

2.3 Knowledge gap and theoretical framework

While existing studies emphasize the benefits and desirable properties of exploitable big data in healthcare, they offer limited insight into the processes through which these properties emerge in practice. Prior research has often examined governance, human skills, and technical capabilities as independent factors, without addressing how they interact dynamically over time. Furthermore, governance capabilities are frequently conceptualized as static enablers, predefined structures or compliance frameworks, that support digital health initiatives from the outset. This perspective underestimates how governance evolves through engagement with other capabilities and emerging coordination needs. This study adopts a dynamic capabilities perspective to examine how healthcare organizations generate exploitable big data through the interplay of dynamic capabilities. The theoretical framework guiding this study is presented in Figure 1.

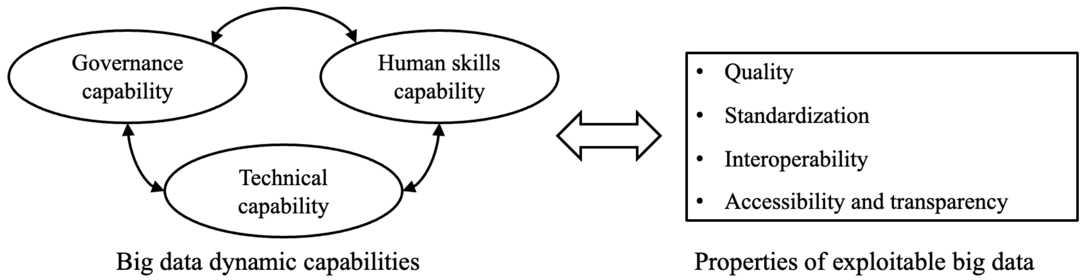

Figure 1. Theoretical framework of the study.

3 Methodology

3.1 Study context and participants: the Health Big Data project

The present study was conducted within the context of the Health Big Data project, a large-scale, multi-year initiative aimed at creating a federated healthcare big data platform. The HBD project, started in February 2020, involves 51 institutes (research hospitals) organized into four inter-organizational networks: Cardiac, Cancer, Neuroscience and Rehabilitation, and Pediatric Networks. The project also incorporates universities and the INFN (National Institute for Nuclear Physics) institute, which contribute through research expertise, academic support, and technical infrastructure for the development and integration of the big data platform. The project's objective is to enable the collection, sharing, and analysis of clinical and scientific patient data on a unified platform, while data remain distributed across the federated network of institutes. Each network and institute contribute data and expertise, and the project is overseen by an Executive Committee with working groups (WGs) for specific domains (e.g., cardiology, oncology).

The architecture of the HBD big data platform is organized into two tiers. Local platforms at each participating research institute are responsible for ensuring the extraction, integration, and interoperability of clinical and scientific data. Network platforms ensure connectivity and interoperability between institutes, as well as advanced analysis of shared data. In the initial years, the project collected and shared heterogeneous data, including omics data (e.g., genomics data, proteomics data), clinical data, clinical imaging and radiomics data, and patient-provided data. In the medium term, data from biosensors, as well as environmental, social, and economic data, will also be incorporated.

This complex, collaborative environment provides an ideal setting to study how dynamic capabilities are deployed to achieve big data exploitability. The authors of the study were embedded in the HBD project, assuming participatory roles in both coordination and operational activities. At the operational level they participated in two streams of activities related to data collection, analysis, and standardization. This provided with firsthand access to meetings, decision-making processes, and day-to-day project work, which is crucial for an in-depth understanding of the socio-technical processes at play (Baskerville & Myers, 2004).

An ethnographic research methodology (Myers, 1999) was adopted to gather rich qualitative data over the course of the project. Ethnography involves the researcher immersing themselves in the context for an extended period to capture the social and organizational dynamics in situ. This approach was well-suited because creating exploitable big data in healthcare is a complex phenomenon that unfolds over time and involves cultural and organizational change. By being on-site (physically or virtually) and participating in activities, we were able to capture nuances that might be missed by more detached methods. Furthermore, the ethnographic methodology has been previously used in several studies investigating how digital technologies are implemented, adapted, and used in complex organizational contexts (e.g., Orlikowski, 1991; Nandhakumar & Jones, 1997; Walsham, 1995).

3.2 Data collection

Our data collection spanned from February 2020 to May 2024 and followed an iterative approach typical of ethnographic research in information systems (Klein & Myers, 1999; Myers, 1999). We drew on multiple sources of evidence, both primary and secondary. Primary sources included participant observation notes from our involvement in project activities, as well as transcripts and minutes from formal meetings, including Executive Committee sessions, Network coordination meetings, and working group discussions. Over the course of the project, we also conducted focused meetings with healthcare professionals and administrators to gather insights into their experiences and perceptions of data collection and sharing. In addition, countless informal conversations helped enrich our understanding of local practices and evolving challenges. Table 2 summarizes the primary data sources, including the number and types of meetings observed and their contributions to the study.

Sources	Details	Contribution to the study
Direct participation	Participation in the activities of WG 3 on data analysis, standardization, and integration	Provided first-hand insights into data analysis and integration processes, directly contributing to the understanding of the challenges and best practices involved in big data standardization.
Meetings with clinical experts for data analysis and dataset validation	5 meetings with institutes from Cancer Network (from December 2022 to March 2023). 8 meetings with institutes from Cardiac Network (from November 2022 to April 2023) 4 meetings with experts from institutes of the Cancer Network (from Sept. 2023 to May 2024).	Enabled the identification of domain-specific data collection practices and validation of dataset standards across clinical contexts. Provided insights into the challenges of integrating clinical data across different networks, highlighting variability in data practices. Supported the theorizing process by providing expert validation of emerging patterns
WG meetings on data analysis and integration	2 meetings (October 2022, October 2023).	Contributed to the understanding of technical and analytical capabilities needed for data integration, including discussions on interoperability standards.
WG coordination meetings	31 monthly meetings involving representatives from all WGs (from September 2021 to May 2024).	Provided an overarching view of project progress, challenges, and coordination efforts, supporting the alignment of activities across different working groups. Facilitated theorizing of patterns by synthesizing input from multiple WGs.
Network coordination meetings	12 network coordination meetings (every 6 months, from October 2022 to October 2023).	Facilitated cross-network knowledge sharing, helping to harmonize practices across different healthcare domains and identify areas for further collaboration. Helped theorizing and validating the interactions of dynamic capabilities.
Overall project coordination meetings	2 meetings (March 2022, June 2023).	Provided strategic guidance for the project, ensuring alignment with overall objectives and establishing priorities for ongoing data collection and integration.

Table 2. Primary data sources in the study.

Secondary sources of data included project documentation such as periodic progress reports, presentation slides and supporting documents from meetings, technical documents outlining the big data platform architecture and data models, standard data dictionaries/models adopted, such as HL7 FHIR (Health Level 7 - Fast Healthcare Interoperability Resources) or OMOP (Observational Medical Outcomes Partnership) specifications, and public communication materials about the project. These documents provided contextual background and details (e.g., data standards chosen, governance policies) that complemented the observational data.

3.3 Data analysis

Our data analysis proceeded iteratively in four steps, enabling us to derive patterns and theoretical insights from the extensive qualitative data. We followed an interpretive approach consistent with ethnographic analysis principles (Klein & Myers, 1999), moving back and forth between data and theory.

Step 1: Coding. We developed a hierarchical coding scheme to systematically organize and analyze the collected data (field notes, meeting transcripts, and project documents). The coding scheme was structured around three primary categories (Level 1): Dynamic capabilities, Capabilities interactions, and Impact on properties of exploitable big data. Within each category, we identified multiple sub-categories (Level 2) and detailed codes (Level 3) to capture nuances of big data implementation. In this coding hierarchy, Level 1 represents broad categories, Level 2 represents sub-categories within each category, and Level 3 lists the specific codes or example concepts used to tag the data. Table 3 presents the hierarchical coding scheme.

In developing the coding scheme, we combined deductive and inductive approaches: we began with broad categories informed by the big data and dynamic capabilities' literature (e.g., looking for governance, human, and technological elements), and then refined specific sub-categories and codes based on themes that emerged from the data through open coding. For example, we expected "governance" and "IT infrastructure" to be important concepts and indeed found many instances related to those, but we also allowed new codes like "stakeholder engagement" or "patient care experience" to emerge as we coded the transcripts. This iterative coding process ensured that our analysis was guided by initial theoretical sensitization and grounded in the project data.

Before coding the full dataset, the research team collaboratively coded a subset of transcripts to calibrate understanding of the codes and achieve inter-coder consistency. We refined code definitions through discussion and then applied the final coding scheme to all qualitative data. Throughout, we maintained memos to capture insights about how different codes might relate (e.g., noting when a "governance decision" code often preceded a "technical upgrade" code in meeting narratives).

Level 1	Level 2	Level 3
Governance capabilities	Governance framework	Decision structures, Policy frameworks, Compliance systems, Risk management
	Process management	Standard procedures, Quality controls, Documentation practices, Change protocols
	Resource organization	Budget allocation, Staff management, Asset deployment, Training programs
Human skills capabilities	Leadership skills	Strategic vision, Decision-making ability, Resource commitment, Change management
	Technical skills	IT expertise, Data analysis abilities, System integration knowledge, Programming capabilities
	Clinical skills	Medical knowledge, Research expertise, Patient care experience, Domain understanding
	Collaborative skills	Cross-functional communication, Team coordination, Knowledge sharing, Stakeholder engagement
	Coordination mechanisms	Inter-department liaison, External partnerships, Network management, Communication channels
Technical capabilities	IT Infrastructure sophistication	Data storage systems, Computing power, Network capabilities, Security infrastructure
	Data management capabilities	Data collection mechanisms, Data integration tools, Data quality control systems, Data validation tools
	Analytics capabilities	Analytics tools availability, Data visualization systems, Statistical analysis software, ML capabilities
	Interoperability infrastructure	Data exchange protocols, API frameworks, Integration mechanisms, Standard compliance tools

Influence of dynamic capabilities	Cross-capability influences	Driver, Enabler, Trigger, Pre-condition, Mediator, Outcome, Catalyst. E.g., "governance decisions drive technical implementation"
Impact on big data exploitability	Quality	Error rates, Validation results, Completeness checks, Consistency metrics
	Standardization and consistency	Data structure, Coding systems, Metadata formats, Documentation standards
	Interoperability	Data exchange protocols, Integration mechanisms, API frameworks, FHIR compliance
	Accessibility and transparency	Data availability, Access permissions, Data sharing policies, Stakeholder visibility

Table 3. Hierarchical coding scheme of the data collected.

Step 2: Pattern identification. Through chronological analysis of the coded data across multiple sources, we identified recurring sequences in how capabilities were activated and interacted. We mapped these sequences against their impacts on big data exploitability properties to uncover the relationships driving successful implementation. For example, we analyzed interactions between governance and technical capabilities through specific instances such as the note from a WG meeting: "*After the Executive Committee established new data sharing policies, the IT department upgraded their systems to include FHIR-compliant interfaces*." The note was coded including governance capabilities (Governance framework/Policy framework) and technical capabilities (Interoperability infrastructure/Standard compliance tools). Such coding allowed us to capture the relationships between governance decisions and technical actions, ultimately enhancing data interoperability.

Step 3: Theory development. We abstracted the empirical patterns into higher-level theoretical constructs to explain the processes enabling successful big data implementation. Our analysis revealed sequential patterns in capability activation and their corresponding impacts on exploitable big data properties. For example, we tracked the sequential pattern of governance decisions influencing leadership commitment and subsequent technical implementation across multiple institutes. For example, the sequence was particularly evident at the Cardiology Care Institute A (CCI-A), where we observed a clear progression over an eighteen-month period. The process began with the initial governance decision to participate in HBD during the first month. This was followed by a demonstration of leadership commitment through resource allocation in the following months. The institute then conducted a comprehensive technical infrastructure assessment, which culminated with the implementation of system upgrades, demonstrating how governance decisions cascaded through organizational layers to drive technical implementation.

Step 4: Validation. To ensure the robustness of our findings, we employed multiple validation strategies that complemented each other. First, we conducted cross-comparisons within our dataset by examining the identified patterns across different institutes and networks in the project. Second, we carried out member checking and stakeholder validation sessions. We presented our interim findings and pattern interpretations to project participants, both at operational working group level and at the coordinating committee level, to assess whether these resonated with their lived experiences. For example, after identifying the governance-to-technology sequence, we discussed it with project coordinators who confirmed seeing the same cause-effect relationship and provided additional examples, strengthening our confidence in that pattern. Third, we maintained reflexivity given our dual role as participants and researchers. We organized internal peer reviews within the research team and assigned one researcher the role of semi-"outsider" to critically assess and challenge interpretations. This helped maintain analytical distance and mitigate potential biases.

Through these methods, we derived a set of findings about how healthcare organizations develop exploitable big data, focusing on the interplay among dynamic capabilities, and revealed how governance assumes multiple roles across different patterns. Findings are illustrated in the following section. A detailed overview of how the methodological steps informed the identification and theorizing of the capability interplay patterns is provided in Appendix A. This includes the coding structure, chronological analysis, and validation traces that underpin each pattern presented in the results.

4 Results

Our analysis of the HBD project reveals that the creation of exploitable big data in healthcare is driven by the interplay of three types of dynamic capabilities: governance, human skills, and technical capabilities. These capabilities do not operate in isolation but interact in complex and systematic ways. We identified four recurrent patterns of capability interplay, each representing a specific sequence of interaction. As summarized in Table 4, these patterns influence key properties of exploitable big data, including quality, standardization, interoperability, accessibility, and transparency.

Pattern	**Capability interplay**	**Pattern theorizing**	**Big data properties impacted**
1. From governance to technology readiness	Governance → Human → Technical	Governance structures create the conditions for leadership action and guide the deployment of technical capabilities.	Quality, Accessibility and Transparency, Standardization
2. From skills complementarity to interoperability	Human → Governance → Technical	Multidisciplinary collaboration reveals domain-specific requirements, prompting coordinated technical integration across domains.	Interoperability, Standardization
3. From stakeholder commitment to standardization	Human → Technical → Governance	Stakeholder commitment activates technical analysis, which identifies gaps and drives the development of standardization guidelines.	Quality, Standardization, Interoperability
4. From data sharing to stakeholder engagement	Technical → Human → Governance	Demonstrated value from integrated data sharing increases stakeholder participation and leads to revised data collection procedures.	Quality, Accessibility and Transparency, Interoperability,

Table 4. Summary of capability interplay patterns and contribution to exploitable big data.

In the following sections, we provide an exploration of each identified pattern, illustrating how the chains of capability interactions facilitate the development of exploitable big data within healthcare organizations.

4.1 From governance to technological readiness

The first pattern that we observed shows how governance capabilities initiate and structure the deployment of technical capabilities by creating a shared framework and allocating legitimacy for action. Specifically, we observed that formal governance mechanisms, such as steering committees, standardized data policies, and decision-making forums, enabled senior leadership to align priorities, commit resources, and mobilize technical teams. This organizational alignment created the conditions necessary for upgrading infrastructure, integrating systems, and enhancing data readiness. This pattern directly impacted the properties of exploitable big data by fostering standardization through policy enforcement, improving interoperability via aligned infrastructure, and enhancing accessibility through integrated and sharable data repositories.

In this pattern, governance capabilities served three interconnected roles. First, they provided direction by defining strategic goals and data-sharing protocols. Second, they enabled leadership commitment, allowing senior actors to justify investments in technology and talent. Third, they structured the technical implementation process, ensuring that infrastructure upgrades aligned with project-wide standards and interoperability goals.

In the HBD project, 15 institutes from different networks underwent an evaluation process based on a maturity model, which allowed for the assessment of the institutes' data collection practices and the sophistication of their data storage systems. Following this evaluation, an invitation was issued to institutes to participate in the creation of a local repository that would be integrated with the network platform of the HBD big data architecture. Participation in this phase was voluntary.

The Cardiology Care Institute B (CCI-B), part of the Cardiac Network, recognized their technological readiness and the strong commitment of top management within their organization. one coordination meeting transcript documented: "*The Executive Committee established clear data sharing protocols, which led to the IT department receiving additional budget allocation for upgrading their storage infrastructure.*" As a result, the institute began the pilot phase of implementing the big data initiative. CCI-B successfully established the data repository, adhering to standard protocols for data storage, and integrated their repository into the big data architecture of the Cardiac Network. This repository became accessible to other institutes within the same network and across different networks.

By aligning governance, leadership commitment, and technological readiness, this first chain of interplay increases the likelihood of successful implementation, ensuring that the resulting big data repository adheres to standardized protocols and integrates seamlessly with the wider network architecture. This integration enhances data accessibility, enabling data sharing and collaboration among organizations within the same network and across different networks.

4.2 From skills complementarity to data interoperability

The second pattern of interplay originates from the human skills capability of involving multidisciplinary stakeholders from different clinical and research organizations and various healthcare domains. The diversity of expertise across these groups surfaced domain-specific factors that could hinder the interoperability of big data. This recognition led to the activation of governance capabilities aimed at coordinating and aligning data practices across stakeholders at local and international levels. Through this engagement, organizations began to harmonize data collection and processing approaches, which in turn strengthened the technical capability to produce standardized and interoperable data.

As stakeholders from clinical, IT, data management, and research domains collaborated in the HBD project, they collectively recognized that the context- and domain-specific nuances of healthcare data could hinder the broad applicability and scalability of big data efforts. For instance, applying the data collection guidelines developed in the Cancer Network to other networks revealed the unique characteristics of data practices in oncology, cardiology, and neurology. Clinical data inclusion criteria for FHIR and OMOP standards, for example, were not equally applicable across all domains. This realization prompted project leadership to activate a governance capability, the establishment of a structured procedure to involve and coordinate with external actors and initiatives at local and international levels, with the goal of achieving standard compatibility by design.

Healthcare organizations were invited to participate in regular plenary coordination meetings organized by the HBD Executive Committee, fostering an organizational platform for knowledge sharing and alignment. Through these coordination meetings, representatives from various healthcare domains and initiatives engaged in discussions, shared experiences, and collectively addressed challenges related to data collection, standardization, and integration. This cross-pollination of perspectives and expertise facilitated the development of consistent and scalable practices for big data initiatives across different healthcare contexts. For example, from a WG coordination meeting notes a representative of the Neurology Care Institute mentioned: "*The collaboration between clinical experts and IT specialists revealed domain-specific data requirements that led to enhanced integration protocols.*"

The interactions within the coordination platforms allowed stakeholders to align data practices across domains. This alignment supported consistent implementation of data standards across organizations and improved interoperability. As a result, the big data infrastructure developed through this pattern could accommodate and integrate data from diverse clinical settings.

In this pattern, human skills capabilities, particularly the ability to navigate complexity through cross-disciplinary collaboration, initiated a broader alignment process. Governance mechanisms served as a bridge, enabling structured coordination across domains. The outcome was an improvement in interoperability, as harmonized practices made it possible to integrate and analyze data from multiple sources. This enhanced the potential for large-scale research and increased the generalizability of findings across healthcare contexts.

4.3 From stakeholders' commitment to data standardization

The third pattern of capability interplay centers on the commitment of stakeholders across different organizational levels, including top management, data officers responsible for data sharing, and clinical and IT personnel. This human skills capability was essential for enabling broad data availability. However, data availability alone is not sufficient for generating exploitable big, data: standardized procedures were necessary to guide data collection that met quality and consistency requirements.

In the HBD project, the Oncology Care Institute A (OCI-A) from the Cancer Network demonstrated a high level of commitment at all levels, as the institute was already involved in a big data initiative focused on collecting cancer-related data. As a result, the institute shared all relevant data variables. This availability triggered the use of technical capabilities for data standardization. An in-depth mapping of OCI-A data against FHIR and OMOP standards revealed that not all collected data could be seamlessly aligned. Variables were then categorized as fully aligned, partially aligned, or not aligned, highlighting potential gaps and inconsistencies.

This analysis highlighted the need for a governance capability: the development of comprehensive guidelines to align data collection practices with the existing standards. Recommendations and guidelines were proposed to extend or refine data collection procedures and ensure stronger adherence to FHIR and OMOP specifications. These guidelines were then shared and discussed with stakeholders, fostering a collaborative validation process. This governance capability facilitated the refinement of data collection practices, directly impacting the properties of exploitable big data. For example, from a working group meeting transcript we noted: "*Clinical staff at OCI-A showed strong interest in data sharing, leading to comprehensive data variable documentation, and subsequent development of standardized collection protocols*."

A similar process unfolded at the Oncology Care Institute B (OCI-B), where initial stakeholder commitment led to organizational reflection. The institute's leadership recognized that the absence of a central data governance body was a key barrier to data interoperability. In response, senior management approved the creation of a Chief Data Officer (CDO) role to oversee digital data collection and ensure alignment with HBD architectural standards. This intervention marked a formal enhancement of governance capability, emerging directly from earlier human and technical interactions.

The impact of this human–technical–governance sequence on big data exploitability manifested in several ways. Data quality improved, as clearer and more standardized collection practices were introduced and adopted. Data consistency was enhanced through the alignment of data collection processes across units and functions. Finally, interoperability increased, as the adoption of reference standards and validated data mappings enabled more seamless integration and exchange of data across systems and networks.

This third chain of capabilities' interplay highlights how human skills capabilities, such as stakeholder commitment, can initiate a sequence of technical refinement and governance innovation drive continuous improvement by providing the initial technical input. Technical analysis revealed existing gaps, prompting the development of guidelines and governance roles to improve data quality and integration. Ultimately, this pattern contributed to the creation of exploitable big data characterized by improved quality, consistency, standardization, and interoperability.

4.4 From data sharing to stakeholders' engagement

The fourth pattern of interplay highlights the sequential relationship between technical, human skills, and governance capabilities, ultimately affecting the quality and interoperability of exploitable big data. This chain originates from the technical capability to collect and share data from different healthcare domains based on validated guidelines and through an interoperable infrastructure. This harmonized approach enabled clinicians and researchers to recognize the value of integrating data across domains and between clinical and research repositories, thereby encouraging the formal engagement of additional organizations. The expansion of participation across multiple stakeholders eventually led to the emergence of new governance practices for managing data collection and sharing.

In the HBD project, the Oncology Care Institute B within the Cancer Network served as an early adopter of the project's technical guidelines for data collection. The institute integrated its patient records with the shared research repository, creating a richer dataset accessible to clinicians and researchers. The successful demonstration of interoperability in this initial case clearly illustrated the value of data integration, particularly for research purposes. This proof of concept was actively disseminated across networks, generating interest among peer institutes.

The visibility of this success created momentum among human skills capabilities, activating stakeholder engagement across various roles and levels. Top management, clinical personnel, IT specialists, and data managers from other institutes recognized the value of the initiative and expressed interest in contributing. This resulted in increased willingness to participate and new commitments to share data, both within the Neurological Network and across other clinical domains. For example, from project documentation notes from one Network coordination meeting, in relation to OCI-B, we recorded: "*The successful implementation of data sharing at the Oncology Institute demonstrated value to other departments, leading to increased participation requests.*"

As additional institutes joined the shared data platform, they encountered variations in local data collection practices. These differences prompted reflection and collaboration among stakeholders, leading to a revision of existing data collection guidelines. New standardized procedures were co-developed to address emerging inconsistencies, drawing on the lessons learned from integrating diverse data sources.

This technology–human–governance sequence impacted on the formation of exploitable big data. Data quality improved, as standardization procedures were refined through real-world feedback from data users and contributors. Data accessibility increased, as more institutes recognized the value of participation and became able to share data effectively. Interoperability was enhanced through both technical alignment with shared standards and the harmonization of processes across networks. Collectively, these improvements fostered scalability, allowing the project to integrate increasingly diverse datasets and expand its analytical and clinical reach.

5 Discussion and conclusion

This study provides insights into how healthcare organizations can develop exploitable big data—data that is high-quality, standardized, interoperable, and accessible—by examining the dynamic interplay of governance, human skills, and technical capabilities. We focus on two underexplored areas of theory development: (1) the processual interplay among big data capabilities, and (2) the evolving role of governance within those interactions. Drawing on an in-depth analysis of the Health Big Data project, we identified four recurring patterns of capability interaction. These patterns illustrate how stakeholder engagement, data sharing, skills complementarity, and governance coordination interact over time to shape the conditions through which big data becomes usable for both patient care and research. Our findings show that governance capabilities play diverse roles across these patterns, at times initiating the process, mediating across domains, or emerging from stakeholder-driven collaboration.

5.1 Theoretical contributions

Our findings offer three key theoretical contributions to the literature on dynamic capabilities and big data in healthcare. First, this study advances a process-oriented understanding of dynamic capability interplay by identifying four distinct patterns through which capabilities interact to enable big data development. While prior research has emphasized the importance of individual capabilities for big data readiness and value creation (e.g., Akter et al., 2016; Gupta & George et al., 2016; Marfo et al., 2017), it has paid limited attention to how these capabilities evolve in relation to one another. We address this gap by demonstrating that capabilities are not isolated levers but operate in mutually reinforcing sequences, where one capability can trigger, enable, or reshape the others. This contribution extends the dynamic capabilities perspective by showing that capability development is not linear or modular, but emergent and systemic, particularly in complex domains such as healthcare (George et al., 2016; Lumor et al., 2021).

Second, we contribute to the literature by linking capability interplay to the emergence of specific properties of exploitable big data. The literature on big data in healthcare often treats properties such as quality, standardization, or interoperability as preconditions or outcomes of technological implementation (Adjerid et al., 2023; Günther et al., 2017; Wang & Hajli, 2017). Our findings challenge this view by showing that these properties are not simply designed or engineered, but rather emerge from sustained interaction among governance coordination, human skills, and technical infrastructure. For example, standardization emerges not only from adherence to technical formats but from the alignment of stakeholder practices and iterative refinement of data collection procedures. By demonstrating how these properties evolve through capability interaction, we extend the literature on big data maturity and readiness with a capability process view of data quality and interoperability (Grover et al., 2020).

Third, we offer a novel theoretical contribution on the multi-faceted and evolving role of governance in big data capability development. In much of the extant literature, governance capabilities are described as predefined structures that enable data collection or ensure compliance (Tallon et al., 2013). Our study builds on this foundation by showing that governance is not static, but a dynamic capability that evolves through practice, emerging, adapting, and responding to ongoing interactions among technical and human capabilities. Across the four patterns, we observed governance acting in multiple roles: as an initiator (e.g., setting strategic direction and allocating resources), as a mediator (e.g., coordinating across clinical, technical, and administrative domains), and as an outcome (e.g., formalizing new roles or procedures in response to emerging needs). This challenges the prevailing view of governance as a fixed enabler and instead positions it as a dynamic capability that is embedded in, and shaped by, the very processes it helps structure. In doing so, we contribute to the literature on IT and data governance (Tallon et al., 2013) by highlighting how governance evolves in response to changing coordination needs and organizational learning within large-scale initiatives focused on data generation, standardization, and use.

These contributions are particularly relevant to the healthcare domain where organizations face unique challenges in generating exploitable big data due to the sensitivity of health information, fragmented data infrastructures, and strong regulatory oversight. Our findings reflect this context by showing that capability development in healthcare requires cross-level alignment, trust-building among diverse stakeholders, and iterative governance processes that respond to emerging clinical, research, and technical needs.

5.2 Practical contributions

From a practical perspective, the findings of this study offer actionable insights for healthcare organizations seeking to generate exploitable big data for clinical and research purposes. Our results underscore the importance of adopting a holistic, integrative approach to capability development, particularly in healthcare, where organizations often face fragmented infrastructures, siloed expertise, and stringent regulatory constraints, and data serves dual purposes of enabling clinical research and supporting patients' care. In such settings, dynamic capability development cannot be reduced to linear implementation plans or isolated investments in technology or personnel.

Rather, healthcare organizations must foster environments in which technical infrastructure, governance frameworks, and human expertise co-evolve in response to emerging coordination needs. This is especially critical in settings that involve multiple institutions, heterogeneous data sources, and varying levels of digital maturity. Our study highlights that the value of big data emerges not simply from the availability of data, but from the capacity to orchestrate interactions across departments, domains, and networks.

Furthermore, our findings suggest that achieving interoperability and standardization across healthcare domains requires more than technical compatibility. It requires structured stakeholder engagement, particularly the alignment of clinical, IT, and research teams with differing priorities and vocabularies. Investing in coordination mechanisms, such as cross-domain working groups, shared validation exercises, or adaptive data collection protocols, can help bridge these divides and build shared ownership of big data efforts.

Altogether, these insights can guide healthcare organizations in designing capability development strategies that are context-sensitive, participatory, and iterative moving beyond isolated technology deployments toward integrated, stakeholder-driven processes.

5.3 Future research

While this study contributes to understanding the processes through which healthcare organizations develop exploitable big data, it also has limitations that open areas for future research. First, the study was conducted within the context of an ongoing initiative, and some findings, particularly those related to data use and long-term value realization, may not yet be fully consolidated. Second, the research was situated in a specific set of healthcare networks within a national context, involving research hospitals, which may limit the generalizability of the findings to other settings with different governance structures, regulatory frameworks, or technological maturity.

Future research could extend this work through longitudinal studies that follow the complete lifecycle of big data initiatives, from early capability development to routine data use and knowledge generation. Comparative studies across different healthcare systems or countries could shed light on how institutional and regulatory differences shape capability interplay and governance roles. Moreover, applying complementary research designs, such as in-depth case studies or quantitative surveys, would help validate and extend the process models developed here.

As digital technologies continue to evolve, future work could also examine how emerging innovations, such as artificial intelligence, blockchain, or synthetic data—artificially generated data that mimics the statistical properties and structure of real-world data and can help overcome privacy constraints—reshape the dynamic capabilities needed for exploitable big data. Exploring how these technologies interact with existing human and governance structures would further refine our understanding of capability orchestration in complex and sensitive domains such as healthcare.

6 Acknowledgments

This study has been partially supported by the Health Big Data Project (CCR-2018-23669122), funded by the Italian Ministry of Economy and Finance and coordinated by the Italian Ministry of Health and the Alliance Against Cancer Network.

The authors would like to express their sincere gratitude to all the research hospitals (IRCCS) involved, as well as to the Alliance Against Cancer, Cardiology, Neuroscience, and Paediatric (IDEA) Networks, for their valuable support.

References

Abbu, H., & Gopalakrishna, P. (2022). Digital transformation powered by big data analytics: the case of retail grocery business. Proceedings of the 55th HICSS Conference, 2022.

Adjerid, I., Angst, C. M., Devaraj, S., & Berente, N. (2023). Does Analytics Help Resolve Equivocality in the Healthcare Context? Contrasting the Effects of Analyzability and Differentiation. *Journal of the Association for Information Systems*, *24*(3), 882-911.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113-131.

Baskerville, R., & Myers, M. D. (2004). Special issue on action research in information systems: Making IS research relevant to practice: Foreword. *MIS Quarterly*, *28*(3), 329-335.

Bedeley, R., & Iyer, L. (2016). Examining the role of business analytics in healthcare value chain. Pre-ICIS SIGDSA/IFIP WG8.3 Symposium: Innovations in Data Analytics Proceedings. 16.

Brown, C., Jensen, T. B., Winkler, T. J., Aanestad, M., Currie, W., Pigni, F., & Shahar, Y. (2014). Leveraging digital innovation in healthcare: harnessing big data, cloud and mobile computing for better health. Proceedings of the 2014 ECIS Conference, Tel Aviv, Israel.

Eschenbrenner, B., & Brenden, R. (2022). Deriving value from big data analytics in healthcare: a value-focused thinking approach. *AIS Transactions on Human-Computer Interaction*, *14*(3), 289-313.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, *59*(5), 1493-1507.

Grover, V., Lindberg, A., Benbasat, I., & Lyytinen, K. (2020). The perils and promises of big data research in information systems. *Journal of the Association for Information Systems*, *21*(2), 9.

Günther, W. A., Mehrizi, M. H. R., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journals of Strategic IS*, *26*(3), 191-209

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Khatri, V., & Brown, C. V. (2010). Designing data governance. *Comm. of the ACM*, *53*(1), 148-152.

Kizito, M. (2020). Ambidextrous IT Governance Enactment in Healthcare: A comparison between the Swedish and Ugandan setting. Proceedings of the 53th HICSS Conference.

Klein, H. K., & Myers, M. D. (1999). A set of principles for conducting and evaluating interpretive field studies in information systems. *MIS Quarterly*, *23*(1), 67-93.

Lumor, T., Pulkkinen, M., Hirvonen, A., & Neittaanmäki, P. (2021). Creating the socio-technical context needed to derive benefits from big data initiatives in healthcare. *Scandinavian Journal of Information Systems*, *33*(2), 1.

Marfo, J. S., Boateng, R., & Effah, J. (2017). A Typology of Big Data Capabilities from Resources to Dynamic Capabilities. Evidence from a Ghanaian Health Insurance Firm. Proceedings of the 2017 ACIS conference.

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. Information & Management, 57(2), 103169.

Myers, M. D. (1999). Investigating information systems with ethnographic research. *Communications of the Association for Information Systems*, *2*(1), 23.

O'Leary, D. E. (2013). Exploiting big data from mobile device sensor-based apps: Challenges and benefits. *MIS Quarterly Executive*, *12*(4), 179-187

Tallon, P. P., Short, J. E., & Harkins, M. W. (2013). The Evolution of Information Governance at Intel. *MIS Quarterly Executive*, *12*(4), 189-198.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509-533.

Wang, Y., & Hajli, N. (2017). Exploring the path to big data analytics success in healthcare. *Journal of Business Research*, 70, 287-299.

Appendix A

The table summarizes how capability combinations interact in each pattern and influence big data properties, following the structure of our analytical steps.

Steps	Description	Description
Pattern 1: From governance to technological readiness		**Pattern 2: From skills complementarity to data interoperability**
Coding	First-level coding revealed frequent co-occurrence of several key codes: • Under Governance capabilities > Governance framework, we frequently coded instances of "decision structures" and "policy frameworks" • Within Human skills > Leadership skills, we regularly identified "strategic vision" and "resource commitment" codes • In Technical capabilities > Infrastructure foundation, we noted recurring codes for "hardware systems", and "data manag. systems"	First-level coding highlighted frequent co-occurrence of: • Under Human skills > Clinical skills: "domain knowledge" and "research expertise" • Within Human skills > Technical skills: "system integration" and "data analysis" • In Governance capabilities > Coordination mechanisms: "inter-department liaison" and "network management"
Pattern recognition (chronological analysis)	• Governance decisions • Leadership commitment • Technical implementation	• Multi-disciplinary team formation, and • Domain-specific knowledge sharing • Technical requirement identification • Integration solution development
	For instance, at the Cardiology Care Institute A we traced this sequence: • Initial governance decision to participate in HBD (Month 1) • Leadership commitment of resources (Month 6) • Technical infrastructure assessment (month 9) • System upgrade implementation (Month 18)	For example, in the Neurological Network, we traced the cycle: • Formation of cross-functional teams (Month 14) • Domain knowledge documentation (Month 18) • Technical specifications development (Month 24) • Integration protocol implementation (Month 31) • Cross-network validation (Month 36)
Impact on properties	• Data Accessibility - Governance protocols enabled data sharing; technical systems ensured compatibility, resulting in accessible, sharable repositories. • Standardization - Governance mandated standards, and technical implementation enforced them, producing consistently formatted data. • Quality - Governance defined quality requirements; technical controls ensured reliability and high data quality	• Interoperability - Domain expertise improved system compatibility; technical solutions enabled exchange; cross-validation ensured practical use. • Standardization - Domain input guided standard design; technical implementation ensured consistency; adoption across networks promoted uniformity.
Pattern theorizing	Governance provides the framework and legitimacy for action, enabling leadership to commit resources and set priorities. This alignment creates the conditions for effective technical implementation.	Diverse expertise contributes to achieving data interoperability across different healthcare domains.
Validation	We validated patterns' interactions through stakeholder interviews. One IT director noted: "*Having clear governance protocols made it easier to justify the technical investments and ensured our upgrades aligned with project requirements.*"	We conducted focus groups with clinical, IT, and research stakeholders involved in coordination meetings. They confirmed the consistency and applicability of standardization practices. One IT manager at NCI noted that early involvement of diverse stakeholders ensured robust, scalable data integration across domains.

Pattern 3: From stakeholders' commitment to data standardization		**Pattern 4: From data sharing to stakeholders' engagement**
Coding	First-level coding highlighted frequent co-occurrence of: • Under Human skills > Leadership skills: "commitment" and "change management" • Within Technical capabilities > Data management: "data collection tools" and "standardization systems" • In Governance capabilities > Process management: "standard procedures" and "quality controls"	First-level coding highlighted frequent co-occurrence of: • Under Technical capabilities > Integration framework: "data exchange protocols" and "interoperability tools" • Within Human skills > Collaborative skills: "stakeholder engagement" and "knowledge sharing" • In Governance capabilities > Coordination mechanisms: "network management" and "communication channels"
Pattern recognition (chronological analysis)	• Initial stakeholder commitment • Comprehensive data sharing • Technical analysis of shared data • Development of standardization guidelines • Enhanced stakeholder engagement	• Initial technical success • Value demonstration • Increased stakeholder interest • Additional data sharing • Enhanced value creation
	For example, at the Oncology Care Institute A, we traced this cycle: • Initial commitment to share cancer data (Month 10) • Complete sharing of data variables (Month 14) • FHIR/OMOP mapping analysis (Month 18) • Development of standardization guidelines (Month 24) • Increased stakeholder participation (Month 32)	For example, in the Oncology Care Institute B, we traced the cycle: • Successful pilot implementation (Month 22) • Documentation of benefits (Month 28) • New institute participation (Month 30 • Expanded data sharing (Month 36) • Network-wide value enhancement (Month 38)
Impact on properties	• Quality - Comprehensive data sharing surfaced issues; guidelines improved consistency; iterative refinement enhanced quality. • Standardization - Initial sharing exposed gaps; technical analysis guided improvements; new guidelines ensured consistency in data collection. • Interoperability FHIR/OMOP alignment boosted compatibility; standardized procedures improved exchange; integration across systems became feasible.	• Quality - Diverse data improved completeness; more cases enhanced reliability; broader participation increased validation. • Accessibility - More institutes joined; data availability expanded; network coverage improved. • Interoperability - Cross-system integration expanded; standard adoption widened; connectivity across networks improved.
Pattern theorizing	Stakeholder commitment drives improvements in data standardization through iterative technical and organizational developments.	Successful data sharing initiatives creates a positive feedback loop of increasing stakeholder engagement.
Validation	Cross-network coord. meetings, where stakeholders confirmed the harmonization of practices. A research lead at OCI-A noted: "*Bringing together perspectives from various domains allowed us to develop consistent practices that could be applied across different networks, significantly improving data integration and interoperability*."	Case studies document on successful data integration initiatives. A clinician at OCI-B noted: "*Seeing added value from shared patient records with research repository highlighted how crucial broad participation and shared data are for enhancing quality and interoperability.*"