Pelin Özüberk

M.Sc. Thesis

AGU 2025

THE IMPACT OF BIG DATA AND DATA ANALYTICS ON ARCHITECTURAL DESIGN PROCESSES

M.Sc. THESIS
SUBMITTED TO THE DEPARTMENT OF ARCHITECTURE
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER

By
Pelin Özüberk
July 2025

THE IMPACT OF BIG DATA AND DATA ANALYTICS ON ARCHITECTURAL DESIGN PROCESSES

A THESIS
SUBMITTED TO THE DEPARTMENT OF ARCHITECTURE
AND THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
ABDULLAH GUL UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER

By
Pelin Özüberk
July 2025

SCIENTIFIC ETHICS COMPLIANCE

I hereby declare that all information in this document has been obtained in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name-Surname: Pelin ÖZÜBERK
Signature :

REGULATORY COMPLIANCE

M.Sc. thesis titled The Impact of Big Data and Data Analytics on Architectural Design Processes has been prepared in accordance with the Thesis Writing Guidelines of the Abdullah Gül University, Graduate School of Engineering & Science.

Prepared By	Advisor
Pelin ÖZÜBERK	Prof. Burak ASİLİSKENDER
Signature	Signature

Head of the Architecture Program

Assoc. Prof Buket METİN

Signature

ACCEPTANCE AND APPROVAL

M.Sc. thesis titled The Impact of Big Data and Data Analytics on Architectural Design Processes and prepared by Pelin Özüberk has been accepted by the jury in the Architecture Graduate Program at Abdullah Gül University, Graduate School of Engineering & Science.

04 / 07 / 2025

(Thesis Defense Exam Date)

JURY:

Advisor : Prof. Burak ASİLİSKENDER ………………………………

Member : Assist. Prof. Ömer Devrim AKSOYAK ………………………………

Member : Assist. Prof. Özlem ATAK DOĞAN ………………………………

APPROVAL:

The acceptance of this M.Sc. thesis has been approved by the decision of the Abdullah Gül University, Graduate School of Engineering & Science, Executive Board dated ….. /….. / ……….. and numbered ……………..……. .

……….. /……….. / ………..

(Date)

Graduate School Dean
Prof. İrfan ALAN

ABSTRACT

THE IMPACT OF BIG DATA AND DATA ANALYTICS ON ARCHITECTURAL DESIGN PROCESSES

Pelin Özüberk
M.Sc. in Architecture
Advisor: Prof. Burak ASİLİSKENDER
July 2025

The increasing proliferation of big data and advancements in data analytics are fundamentally transforming architectural design processes. This thesis investigates the role and impact of big data within the architectural field, focusing on how vast datasets and analytical techniques are influencing traditional and digital design workflows. Through a detailed exploration of the concepts of data, information, and knowledge, the study outlines how data-driven approaches reshape the way architects conceptualize, develop, and optimize designs. The research examines the transition from traditional intuition-based design models to digital, adaptive, and performance-oriented methodologies, highlighting the critical principles and techniques of data analytics relevant to architecture. It further analyzes contemporary practices such as adaptive design, generative systems, and performance-based strategies that are enabled by big data integration. The study concludes by reflecting on the opportunities and challenges posed by big data, emphasizing the necessity for interdisciplinary collaboration, ethical considerations, and the evolution of architectural education to equip future practitioners. The work also suggests future research directions, particularly in the development of intelligent, user-responsive architectural environments.

Keywords: Big Data in Architecture, Architectural Design Process, Data-Driven Design, Architectural Knowledge, Design Process Transformation

ÖZET

BÜYÜK VERİ VE VERİ ANALİTİĞİNİN MİMARİ TASARIM SÜREÇLERİNE ETKİSİ

Pelin Özüberk
Mimarlık Anabilim Dalı Yüksek Lisans
Tez Danışmanı: Prof. Burak ASİLİSKENDER
Temmuz 2025

Büyük verinin artan yaygınlaşması ve veri analitiğindeki gelişmeler mimari tasarım süreçlerini kökten dönüştürmektedir. Bu tez, mimari alandaki büyük verinin rolünü ve etkisini araştırıyor ve geniş veri kümelerinin ve analitik tekniklerin geleneksel ve dijital tasarım iş akışlarını nasıl etkilediğine odaklanmaktadır. Veri, bilgi ve bilgi birikimi kavramlarının ayrıntılı bir incelemesi yoluyla bu tez, veri odaklı yaklaşımların mimarların tasarımları kavramsallaştırma, geliştirme ve optimize etme biçimini nasıl yeniden şekillendirdiğini ana hatlarıyla açıklamaktadır. Araştırma, geleneksel sezgiye dayalı tasarım modellerinden dijital, uyarlanabilir ve performans odaklı metodolojilere geçişi incelemekte ve mimariyle ilgili kritik veri analitiği ilkelerini ve tekniklerini vurgulamaktadır. Ayrıca, büyük veri entegrasyonunun sağladığı uyarlanabilir tasarım, üretken sistemler ve performansa dayalı stratejiler gibi çağdaş uygulamaları analiz etmektedir. Çalışma, büyük verinin ortaya koyduğu fırsatlar ve zorluklar üzerine düşünerek, disiplinler arası iş birliğinin gerekliliğini, etik hususları ve gelecekteki uygulayıcıları donatmak için mimarlık eğitiminin evrimini vurgula. Çalışma ayrıca, özellikle akıllı, kullanıcıya duyarlı mimari ortamların geliştirilmesinde gelecekteki araştırma yönlerine dair önerilerde bulunmaktadır.

Anahtar kelimeler: Mimarlıkta Büyük Veri, Mimari Tasarım Süreci, Veri Odaklı Tasarım, Mimari Bilgi, Tasarım Süreci Dönüşümü

Acknowledgements

Firstly, I would like to thank TÜBİTAK for supporting my studies through the 2210-C Yurt İçi Öncelikli Alanlar Yüksek Lisans Burs Programı. I am also grateful to my advisor, Prof. Burak Asiliskender, for always making time, for his belief in me, and his support throughout this thesis.

Additionally, I would like to thank my family, especially my spouse, Mustafa Özüberk, for his understanding and unwavering support during my master's studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ADP	Architectural Design Process
AI	Artificial Intelligence
BIM	Building Information Modelling
CAD	Computer Aided Design
DIK	Data Information Knowledge
DIKW	Data Information Knowledge Wisdom
IFC	Industry Foundation Classes
IoT	Internet of Things
VAE	Variational Autoencoders
VR	Virtual Reality

Chapter 1

Introduction

Over time, the concept of VUCA, an acronym referring to volatility, uncertainty, complexity, and ambiguity, has gained increasing relevance. Originating in military language in the late 1990s, it has since been adopted to describe the dynamics of an increasingly unstable and rapidly evolving global landscape, particularly within the business and organizational context (Lawrence, 2013). These conditions come with many changes in every aspect of life, from personal lifestyles and professional environments to academic and research settings. However, as we begin to make sense of these conditions and uncover the opportunities they present, we might realize this can help us to navigate the aspects of our work, design, or research processes more confidently.

In this contemporary era, when the concept of data is overly complex, fast, and intertwined, returning to the beginning and understanding the concepts with their first meanings and the conditions in which they emerged can help us interpret this confusion today. Rescher (2003) argues that the pursuit of knowledge is not a choice, but a necessity rooted in human nature. He emphasizes that individuals are compelled to seek understanding because their well-being depends on cognitive orientation in the world. When this orientation is lacking, the result is psychological and even physical distress. Rescher (2003) emphasizes that humans require information to navigate their environments, resolve uncertainty, and maintain a sense of consistency. In this view, the drive to make sense of things is not simply intellectual curiosity but an inherent feature of human life.

Architecture, as both a spatial and cultural practice, responds to this need by producing environments that help people feel situated, connected, and oriented. The built environment plays a critical role in shaping how individuals interpret and move through the world. Architecture does more than fulfill functional requirements, it also structures experience, anchors memory, and supports shared forms of understanding. In this way, design processes in architecture are intimately linked with the human-need to know.

Kitchin (2014b) contributes to this discourse by tracing how, across history, institutions such as governments, scientific communities, and businesses have relied on data as a tool for observation, regulation, and decision-making. He notes that while data collection was once a time-intensive and limited process, today it is extensive, automated, and capable of producing detailed, real-time accounts of the world. This shift has not only transformed the scale and speed at which information is produced but also the ways in which knowledge is constructed, applied, and circulated.

At the same time, the rise of Big Data has significantly changed how knowledge is produced and used. With its large scale, speed, and variety, Big Data goes beyond traditional ways of collecting and understanding information. It does not just reflect what is happening in the world, but also influences how decisions are made and how systems are designed. For architecture, a field that has often relied on experience, observation, and intuition, this shift creates both new possibilities and new questions. It encourages designers to think about how data can be included in the design process, and how it might support or challenge the ways we have traditionally understood and shaped the built environment.

As digital tools, computational models, and data-driven systems become more embedded in architectural workflows, these developments introduce new conditions for how spaces are conceived and evaluated. The discipline must now balance its historical knowledge base and creative traditions with the emerging challenges and potentials of working in data-rich environments. This thesis explores that intersection, examining how architecture participates in the broader shift toward data-informed practices while remaining attentive to its epistemological foundations and human-centered values.

1.1 Purpose of the Thesis

This thesis focuses on the growing use of big data and data analytics in architectural design processes. It aims to understand how data is influencing contemporary design methods and to explore what this shift means for the role of the architect. As architecture moves toward a more technologically integrated practice, the question arises: How can designers use data without losing the creative and critical aspects of their work? This study does not simply aim to promote data-driven design but to critically examine its potentials and limitations in real-world applications.

The main goal is to investigate how big data is being incorporated into the stages of architectural design process. This includes an exploration of the kinds of data that are being used, how this data is collected and interpreted, and what tools or platforms are enabling these processes. Additionally, the thesis addresses how the presence of data may influence the relationship between the designer, the space, and the user. It questions whether data is helping to create more inclusive and efficient environments or if it risks oversimplifying complex spatial experiences.

To guide this investigation, the following research questions have been formulated:

Table 1.1 Research Structure (created by author)

Research Questions	Objectives	Assumptions
How is big data currently used in architectural design?	To identify key applications of big data in design workflows.	Big data supports decision-making by adding more layers of spatial insight.
What challenges arise in integrating data analytics in design?	To explore both technical and conceptual barriers.	Architects need better training, tools, and frameworks to use data effectively.
What potential does data hold for shaping future design logic?	To reflect on new directions in data-driven architecture.	The role of data will become increasingly central in architectural thinking.
How does data influence the ways architects represent and communicate design ideas?	To analyze how data-driven workflows affect design language, tools, and drawing conventions.	Data is not only an input but also a representational logic in design processes
In what ways does data reshape architectural knowledge and authorship?	To explore the epistemological and disciplinary implications of integrating data in design.	The role of the architect is shifting from form-maker to data interpreter or strategist.

(Table 1.1) presents the core structure of the research and shows how the study moves from current practices to broader implications. The questions were shaped by a combination of preliminary readings and the observation that while many architectural offices now use software tools that include data features, the integration of these tools is not always deep or strategic. In some cases, data is used in surface-level ways, such as visualizations or sustainability checklists, without influencing the core spatial concept.

This research positions itself at the intersection of architectural design theory, digital representation, and data-informed practice. While much of the existing literature on big data in the built environment focuses on smart city infrastructures, technical optimization, or post-occupancy analysis, this thesis contributes to a smaller but growing body of work that explores the conceptual and epistemological implications of data in the design process itself. By focusing on how data is framed, visualized, and translated into spatial logic, the study aims to offer a design-centered perspective that is often overlooked in more technically driven or policy-focused discussions. In doing so, it aligns with emerging architectural scholarship that critically engages with computation not only as a tool but also as a design ideology and representational shift.

What this thesis seeks to uncover is whether data can be more than a supporting tool: Can it be a partner in creative thinking? Can it expand the ways we define, understand, and experience architecture? In doing so, the research also touches on themes of authorship, agency, and the future identity of the profession.

1.2 Method of the Thesis

The methodology of this thesis can be classified as qualitative and interpretive, with elements of conceptual/theoretical research and visual analysis. To explore the questions outlined above, this thesis adopts a two-phase methodology: (1) a conceptual and theoretical framework built through literature, and (2) a visual and representational analysis of how data appears within architectural design processes and communication tools.

In the first phase, a wide range of academic literature is reviewed to map how data and computation have entered architectural discourse. This includes works that discuss data-driven design, computational workflows in architecture, and the evolving relationship between digital tools and architectural thinking. The literature covers both the technical side, such as how data is generated and analyzed, and the conceptual side, including debates about creativity, authorship, and ethics in data use. The aim of this phase is to build a foundation for understanding not only what big data is, but also how it has been framed within the design field. This part of the research also identifies gaps in the literature, especially the need for more design-focused critiques of data systems.

The second phase centers on examining how data is represented and interpreted in architectural design workflows. Rather than selecting full-scale case studies of completed buildings or masterplans, the thesis focuses on design processes, particularly on visual and narrative tools that show how data is embedded in decision-making. These include diagrams, project illustrations, conceptual workflows. This phase also draws on mostly published academic articles that explicitly discuss how data has been used within design logic or project development. These sources are used not only for factual reference but also to understand how practices frame and communicate the role of data in design. The aim is to read these materials critically and to understand how data influences the conceptual and communicative stages of architectural work.

This analysis is guided by the following considerations:

- What kinds of data are being used or visualized? (such as environmental data, behavioral patterns, program distribution)
- How is this data being processed and interpreted in the design narrative?
- What visual techniques (such as layering, coding, or mapping) are used to embed data into drawings or diagrams?
- What role does data play in shaping spatial strategies, design arguments, or conceptual logics?
- How do these practices reflect broader shifts in the role of the architect, as a designer, analyst, or mediator?

Importantly, this thesis does not produce its own dataset or conduct original data modeling. Instead, it focuses on interpretation and critique, looking at how data has been used and visualized by others, and what these practices reveal about the evolving relationship between data and design.

The thesis is organized into five main chapters, each building upon the previous to form a coherent progression from context to critical reflection. Chapter 1 sets the foundation by introducing the research background, framing the key questions, and outlining the methodology. Chapter 2 establishes the conceptual ground by exploring the notions of data, information, and knowledge, and how these have been interpreted and applied within architectural discourse. Building on this foundation, Chapter 3 focuses specifically on the idea of big data, examining its entry into architectural thinking and how it has influenced disciplinary boundaries, roles, and methods. Chapter 4 brings these discussions into a more applied context by analyzing examples of data-influenced design processes, reflecting on how these practices illustrate broader shifts in architectural workflows, not only in how data is utilized, but also in how the design process itself is being redefined through digital and computational thinking. Finally, Chapter 5 concludes the thesis by summarizing the key findings, reflecting on the implications of data integration in architectural design, and suggesting potential directions for future research and practice.

This structure allows the thesis to move from general context to specific conditions and finally to critical reflection. It supports a narrative flow that mirrors the evolving design logic many architects follow today, starting from foundational ideas, moving into digital tools and technologies, and finally looking at practical illustrations of design development. The reason the thesis begins with data, information, and knowledge (before addressing big data directly) is to establish a layered understanding of how architects engage with meaning, scale, and interpretation in data use. This sequence helps clarify that big data is not just a technical tool, but part of a broader cultural and epistemological shift that reshapes how architects think, communicate, and design.

1.3 Scope of the Thesis

The scope of this thesis is shaped by both thematic and practical boundaries. Thematically, it focuses on the intersection of architectural design and data analytics, with an emphasis on how data influences spatial decisions and design decisions. While the primary interest lies in projects that use data during the early and middle stages of the design process, the study also considers aspects of later stages, such as technical documentation and delivery phases, when they help to clarify the overall role of data across the full arc of the design workflow. However, these later phases are not examined in technical depth but rather as part of a broader understanding of how data flows through a project timeline.

In terms of content, this thesis covers discussions across multiple design scales and fields, primarily architectural and urban design but also includes product and industrial design, computational and generative design, as well as smart environments and performance-driven systems. It excludes interior-only designs or speculative proposals without explicit reference to data-driven thinking. Similarly, projects that are primarily engineering-oriented with minimal architectural or design engagement are outside the scope. The main focus is on design processes where data actively informs decisions on form, function, performance, spatial layout, or user interaction.

Although no strict time-frame is applied to the selection of projects, most of the examples reflect developments from the last decade, a period marked by the increased accessibility of digital tools, performance software, and urban data platforms. The recency of these projects allows the thesis to engage with contemporary practices that are still evolving and actively shaping architectural discourse.

Geographically, the thesis does not limit itself to a specific region but gives priority to projects that emerge from contexts with digital infrastructure capable of supporting data-driven design. These typically might include examples from Europe, North America, and East Asia, though projects from other regions are not excluded. The selection criteria are based more on conceptual relevance and design clarity than geographic distribution.

Finally, the thesis does not aim to address the full technical complexity of data science or software development. It avoids detailed discussions of algorithm design, programming, or system engineering. Instead, it centers on how designers engage with data as a conceptual and visual input, and how this engagement reconfigures design thinking and representation practices.

By setting these boundaries, the thesis remains focused and manageable while still addressing a broad and timely topic. It offers insight into how architects can work with data today and how they might rethink their role in the near future.

Chapter 2

Data, Information and Knowledge in Architectural Discipline

In a time where digital transformation and data-centric methodologies are rapidly reshaping architectural practice, it holds a critical value to begin this thesis by revisiting the foundational concepts of data, information, and knowledge. These are not only technical terms but also epistemological foundations that define how we think, design, and evaluate the built environment. To understand how architectural design is being reshaped by digital tools, big data, and artificial intelligence, it first becomes necessary to pause and ask what these terms actually mean within the context of architecture.

This chapter begins the thesis because understanding what data is, and how it differs from information and knowledge, is essential to grasp the significance of data-driven processes in architecture. Architectural design has always relied on a variety of knowledge types: spatial, material, cultural, and experiential. However, the rise of digital infrastructures and computational thinking demands that we re-express this knowledge in forms that machines can interpret, analyze, and even generate. As such, the architectural discipline faces both opportunities and challenges in redefining its epistemological foundations.

This chapter explores how data, information, and knowledge have been conceptualized historically and how they function today in architectural design. It discusses how these concepts are interpreted across disciplines and introduces frameworks such as the DKIW (Data, Knowledge, Information, Wisdom) model to unpack their interdependencies. The discussion pays particular attention to how architectural knowledge is created, shared, and transformed in the context of big data and digital workflows, and what kinds of data engage in the design process, from intuitive sketches to environmental simulations.

By establishing this foundation, the chapter prepares the foundation to better understand the transitions explored in later chapters, especially those dealing with big data, artificial intelligence, and the transformation of the architectural design process. In doing so, it positions architecture not only as a field of form-making but also as a domain of knowledge production, responsive to technological, cultural, and epistemic shifts.

2.1 Understanding Data and Information

There are various definitions of data. This means that the scope can also change across different domains. A brief review of the existing literature reveals a contradiction that is easily noticeable in modern discussions: even with the growing popularity of the term 'data' in academic research and various fields, there is no single, universally accepted definition for it (Avanço et al., 2021, p. 6). Merriam-Webster (2025) defines the etymological origin of the term data, tracing it back to the Latin term datum, which is defined as ‘something given or admitted especially as a basis for reasoning or inference’. Furthermore, its earliest recorded usage can be traced to the mid-17th century.

The first meaning of data occurs as facts from sacred writings or a ground basis for arguments, where there was no room for interpretation. Throughout the end of the century, it has been established and evolved as outcomes of an ‘experiment, experience or collections’ (Rosenberg, 2013, p. 33). This definition gives the impression of its role as a result rather than an assumption. That point of view is critical where we think, search and define the meaning of it. It reinforces the definition of ‘something given’ as it implies the relational link with a result of an act, as Rosenberg (2013) emphasizes ‘Data means that which is given before argument. Consequently, the meaning of data must always shift with argumentative strategy and context, and with the history of both’. This perspective highlights that data is not fixed in meaning but instead gains significance through the lens of interpretation and the context in which it is used.

In present-day discussions, data is acknowledged as a key foundation for argumentation; however, our primary notion of data as information in numerical representation is based on the advancements from the late eighteenth century (Rosenberg, 2013, p. 33). He also offers an alternative interpretation as follows: ‘It may be that the data we collect and transmit has no relation to truth or reality whatsoever beyond the reality that data helps us to construct. This fact is essential to our current usage’ (Rosenberg, 2013, p. 37). Without undermining the importance of data itself, this statement highlights the status of it is a building block for both mental and any other external processes. From there, it leads to the way of knowledge and information. But also, it holds an important place on its own without interrelations to these terms. Data is not fundamentally equivalent to truth, nor is it assigned the role of carrying absolute validity. Rather, it serves as a representation of an outcome, independent of any obligation to assess the credibility of its source. The responsibility to interpret, validate, and critically engage with data lies with us as scholars and researchers.

This changes the perspective of the term in a context. It opens a discourse that the perceptions we may accept do not necessarily have to be distant from bias. It comes from our interpretations, this also questions the term of ‘something given’ as if a result of an observation comes from interrelated inquiries, would it be the justice to describe it as something given just like appearing from nowhere? Though it may fit on some occasions where the data comes from or where we can directly observe it as a fact. In that situation, we may conclude the meaning, and therefore the data itself may shift according to the situation or inquiry we extract it.

This reasoning shows itself in different fields of science. In the arts and humanities, it is a more sensitive and complex topic. Because the areas it contains are mostly on analytical frameworks such as semantics, rhetoric, phenomenology, and such, it can create contrasts with humanistic inquiry. Within the notion of the digital humanities, researchers argue that the meaning is more of an active notion rather than an observer-independent one, which is given perspective (Drucker, 2011, as cited in Schöch, 2013). This follows with the commentary that, on the contrary, it can be manufactured, created, and open to interpretation. Social sciences add the perspective that the notion of data does not always have to relate to an organized object. Rather than the term itself, it is more important to speak up about the collective nature of data (Bordignon & Maisonobe, 2022). This point blends nicely with the nature of human relations.

As researchers, we may attach varying meanings to the term "data," shaped by our disciplinary perspectives and conceptual frameworks. Bordignon and Maisonobe (2022) delve into this subject by examining how the term *data* are used within scholarly articles. They concluded with the idea that rather than the term itself, adjectives and verbs that define data hold greater value since the word of data is a discussion and connected with the framework where it is based. This shifts the focus from rigid definitions to a deeper understanding into what data does, why it matters, and how it functions within our disciplines. As Rosenberg (2013) explains, the emergence of modern economics and empirical science brought about new ways of constructing arguments and new expectations around facts and evidence. The evolution of such terms illustrates broader transformations in how knowledge is understood and confirmed. The significance of data lies not only in its informational content but also in its cultural, historical, and linguistic roots. Zins (2011), for instance, draws on earlier observations to attention that misuse of the term *data* often stems from ignorance of its Latin roots, *datum* meaning "the given" and *data* being its plural form. These insights call for a more nuanced understanding of data, not as a fixed matter, but as an expanding influence in contemporary conversation, a pixel in our cognitive landscape, and a key to unlocking the reasoning processes embedded in our professions.

Understanding data today requires moving beyond the notion of it as a neutral or objective reflection of the world. As Gitelman (2013) and Loukissas (2019) argue, data is always the outcome of situated processes, shaped by the contexts, technologies, and human decisions that bring them into being. They are assembled through intricate networks of sensors, infrastructures, standards, algorithms, and expert practices, all working together to produce, validate, and circulate what becomes recognized as data (Gray et al., 2018).

Rather than existing independently, data are entangled within what Crawford (2021) describes as *data assemblages*, dynamic circumstances of technical systems, social actors, and material environments. These assemblages frame what counts as data and evolve over time, mutating as technologies advance, institutions shift, and new forms of knowledge and power emerge (Amoore, 2020). These assemblages can vary from ‘forms of knowledge’ to ‘system of thought’ or ‘places’ to ‘finance’ and other different groups, as Kitchin (2014b) represents. Thinking of data in this way challenges the persistent idea that it can ever be fully raw, complete, or universal. Instead, it reinforces its partial nature and highlights how data practices actively construct the realities they claim to represent. To engage critically with contemporary data environments, it is essential to understand the complex, relational infrastructures through which data are continuously produced, interpreted, and reshaped.

Similarly with data, information also has been conceptualized in various ways depending on the perspective taken. Some view it simply as connected data, others as data enriched with meaning, as a meaningful signal extracted from noise, or as processed primary data transformed through analysis (Kitchin, 2014b). These definitions highlight the complexity of distinguishing data from information, as well as their relationship to knowledge. However, the view that treats data as the raw material for information, and information as the raw material for knowledge has been criticized. Zins (2011) argues that this linear progression is problematic; information is not simply a transitional phase between data and knowledge. He further challenges the alternative approach that treats information and knowledge as synonymous, describing this too as a conceptual fallacy. According to Zins (2011), in the subjective domain, information is perceived as a "type of knowledge," while in the objective domain, it is defined as a "set of symbols" that represent empirical knowledge. This duality illustrates that information's nature is deeply influenced by the context in which it is considered.

Building on a more formal structure, Floridi (2010) offers a data-based definition of information, emphasizing that information emerges from meaningful and well-formed data. Importantly, Floridi (2010) notes that even the absence of data, such as silence, still qualifies as data. Yet, for data to become information, it must be syntactically well-formed; in other words, it must follow to the structural rules of a given code or language (Floridi, 2010). Meaning, in this sense, depends not only on the data itself but on its alignment with predefined linguistic or symbolic structures. Floridi (2010) further elaborates that information can constitute various types of data, including primary data, secondary data, or metadata. When data is properly structured and meaningful, it becomes what he calls *semantic content*. Notably, information does not always require an intelligent agent for its creation; it can exist naturally in the environment, with data carrying meaning inherently, Floridi further completes.

Through these perspectives, it becomes clear that information is not a simple, uniform concept, but rather a layered and context-dependent phenomenon. Understanding these nuances is crucial for disciplines such as data science, architecture, and information theory, where the distinction between data, information, and knowledge constructs critical practices.

2.1.1 Data Types and Formats

Understanding data may begin with its rhetorical and ontological discussions, but there are several perspectives, as Kitchin (2014b) mentions that it can be socially constructed, carrying ideological biases or serving as a public good. The main idea is that data may never be purely objective or neutral. Many characteristics are involved on the categorization of data. These definitions lead us to the (Table 2.1).

Table 2.1 Data Kinds (based on Kitchin, 2014b)

DATA KINDS	**By Form**	• Quantitative Data • Qualitative Data
	By Structure	• Structured Data • Semi-Structured Data • Unstructured
	By Source	• Captured Data • Exhaust Data • Transient Data • Derived Data
	By Producer	• Primary Data • Secondary Data • Tertiary Data
	By Type	• Indexical Data • Attribute Data • Metadata

Definitions of kinds of data follow as (Table 2.2), and these forms of data reveal that it can be shaped and categorized by many factors. Riswadkar's (2006) educational report on data types also states that different categorizations of data are possible, and this classification is demonstrated through the Sciences and Social Sciences. In Sciences, defining factors which data can be aligned by are: 'time factor', 'location factor', 'mode of generation', 'nature of quantitative tasks', 'terms of expression' and 'mode of expression' whereas in Social Sciences it is arranged by: 'scale of measurement', 'continuity', 'reference to time', 'characteristics', 'number of characteristics' and 'origin'. Kitchin (2014b) also mentions that there are many other ways data can be seen and contextualized. Data can be framed 'technically', 'ethically', 'politically', 'spatially', or 'philosophically'. Which brings us to the conclusion that it is a very rich phenomenon, and the process of conceptualizing data involves more than what is typically recognized by the fields it has been used.

Table 2.2 Data Kinds Explanation (based on Riswadkar, 2006)

• **Quantitative Data:** Numerical and typically large in scale, describing physical traits or representing the non-physical aspects of phenomena. • **Qualitative Data:** Non-numerical, including texts, images, art, videos, sounds, and music
• **Structured Data:** Fits neatly into a defined model, making it easy to organize, store, and share. • **Semi-Structured Data:** Lacks a fixed schema but still has some structure. • **Unstructured:** Has no consistent model or recognizable structure.
• **Captured Data:** Captured data are collected through direct measurement and intentionally. • **Exhaust Data:** By-products generated by devices or systems. • **Transient Data:** Temporary and exist only for a short period, not stored permanently. • **Derived Data:** Result from further processing or analysis of captured data.
• **Primary Data:** Collected directly by researchers using their own methods. • **Secondary Data:** Created by others and made available for reuse. • **Tertiary Data:** Typically published by statistical agencies to protect the confidentiality of the individuals represented.
• **Indexical Data:** Allow identification and linking through unique identifiers like passports etc. • **Attribute Data:** Describe characteristics of a phenomenon but are not used for identification. • **Metadata:** data about data, either detailing content, such as field names and descriptions, or describing the entire dataset.

2.1.2 Data in Digital and Big Data Era

Emerging information technology and the digital era added new layers to the notion of data relevant to the original concept from the late 18th century (Rosenberg, 2013, p. 34). This corresponds with the idea that the original notion of data is in a similar way to the construction of digital era data systems. We can say that the digital era gave a universal aspect to the concept of data. Binary systems that allow to representation of numerical data, texts, images, or even audio in a format create a system that constantly grows the informational sources. Associated only with a science-based concept, data became a universal understanding with a broad aspect concerning informational science and technology (Avanço et al., 2021, p. 7).

Starting with this digital and informational revolution, the perception of data appears to indicate a major epistemological evolution: the shift from empirical induction to neural connectivity aligns with the change from scientific data to informational data, which is now characteristic of the Digital Era (Avanço et al., 2021, p. 7). Just as the semantics in scientific fields change according to the content and create integrated meanings, it can also accommodate and exchange with technological developments. This shows the flexible structure of the concept of data, which emerges from people's cognitive process, contrary to the rigid and clear definitions made at the beginning.

The idea of data itself has been radically transformed by disruptive technologies that restructured how data is produced, analyzed, stored, and shared, moving from something rare and tightly controlled to inexpensive and increasingly open and accessible (Kitchin, 2014b, p. 17). This shift is driven by the latest generation of digital technologies, like smartphones, cloud services, and sensor networks, that now design nearly every side of daily life and turn them into data, feeding into data-driven platforms (Kitchin, 2014, p. 17). The volume and reach of this transformation have sparked strong enthusiasm around the value and possibilities of big and small-scale data, though this optimism is sometimes exaggerated or uncritically accepted. Yet, such transformations are not without criterion: previous historical moments, such as the invention of the printing press, also triggered similar information explosions that reshaped how knowledge was created, organized, and interpreted, marking a pattern in the evolution of data practices as Kitchin (2014b) explains.

2.1.3 Architectural Data and Effects of Big Data

In the architecture, we should also consider different definitions of data. Because the architecture stands on social, modern sciences, and humanities, and it therefore contains different types of data. This diverse type of architectural data requires classification and frameworks to demonstrate. If we cannot represent such a variety of different data, we cannot control it, which affects the efficiency of the design process.

Design data refers to the broad spectrum of information that guides and shapes the development of a design solution. This encompasses not only clearly defined requirements and technical specifications but also less tangible elements such as personal experience and intuitive understanding (Cross, 2006). In the context of architecture, this data takes on a particularly diverse and multifaceted form. Architectural design data integrates various types of information, ranging from spatial and functional considerations to environmental factors and user experience, all of which interact to influence the outcome of the design process (Kalay, 2004).

Menendez et al. (2020) categorize data in design into two main types: quantitative and qualitative. Quantitative data, according to the authors, usually numerical, tells “what” is happening and supports objective analysis, such as building simulations assessing energy use or structural performance. Qualitative data explains “why” something is happening by providing non-numeric insights into experiences, behaviors, or motivations, crucial for designing responsive environments that engage human emotions. They further elaborate that each type holds specific value in different design applications. Some projects focus more on quantitative data for evaluation and analysis, while others emphasize qualitative data for experiential aspects. Often, combining both types gives a comprehensive understanding, enabling designers to tackle complex challenges effectively by integrating measurable facts with experiential insights.

Architecture includes different knowledge types from a variety of areas, such as *physical properties of materials* to *principles of visual perception*. This is leading to the formulation of different systems of inquiry and research methods (Groat & Wang, 2013, p. 65). Since this framework for inquiry is inherently related to inquiring about data, following Groat & Wang’s (2013) models can naturally help us to discover different data types. This foundation is about traditional data types and needs to be bridged with modern data science classifications. When following this, we also need to realize that research methods are rather complicated and do not have clear boundaries. Even it is not the direct depiction of data types, the framework Groat & Wang (2013) proposed for architectural design research: *continuum of research paradigms* shows the complex nature of architecture from technical to historical perspective and indicates the vast array of data types. (Table 2.3) illustrates the philosophical constructions that guide different types of research approaches, specifically through the lenses of epistemology (how we know) and ontology (what is real). It spans from objective paradigms (like positivism/postpositivism) to subjective ones (like constructivism), with an intersubjective position in between. While the table does not explicitly categorize data types, it strongly informs how data is understood, collected, and interpreted in architectural research.

Table 2.3 Continuum of Research Paradigms (Groat & Wang, 2013)

Objective ⟷ **Subjective**

	Positivism/Postpositivism		**Intersubjective**	**Constructivism**	
Epistemology	Knower distinct from object of inquiry	Knowing through distance from object	Knowledge framed by understanding sociocultural engagement	Knowledge co-constructed with participants	Knowledge perpetually provisional
Ontology	Assumes objective reality	External reality revealed probabilistically	Diverse realities situated in sociocultural context	Multiple constructed realities	Infinite realities

In architectural research, the chosen paradigm shapes how data are understood and used. *Positivism/post-positivism* assumes an objective reality and relies on quantitative, structured data like energy metrics or spatial measurements, collected independently from the observer. *Intersubjective* approaches recognize diverse social contexts and use mixed data, combining numerical data with qualitative insights, such as surveys or observational studies. *Constructivism* sees knowledge as co-created, embracing qualitative, thick data like interviews and narratives. Overall, objective paradigms favor measurable data, while subjective ones prioritize experiential, context-rich information, influencing design interpretation and decisions. Conceptual mapping of these possible connections is depicted in (Table 2.4)

Table 2.4 Continuum of Research Paradigms and Data Types (created by author)

Objective ⟷ **Subjective**

Positivism/Postpositivism	**Intersubjective**	**Constructivism**
Quantitative/Structured Data	**Mixed/Contextual Data**	**Qualitative/Thick Data**
• Energy simulations • Daylighting metrics • BIM data	• Post-occupancy evaluations • User surveys with metrics and open-ended responses • Behavior mapping	• Ethnographies • Interviews • Community narratives, • Participatory design workshops

Various sources discuss data types that are employed in architecture, but they are dispersed even the sources themselves may be detailed in their course. Exploration of data types also changes according to the type of architectural approaches embraced (Qabshoqa, 2017). Identifying types of data involved in the architectural design process is crucial because certain type of data requires certain types of processes, thus making it easier to navigate the parts leading to the design product.

Although we know which types of data exist in architecture, we do not have a complete framework that demonstrates this variety of design data across the phases of the architectural design process. Still, some frameworks attempt to capture it. Qabshoqa (2017) classified types of data in architecture based on the aspects of *quantification, format, time,* and *perception* as illustrated in (Table 2.5).

Table 2.5 Types of Data in Architecture (Qabshoqa, 2017)

DATA TYPES	**Quantification**	• Quantifiable • Unquantifiable
	Source	• Digital • Analogue
	Time	• Past • Present • Future
	Perception	• Visual • Auditory • Kinaesthetic • Sense of Smell • Sense of Taste

Based on the Qabshoqa's (2017) classification, we can further extend the depiction by presenting an integrated typology of data used in architecture, merging sensory, temporal, and quantification-based classifications with structural and technical data typologies as depicted in (Table 2.6).

Table 2.6 Extended Types of Data in Architecture (based on Qabshoqa, 2017)

Category	Subtypes	Examples
Quantification	• Quantitative • Qualitative	Square meter calculations are *quantitative*, Design descriptions are *qualitative*
Source	• Digital • Analogue • Captured Data • Exhaust Data • Transient Data • Derived Data	A 3D laser scan is *digital*, A hand-drawn sketch is *analogue*, Sensor data from smart buildings is *captured*, Server logs from a building are *exhaust data*, Real time occupancy data is *transient*, Climate comfort scores calculated from raw weather data are *derived*.
Time	• Past • Present • Future	The architectural data embedded in historical architecture and literature regarding dimensions, styles and narratives are *past data*, Current user movement in a space is *present data*, Projected energy consumption is *future data*.
Perception	• Visual • Auditory • Kinaesthetic • Sense of Smell • Sense of Taste	Facade material color perception is *visual*, Acoustic simulation output is *auditory*, Movement comfort in stair design is *kinaesthetic*, Air freshness rating are based on *smell*, *Taste* perception is rare but may apply in food-related design.
Structure	• Structured • Semi-Structured • Unstructured	A BIM database model is *structured*, JSON logs from Iot Sensors are *semi-structured*, Design sketches and photos are *unstructured*.
Producer	• Primary • Secondary • Tertiary	Survey responses from building users are *primary data*, Government census data reused for planning is *secondary*, A report summarizing other reports are *tertiary*.
Function	• Raw Data • Processed Data • Metadata	Direct sensor output is *raw*, Airflow simulation results are *processed*, BIM model file details like author and date are *metadata*.
Type	• Indexical Data • Attribute Data • Metadata	A geo-tagged photo is *indexical*, Room temperature is *attribute data*, File creation time is *metadata*.
Context	• Experimental • Observational	Wind tunnel test results o a model are *experimental*, Time-lapse footage of a site is *observational*.

However, the classification of data in architecture is not only a matter of categorization but also one of epistemological framing. The nature of data, whether it is objective, subjective, sensory, computational, or emergent, plays a critical role in how it is collected, interpreted, and applied within the design process. As shown in the expanded table, data may be sorted according to quantification, source, time, structure, and perception, among others. These typologies provide a useful foundation, yet it is essential to recognize that data classification is context-dependent and dynamic. Different framing lenses may lead to alternate taxonomies. For instance, data can be classified according to its technical properties (structured vs. unstructured), functional role (raw vs. processed), or semantic intent (attribute vs. indexical).

More on the topic of epistemological difference, in terms of methods of data collection for designers, Melendez et al. (2020) states that data collection has historically coexisted as analog and digital approaches. Analog methods include direct observation, experiential research, and surveys, focusing on capturing qualitative insights into human behaviors and environmental contexts. In parallel, advancements in technology have introduced digital methods such as sensors, smartphones, and social media platforms, which not only record but also generate substantial amounts of data in real time. These technological tools enable more precise, large-scale, and dynamic data collection, empowering designers to make informed and responsive decisions throughout the design process. This integrated approach allows for a comprehensive understanding of both environmental conditions and human needs. Furthermore, the origin of data, whether it stems from analogue traditions or digital ecosystems, adds an additional layer of complexity, especially as hybrid workflows become increasingly common in architecture.

Moreover, as digital technologies evolve, the kinds of data in use are expanding. BIM platforms introduce object-oriented and combined datasets, AI systems require training and predictive data, and big data ecosystems thrive on high-volume, heterogeneous, and real-time data streams. These forms often resist simple categorization and instead reflect a relational or networked nature, where data exists as part of a larger system of interdependencies. The classification of architectural data must remain flexible and reflective of the goals and tools at hand. Whether framed through its nature (e.g., sensory, qualitative, computational), its properties (e.g., format, structure, scale), or its function within digital applications, understanding how data is defined and deployed is essential for constructing a coherent and purposeful design process in the age of data-driven architecture.

With the growing impact of big data, this lack of representation becomes clearer. Since there is no complete framework that shows all data types take place in architecture, it can get confusing. Design inputs, or we can name them design data, have been a fundamental part of the architectural design process. Whether it is a visual input or a sketch, these are all part of the inquiry process. There are different types of data inputs in every stage of the process. But some of them fall under the category of quantitative or visual, or just at the tacit knowledge level. Because of that, we may not always fully grasp or determine the inputs. Clarifying different types of data in the different processes of architectural design can help us, as it may lead to a more aware design process and a design product.

And why is the topic of data types crucial? It is not because of the trend that has been going on for the decades. Sure, it draws attention to the subject as everyone says big data is the future of companies or the research communities. But it has been there from the beginning, even before the term of data making its appearance in the books. But with the digital revolution, data started to pile up. Of course, just the size of the data is not significant on its own. But the managing and drawing insights from this source is what gives meaning to it.

Data in architecture has been associated with technical data for a long time. Ernst Neufert published his reference book in 1936 with the title 'Architect's Data', which provides spatial requirements in building design and site planning. Tools that operate with technical data are capable of developing structures with different levels of detail (Hois et al., 2009). But technical data only performs as a guideline or constraint and needs to be connected to meaningful designs. Technical data needs to be supported by qualitative and conceptual inputs (Hois et al., 2009). Qabshoqa (2017) represents various architectural approaches as *design as ideology, design as profession,* and *design as service,* which he emphasizes data and information interpretation, and types vary accordingly to different perceptions. When architecture is being handled as ideology, it is associated with cultural and ideological values, and the data that comes with approaches are *qualitative, symbolic, philosophical,* and *unquantifiable* (Qabshoqa, 2017). Whereas in professional and service-based approaches, it comes with economic, marketing, and building performance and service data. Qabshoqa's (2017) proposal reveals that data is more than an input and has value-driven attributes and helping to develop values.

Research within the field of BIM and AI with IFC (Industry Foundation Class) or investigating the impact of big data on these fields is highly popular. IFC is a standardized data model specifically designed to represent detailed information about buildings and construction projects. It allows different stakeholders (like architects, engineers, contractors, and facility managers) and their various software tools to share and exchange building information seamlessly (Panagoulia, 2023). Because the application and usage of these tools mainly rely on data or information management, we can detect which data types are highly integrated and in use currently through recent studies.

As Chen (2021) describes, design fields include environments such as *humanistic features, geographic information, construction technology,* or *entity functional space,* and include both *quantitative engineering data* and *qualitative analysis of users, psychological feelings,* and *evaluation information*. Data and information management and optimization are highly popular in fields building management, smart city strategies to maintain and diagnose the intelligent building systems.

We can generalize the data types in architecture, under the influence of AI, big data, and IoT (internet of things), and its application areas, such as smart building systems, urban mobility networks, or intelligent environmental monitoring systems. The Internet of Things (IoT) describes a system where physical objects are equipped with technologies like sensors and software, enabling them to connect to the internet and share data with other connected devices (Atzori et al., 2010). Digital technology also affected the cultural-heritage discipline by expanding available tools, sharing data more widely, and introducing digital ecosystems (Néroulidis et al., 2024).

In their systematic literature review for generative AI in architectural design, Jang et al. (2025) explored sixty papers in the last ten years and revealed the data types and contents used in AI models for architectural design. The paper identifies primary data types used in AI models in architecture as: *visual, categorical, text, graphic data,* and *3D models* or *combination of image, text, and graphics.*

Sönmez (2018) stresses that AI depends on large datasets, and architecture can be advanced by mining the rich knowledge embedded in existing buildings and designs. AI can help turn them into usable, learnable data for intelligent design systems. This indicates the discipline of architecture is experiencing a shift from treating design examples as static references to viewing them as dynamic data sources. Depiction of the shift from traditional to future design systems can be seen in (Table 2.7) This transformation, driven by AI and data-centric approaches, redefines examples as formal, semantic, perceptual, and contextual data that can be processed, learned from, and used to develop intelligent design technologies.

Table 2.7 Evolving Types of Data in Architecture (created by author)

Traditional Architecture	Current Architecture	Future Design Systems
Hand-drawn plans and physical models	Formal geometric data (CAD, BIM, 3D models)	Dynamic generative models that can adapt and optimize during design.
Visual references (photos, sketches)	Semantic data (classified elements: walls, windows, etc.)	Automatic extraction of detailed building ontologies and relations.
Spatial intuition and qualitative judgments	Perceptual data (measurable qualities: openness, privacy, flow)	Predictive models for user experience and spatial behavior.
Case studies and precedents (unstructured)	Contextual data (adjacency patterns, typologies, cultural contexts)	Context-aware design systems that reason like human architects.
Verbal heuristics, tacit design knowledge	Machine-learning models trained on larg example datasets	Continually learning AI design assistants that evolve with new cases and styles.
Style based on subjective judgment	Style classification models (pattern recognition, feature extraction)	Generative systems capable of creating or remixing styles automatically.

As Bernstein (2018) states, even technology does not replace the act of thinking, it is changing the frame it occurs. The first digital turn in architecture was characterized by computer-based geometric modeling. Followed by the second digital turn, which is cloud-based, and its characteristics were big data-enabled machine learning and AI. These significant eras were also bridged by BIM technologies (Bernstein, 2018). Bernstein (2018) explains that in the pre-digital era, architectural data or representation methods were mostly characterized by hand-crafted drawings and personal experience. Representation of buildings relied on plans, sections and such, allowing architects to communicate complex spatial ideas through two-dimensional drawings. The reasoning process was also rather manual, depending on mathematical formulas and human judgment to solve rather *tame problems.* Collaboration happened through physical forms, printed documents, and in-person meetings, which can count as another form of traditional data sources.

Landscape of the architectural data has been reshaped by digital technology (Bernstein, 2018). CAD technology brought the ability to draw more efficiently, but BIM fundamentally started the true transformation by enabling to linking of geometry with metadata about materials, systems, and performance. Data now travels directly from the architect's digital model to machines on construction sites. The collaboration evolved into real-time, cloud-based platforms. This decentralization gave data fluidity across disciplines. As time progresses, Bernstein (2018) suggests that architectural practice is shifting toward an outcome-based design with predictive data and machine learning, as it is visualized in (Figure 2.1). Historically, data and information served firstly to represent design intent through hand drawings and physical models (Bernstein, 2018). Early digital drafting tools enhanced production speed, but they did not alter the symbolic role of data. With the arrival of BIM, data became dynamic. As architecture embraces machine learning and big data, Bernstein (2018) foresees a profession organized around an outcome-based delivery, where the value of data lies not in representation but in performance and measurable results.

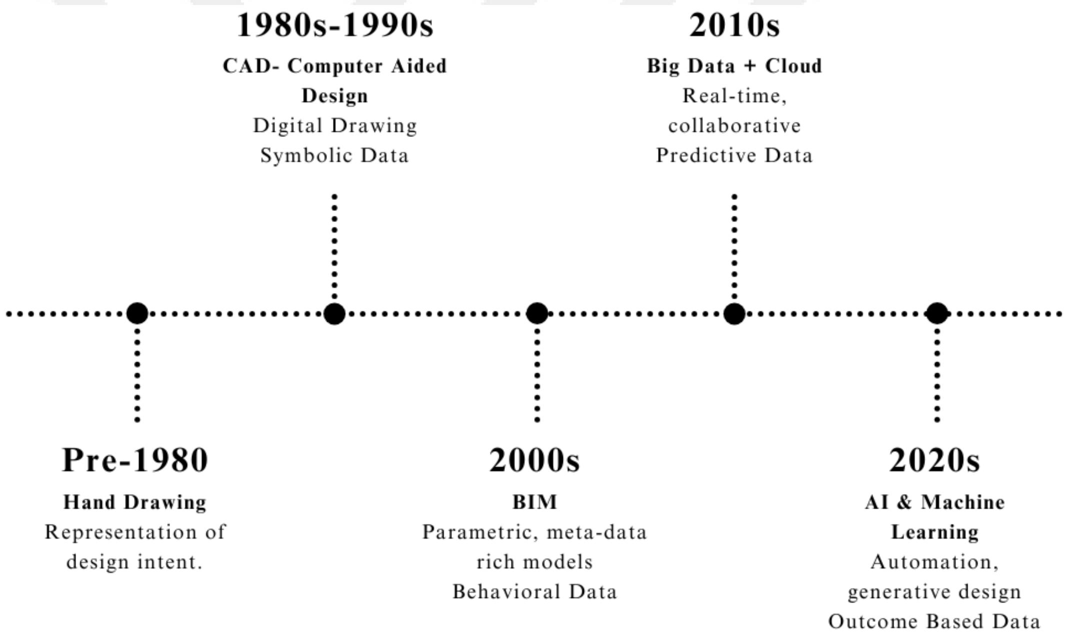

Figure 2.1 Timeline of the Architectural Data Evolution (based on Bernstein, 2018)

Although Mortati et al. (2023) primarily focuses on design thinking, which is still relevant due to its valuable insights into how different forms of data, particularly big data and thick data, can inform and enhance design processes. These insights are highly relevant to architectural design, where the integration of diverse data types is increasingly shaping how spaces are conceived and developed. Design thinking is predominantly related to qualitative data, such as user observation or participant interviews (Mortati et al., 2023). The architectural design process also carries similar features in terms of the qualitative data types it holds. Michelli et al. (2019) describes this type of quantitative data as the main tools and methods for sourcing data, used in design thinking as *ethnographic methods, personas, journey maps, brainstorming, mind maps, visualization, prototypes,* and *field experiments.* Mortati et al. (2023) groups this type of data in design thinking research as *thick* or *enriched data*. Still, the outcomes of the impact of big data on design thinking are relatively limited. The main sources that sustain design thinking is mostly based on qualitative aspects. This hypothesis also validates on some parts of the architectural design process, as the process of ADP leans on qualitative data or thick data on some levels.

Thick data refers to forms of data that go beyond simple observation and quantification. It encompasses the context, emotions, intentions, and cultural meanings behind human behavior, offering a richer and more nuanced understanding of social phenomena. The concept first appeared in philosophical discourse as "thick description," introduced by Gilbert Ryle, who emphasized that human behavior should be interpreted not only by its observable form but by the underlying intentions and meanings (Harrison, 2013, p. 860). Anthropologist Clifford Geertz later expanded this idea, framing thick description to uncover the symbolic layers of cultural practices (Geertz, 1973).

In the context of design, thick data is seen as experiential material that can take many forms, from tangible textures to natural visual patterns, serving as inspiration for creative work (Mortati et al., 2023, p. 3). These inputs are deeply connected to emotional and sensory responses, forming a basis for intuition and affective engagement in the design process.

Architecture, in particular, is inherently multisensory. The way people experience a space involves not only vision but also sound, touch, smell, and the perception of spatial scale through the entire body. Sensory responses emerge from interactions between material qualities and bodily presence, forming a comprehensive perceptual field (Holl et al., 2008, p. 30). However, modern scientific approaches have often emphasized reductionism, aiming to understand complex phenomena by breaking them down into fundamental parts and treating the whole as the sum of these components (Pallasmaa, 1996, p. 449). This mindset has influenced architectural theory and education, which frequently prioritize visual aesthetics and formal composition. Based on principles like those taught at the Bauhaus, architecture has often been reduced to visual experimentation with form and space, disconnected from real-life sensory experience or deeper cognitive and emotional resonance (Pallasmaa, 1996, p. 449).

Thick data provides a critical counterpoint to these reductive tendencies by emphasizing meaning, context, and human experience. It advocates for a more integrated and embodied understanding of design, one that reconnects with the full range of perceptual and cultural dimensions that shape how we engage with built environments. Since the thick data needs assistance with the growing impact of technology, different types of data need to be integrated into the design thinking process. In this situation, big data becomes highly relevant. Mortati et al. (2023) approaches design thinking projects with a blend of qualitative data and quantitative analysis methods and reveals which types of data are used at which stage of design thinking projects. Thick data in design processes is related experiential learning process targeted for enhancing creativity. Michelli et al. (2019) states that, instead of seeing data as something technical, designers can creatively work with data. Data in design today offers new ways to understand how people behave and how spaces are used.

2.2 Understanding Knowledge

The definition of knowledge, like the meaning of data, is controversial and includes different comments from various colleagues, and before we dive into the semantics, we can identify that the beginning of the usage of writing systems gave knowledge a cumulative and transferable aspect (Liew, 2013, p. 3). Hume (1979) mentions that sensory experiences lead to the cumulation of knowledge, attributing it to a more experiential notion. From a different period, Latour (1987) describes it from a scientific perspective and calls it a ‘social construct’ that has been fed from scientific frameworks. In the Data-Knowledge-Information trio, (Floridi, 2011) says it is contextualized and meaningful data implemented in information. As a perspective in contemporary discussion, (Zins, 2007) published an article in which he gathered the results of a study on information science, the paper includes 130 different definitions of ‘data’, ‘information’, and ‘knowledge’ from 57 scholars in the major fields of information science. This article draws a broad picture of current discussions and gives directions for blind spots and further implementations.

The idea of knowledge is flexible and diversified, and the classification depends on the issue that people want to acquire. It requires a certain type of way to reach or have access to so-called desired information (Rescher, 2003, p. 2). This mechanism manifests itself in the formation of knowledge types as illustrated in (Figure 2.2).

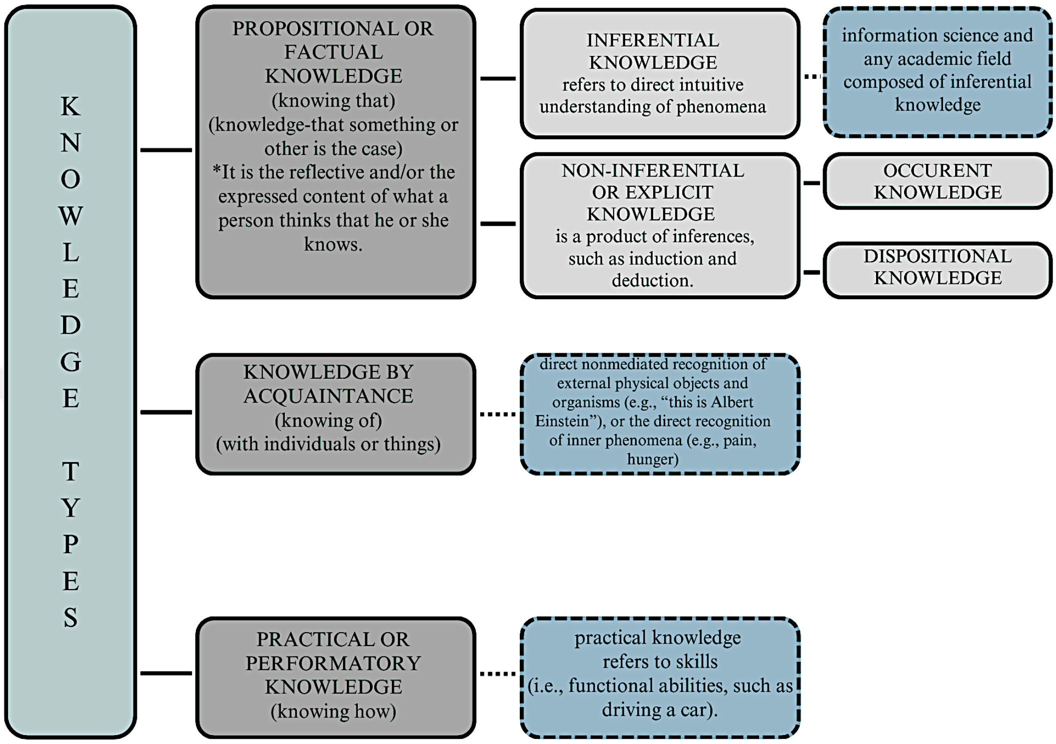

Figure 2.2 Knowledge Types (based on Zins, 2007 and Rescher, 2003)

Propositional knowledge is a cognitive process that resulted from the interaction of people and facts (Rescher, 2003, p. 4). And it can be divided into two categories, which are inferential and non-inferential (explicit) knowledge. Explicit knowledge is more time-responsive, so to speak. It is more about actively paying attention or giving answers or producing actions on demand (Rescher, 2003, p. 8). And it takes two forms as 'Occurrent' and 'Dispositional' knowledge. For inferential knowledge, it is a deeper type of knowledge that develops in a course and that we can call process-dependent. Instead of giving active answers, it can start from the outputs of the information obtained from occurrent and dispositional knowledge (Rescher, 2003, p. 8).

To begin with, these discussions of describing knowledge are taking part in the propositional knowledge specifically, which it has been referred to as 'that knowledge', traditional epistemology focused on this. Approaches to describing knowledge appear in two categories. The existence of the knowledge can manifest itself as a thought in the subject's mind or may appear as an object (Zins, 2006, p. 449). Plato's approach to it appears as 'justified true belief', is a thousand years old, but Edmund Gettier challenged it in his 1963 philosophy paper, where he describes cases that justified true belief may not always be described as knowledge. After the Gettier Problems, different attempts for the definition of knowledge have been proposed to eliminate these cases and still capture knowledge.

The second approach describes knowledge by detaching it from the individual and represents it as an independent object that results from objective propositions (Zins, 2006, p. 450). Objective knowledge definitions also exist with counter-ideas that suggest 'knowledge is not an independent object' or an agnostic viewpoint state 'we can't know objective knowledge exists outside of the subjective mind'. Along with these discussions, Zins (2006) suggests that since we might not be sure knowledge is true, it is fairer to call it 'Universal', 'Collective' or 'Inter-Subjective Knowledge'. In the end, these two approaches are complementary, and the construction of universal knowledge depends on the interpretation of subjective minds. It is externalized by subjective knowledge with devices such as documents, records, or physical objects (Zins, 2006). We can observe the realms of knowledge in (Figure 2.3).

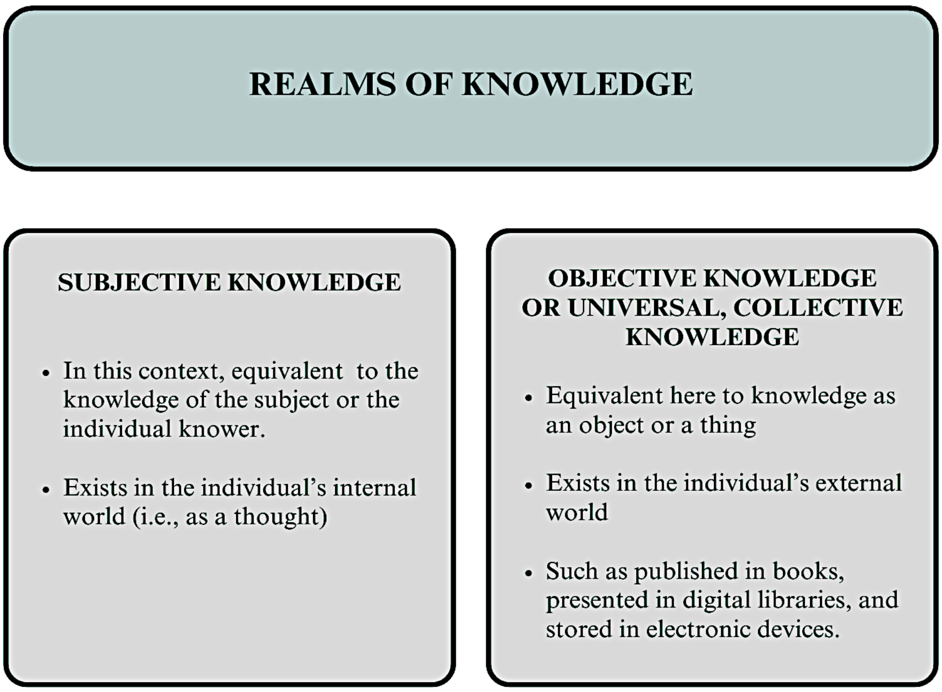

Figure 2.3 Realms of Knowledge (based on Zins, 2007)

These approaches to define knowledge guide users' cognitive behaviour and their research ability. They also led to the way of information retrieval and thus affected the mechanism of information science (Hjørland, 1998, p. 606).

Rescher's (2003) thoughts on knowledge are also intriguing, as he counters the idea that knowledge is a 'mental activity'. Rather than something you can acquire by being in the process, he claims 'to know' is related to being a product of a matter. Differentiating the verb of 'knowing' from such as wandering or thinking, one can know something without engaging a person's mental process. The act of knowing does not mean anything; it is a cognitive relation with a fact.

We need to challenge the limits of knowledge, instead of focusing on what we know, we can question ourselves with directive sentences such as 'I cannot answer that' (Rescher, 2003). This state of mind can direct us and may help us face our lack of or limits of knowledge. If the knowledge is what we acquire from questions we can resolve (Rescher, 2003, p. 3), then this area can be challenged by framing new questions within an epistemological context.

Pursuing knowledge is sort of a natural process, because we feed on its outputs, as we go along with our lives, it keeps posing questions and demanding answers (Rescher, 2003, p. 4). Since it is in our nature to understand, we will keep following that instinct to satisfy it. As long as a mystery or some matter is demanding to be solved, that is what keeps our life meaningful. Knowledge is an urge that leads us to acquire information (Rescher, 2003, p. 5). We want to make sense, collect information, and most importantly, understand the environment we live in. 'Cognitive orientation' is what keeps our minds at ease, whereas 'cognitive disorientation' poses stress for our being (Rescher, 2003, p. 5). In the end, all of these activities are what keep us going for research, innovation, and most basically to survive.

2.2.1 Data Information and Knowledge Models

Data, Information, and Knowledge models are conceptual approaches that are used to demonstrate the relationship between them concerning each other. As we explored earlier, there are various approaches and subheadings for defining each of them. Earlier depictions of the DIK/ DIKW models established as a pyramid, hierarchically, but the DIKW Pyramid evolved into other frameworks over time as scholars criticized its hierarchical structure by amplifying that the pyramid assumes a linear, step-by-step progression, but in reality, these categories overlap and interact dynamically. Along with them, many other studies are criticizing the decision-making processes in business management, knowledge management, or domain-specific areas. The model had evolved from static categories to action, decision making, and real-world applications (McElroy, 2003; Rowley, 2007; Snowden, 2002; Snowden, David J. & Boone, Mary E., 2007).

Cleveland (1982) introduced the DIK model, which was developed in information science without the ‘Wisdom’ layer. Russell Ackoff (1989) was one of the first scholars to explicitly add "Wisdom" to the pyramid. He emphasized that wisdom is not just about knowing but about beneficially applying knowledge. Defines data, information, knowledge, understanding, intelligence, wisdom, and emphasizes that the development of wisdom is crucial to gain value. This draws attention to the way we obtain value and how we define the terms, as it leads to the way of wisdom. Although Ackoff did not provide a visual diagram, the pyramid representation has since become a widely used schematic to represent the hierarchy as seen in (Figure 2.4).

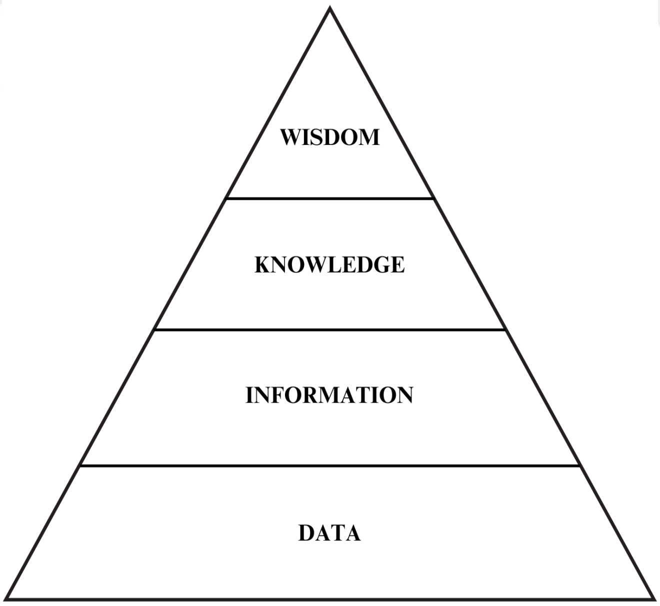

Figure 2.4 Visual Representation of the DIKW Hierarchy (adapted from the conceptual model by Ackoff, 1989)

Literature of the wisdom in information systems and knowledge management is limited to numbered researchers, even wisdom is at the top at the knowledge hierarchy (Rowley, 2007, p. 174). Similar limited definitions also applied in psychology (Liew, 2013, p. 5). Rowley (2007) comments it could be because the community views wisdom out of their realm or it is thought too vague to be included. Or it is associated more to the humanistic features than systems. For its limited discussions, ‘generally defined in terms of value-driven (ethical) intelligent behavior’ (Baskarada & Koronios, 2013, p. 8). By combining available definitions, Rowley (2006) defines it as ‘The capacity to put into action the most appropriate behaviour, taking into account what is known (knowledge) and what does the most good (ethical and social considerations)’. Both in practical and theoretical discussions, wisdom is a multi-layered concept that maintains its link to the knowledge and their interactive nature for the better.

Ackoff’s views on Data, Information, Knowledge, and Wisdom can also be interpreted from the perspective of time. Representation of this approach shows that data, Information, and Knowledge are part of understanding the past (Qabshoqa, 2017, p. 34). Specifically for Knowledge, it also leans into the Present as we actively perform it (Clark, 2004). The Figure Continuum of Understanding, as illustrated in (Figure 2.5), represents that we can gather knowledge with context and understanding. Cumulation of context leads to the experiences one can make use of and, as the understanding of a subject grows, one is more capable of handling experiences through researching, absorbing, doing, interacting, and reflecting.

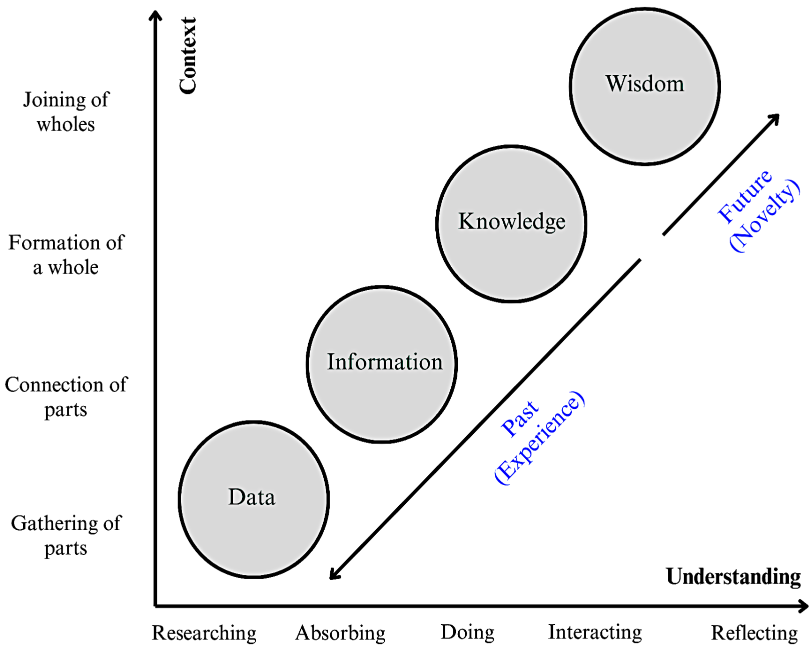

Figure 2.5 The Continuum of Understanding (Esterbrook, 2012)

Also, Bellinger et al. (2004) further refined the model, depicted in (Figure 2.6), emphasizing that wisdom involves ethical judgment and deep understanding, not just an accumulation of knowledge. Adding a wisdom layer brought the perspective of applying knowledge effectively, highlighting decision making of information use. It is not just a 'top' layer but an ongoing cycle of refinement. They have offered a model that emphasizes continuous feedback loops instead of a strict hierarchy.

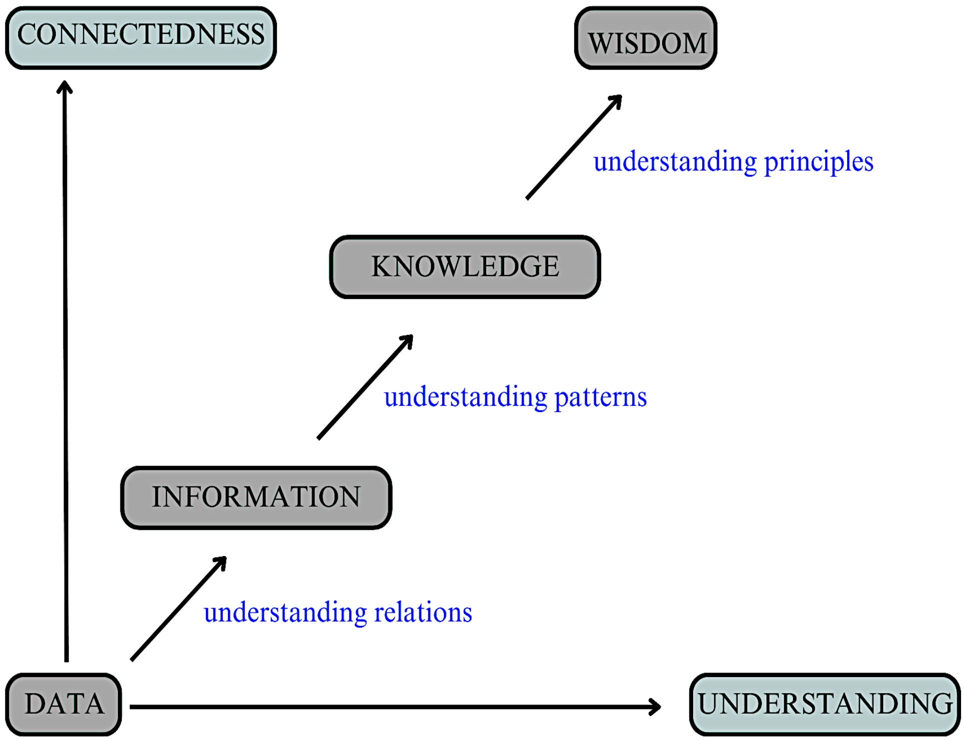

Figure 2.6 The Transitions from Data to Wisdom (Bellinger et al., 2004)

Chaffey and Wood (2005) discuss the relative value of each stage within the hierarchy, in (Figure 2.7), suggesting that as one moves from data to wisdom, the value increases. This conceptualization underscores the transformation process from mere data to actionable wisdom, highlighting the increasing value and utility at each stage.

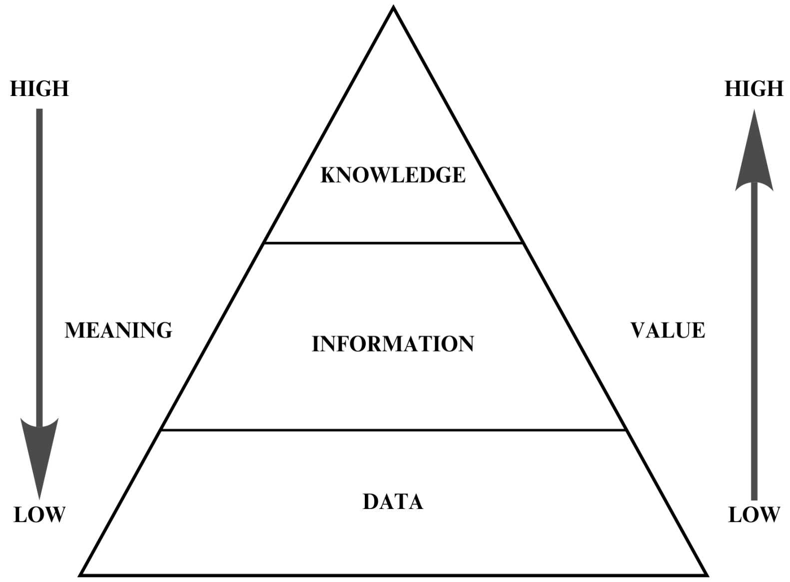

Figure 2.7 DIKW Hierarchy (Rowley, 2007)

Rowley (2007) criticised the current DKIW models that include a limited perspective on the wisdom or how it culminates and questioned the rigid separation between levels. Wisdom is not just a conceptual knowledge but is linked to real-world decision-making. In the suggested model, which can be seen in (Figure 2.8), the concepts such as meaning, value, and transferability as those terms mostly seen in definitions of knowledge and information.

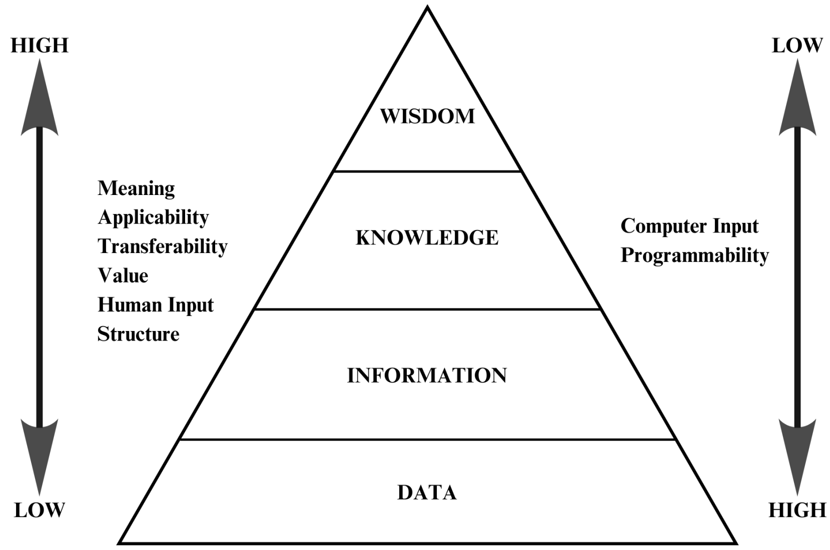

Figure 2.8 DIKW Hierarchy Interpretation of Rowley (Rowley, 2007)

Choo (2005) contributed to the hierarchy by emphasizing the role of sense-making, knowledge creation, and decision-making in organizations. His work, in (Figure 2.9), aligns with the idea that data transforms into information, then into knowledge, and ultimately into wisdom through processes such as interpretation, context application, and learning.

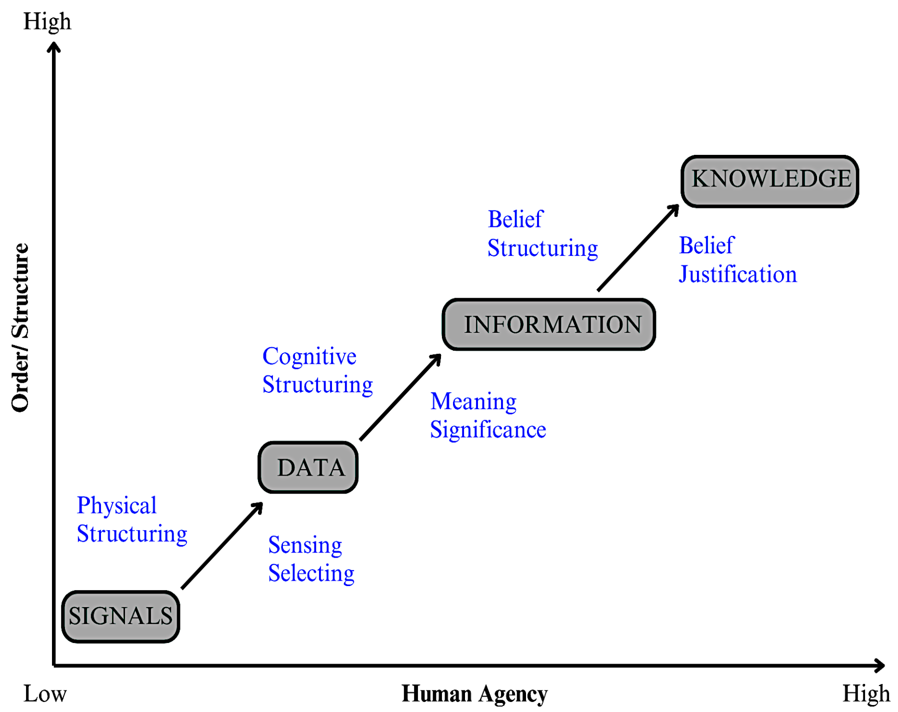

Figure 2.9 DIKW Hierarchy Interpretation of Choo (Rowley, 2007)

Along with different approaches in DIKW, we could say there are some divisions for constructing these models. Subjective and Universal Domain are the main theoretical grounds for formulating DIK models (Zins, 2007, p. 488). In his article, Zins (2007) analyzed 45 citations and spotted five different models for the most common DIK definitions which depicted in (Table 2.8). The most common one, namely Model-1, aligns with the Information Science. It takes Data and Information under the Universal Domain, which is seen as an external phenomenon. Knowledge is less explored, which is seen in the Subjective Domain therefore an internal matter. On the other hand, formulating definitions of DIK and being reflective on the subject requires the cooperation of these two domains.

Table 2.8 Models for Defining Data, Information and Knowledge (Zins, 2007)

Model 1		Model 2		Model 3		Model 4		Model 5	
UD	SD	UD	SD	UD	SD	UD	SD	UD	SD
D		D		D		D	D	D	D
I			I	I	I	I	I	I	I
	K		K	K	K		K	K	K

UD = Universal Domain SD= Subjective Domain D= Data I= Information K= Knowledge

Exploration of the individual terms and their conceptual formulations, points that this is a highly versatile field. There are many definitions and models defining data, information, knowledge and wisdom, yet circular definitions of the terms are problematic (Liew, 2013, p. 2). Exploring their inter-relationships does not mean defining them. This is a significant view on expanding or evaluating information science. Internalizing the right set of mind and minding the inter-relationships would bring more value to the area than numerous definitions.

2.2.2 Knowledge and Architectural Design in The Era of Big Data

With the Big Data era, knowledge creation methods have undergone a change. Small samples and theoretic models were the basis of traditional knowledge, whereas, in the era of big data, there is a movement towards data-driven knowledge where huge sets of data are analyzed in order to identify patterns without explicitly beginning with a theory (Kitchin, 2014b). Although there are promising new opportunities in big data, there are also significant questions raised over data quality, bias, ownership, privacy, and the politics surrounding data control. Big data is not neutral; it is under the command of whoever collects it, how it is processed, and for what purpose (Kitchin, 2014b).

Traditionally, knowledge exists in often physical forms and therefore is more static, central, and limited. But in the Big Data World, Floridi (2014) refers to it as *infosphere* which means that knowledge is now co-existing side by side with machine-created knowledge. Such a new approach introduces a new vision in design and architecture. As humans, Floridi (2014) refers to as *informational organisms*, inhabiting this interactive *infosphere*, built environment and world of information and now coexisting in this new environment. The fluidity of the new environment provides designers the opportunity to challenge the old mindset of the space and depict the means of collaborating with data-driven agents. When knowing tramforms into navigating and co-creation rather than following a traditional static mindset, architectural design is also shifting. Spaces we inhabit turn into physical and digital or *phygital* environments (Floridi, 2014). It demands a new architectural strategy that addresses both physical requirements and digital presence too. Because how we inhabit environments is also changing from physical to informational physical.

In a world of over-connection and an excessive amount of information, it is increasingly important to curate knowledge and determine what is significant (Floridi, 2011). Previously, knowledge developed through theoretical methods, beginning with questioning and confirming or rejecting theories. Now, according to Anderson (2008a), we need not start with theories; patterns can emerge directly from data analysis. Architects also face the challenge of design flows of information, as architectural design can now be approached through data-driven methods which reveal patterns.

Oxman (2006) refers to how architectural design evolves following digital technologies and provides a model of *digital design thinking*. As digital tools (CAD, BIM, parametric design, VR, generative design) entered, architects now make complex data a part of the design process (Oxman, 2006). It also altered how knowledge is perceived as ‘dynamic knowledge’ in which architects are now working on live data in a more flexible and adaptive manner.

All of the data or knowledge-based settings and newer technologies also bring new skills and ethical issues. Large quantities of data that arise in numerous settings also bring the significance of bias, sourcing, and choosing it. Main theoretical methods are also shifting. What is upon designers in architectural design is to adopt concepts like data-driven decision-making, ethical utilization of data, interdisciplinary collaboration, and knowledge integration instruction.

2.2.3 The DKIW Hierarchy in Architectural Design and Effects of Big Data

Floridi (2010) argued that we are entering a *fourth revolution*, one that reshapes not just our place in the natural environment, but also the role of architecture. As reality shifts toward an informational form, architecture follows. According to Floridi (2010), we are moving into environments that are increasingly synchronized, delocalized, and correlated. Just as earlier agricultural and industrial revolutions transformed our built world, Floridi (2010) suggests that the information revolution will bring similar, lasting changes.

As an interpreted version of DKIW hierarchy in architecture, Qabshoqa (2017) illustrates the hierarchical relationship between data, knowledge, and value within architectural design in (Figure 2.10). The model is based on the DIKW pyramid and highlights how raw data, when systematically collected and organized, becomes information. Through analysis, synthesis, and interpretation, this information transforms into knowledge. Value, in this context, is established only when profound knowledge is applied. The model emphasizes a top-down process where processing data leads to deeper understanding and value creation in design decisions. Conversely, a bottom-up approach allows observation of data types based on predefined values. The model underscores that consciously operating data to drive the entire design process and business is essential for realizing value from data in architecture.

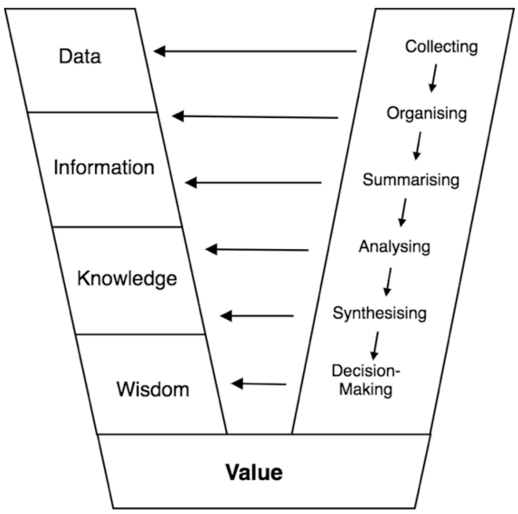

Figure 2.10 Data, Knowledge, Value Model (Qabshoqa, 2017)

Big Data, though not explicitly mentioned in the transformation of the DKIW hierarchy, plays a significant role in reshaping it. It changes how data is sourced and structured, and how decisions are made at every level which can be observed in (Table 2.9). The DKIW model's complex flows are naturally intertwined with Big Data, but this does not mean architectural intuition or ethics are being replaced. Instead, Big Data redefines the field while leaving core human judgment intact.

Table 2.9 Transformation of Data, Information, Knowledge and Wisdom Under the Influence of Big Data (created by author)

	Before Big Data	Early Big Data	Advanced Big Data	Next-Generation Big Data
Data	Fragmented, manually collected data (e.g., physical si surveys, hand-measured sketches)	Raw, unstructured digital data (e.g., early climate records, basic GIS datasets, sensor logs)	Structured and semi-structured integrated data (e.g., BIM metadata, occupancy data, social media data)	Real-time, context-aware, AI-processed data (e.g., Digital Twins, live user feedback, predictive analytics)
Informatior	Static reports and drawings (e.g., CAD drafts, hand-drawn maps, paper schedules)	Aggregated into basic digital formats (e.g., simple site analysis reports, early GIS layers)	Dynamic dashboards and analytical models (e.g., BIM analysis reports, energy simulation outputs)	Immersive, interactive environments (e.g., VR/AR city models, real-time simulation feedback)
Knowledge	Experience-based interpretations (e.g., relying on accumulated personal design knowledge)	Manual expert analysis supported by digital tools (e.g., architects cross-checking GIS data with survey data)	Collaborative knowledge systems (e.g., shared BIM environments, interdisciplinary data coordination)	AI-driven recommendations combined with human critical judgment (e.g., using generative design but filtering outputs for cultural fit)
Wisdom	Intuition and tradition-based decision-making (e.g choosing site orientation based on customary practices)	Human expertise enhanced by basic digital insights (e.g., early energy modeling influencing material choices)	Data-supported but human-led strategic decisions (e.g., choosing building form based on multiple simulation outputs)	Human wisdom is indispensable for ethical and contextual decisions (e.g., critically evaluating AI proposals for social sustainability and urban fairness)

Sun et al. (2022) highlight that beyond Big Data, the deeper layers of data, information, and knowledge models are just as critical for decision-making, especially across broad systems like society. For architecture, this means that it is not enough to manage large datasets like environmental inputs; we must also consciously expand and structure the higher tiers of the DKIW hierarchy. As Sun et al. (2022) point out, DKIW concepts are still underused in both academia and practice, despite the popularity of Big Data analytics. This gap signals an opportunity: structured knowledge models could strengthen architectural design processes, especially as we move toward more intelligent analytics.

This shift is already visible. Oxman (2017) mapped the evolution of design thinking, showing how it moved from physical materials to computational models. Through practices like *sketching by code*, Oxman (2017) revealed how design skills and cognitive approaches are fundamentally changing. In parametric design thinking, the DKIW layers transform: data and information are translated into computational logic, and knowledge is embedded directly into the design. Yet wisdom, the top layer, remains distinct. Even as machines begin to support parts of the process, true wisdom still requires human judgment.

Of course, this connection between Big Data, DKIW, and architecture is not without challenges, as Peters et al. (2024) notes that the DKIW hierarchy faces criticism in the age of AI. Some argue it no longer fully captures the complexities of knowledge and wisdom today. Acknowledging these critiques matters; it reminds us that wisdom remains deeply human and social, and that any framework we use must be flexible enough to adapt. Even so, the DKIW model still offers a valuable lens and might yet inspire novel approaches to integrating data, knowledge, and ethics into architectural design.

Chapter 3

Understanding Big Data

The concept of Big Data is framed within the larger context of technological and societal transformation, where digital infrastructures and connected devices have made data generation a pervasive part of everyday life. This shift is not only technological but also epistemological, reflecting changes in how information is produced, interpreted, and valued. Definitions of Big Data are explored through multiple disciplinary lenses, showing how early descriptions focused on data volume and computational limits, while more recent frameworks highlight dimensions such as velocity, variety, veracity, and value. These evolving definitions reveal the complexity of categorizing Big Data and underline its interdisciplinary and dynamic nature.

Historical developments trace the emergence of Big Data from early computational advances in the mid-20th century to the rise of large-scale digital infrastructures. Military and commercial efforts played central roles in shaping data practices and platforms, while broader cultural and institutional shifts influenced how data was integrated into governance, science, and daily operations. The scale and speed of data production led to new methods of storing, organizing, and interpreting information, often accompanied by ethical concerns such as privacy, control, and accountability. As data became more central to decision-making, questions emerged about the authority to collect and interpret it, and about the consequences of its widespread use.

Management strategies for Big Data are presented through its collection, storage, and analytical processing. Tools and infrastructures such as data lakes, distributed systems, and machine learning techniques enable the handling of complex and voluminous datasets. Traditional methods are shown to be insufficient for managing such data environments, leading to the development of new computational frameworks. The evolution of data science reflects this shift, combining statistical, algorithmic, and domain-specific expertise to generate insights. Alongside technical advancements, concerns about transparency, theoretical grounding, and interpretive limitations are raised, especially in scientific contexts where data-driven methods sometimes replace hypothesis-driven inquiry.

While Big Data has introduced significant possibilities for innovation across disciplines and sectors, critical perspectives draw attention to its philosophical and methodological implications. Data alone is not assumed to be self-explanatory. Instead, its value depends on the ways it is framed, analyzed, and understood. This positions Big Data not only as a technical phenomenon but also as a socio-cultural one that reshapes how knowledge is constructed and applied.

3.1 Overview of Big Data

The arrival of advanced information and communication technologies (ICTs) has fundamentally reshaped how data is generated, collected, and utilized across various sectors. This transformation is evident in the expansion of digital devices and platforms, ranging from mobile technologies and cloud computing to social media and the Internet of Things, which are increasingly spreading through everyday life, capturing data from work, consumption, communication, and beyond (Kitchin, 2014a). As a result, society is witnessing what some have called a "data revolution," a period marked by the widespread embrace of data-driven processes and promises.

This revolution is not occurring in a sterile environment; rather, it reflects historical moments of revolutionary acts in the organization and interpretation of information, as Kitchin (2014b) indicates. Much like the paradigm shifts triggered by the invention of the printing press, today's explosion of data produces new modes of knowledge production and understanding, driven by both the capabilities and the challenges of processing massive volumes of information. Among these shifts, the culture of what Kitchin (2014b) calls "data boosterism" has emerged, a widespread curiosity for the potential of big, open, and even small-scale data to transform decision-making and innovation. While some of this optimism is well-founded, much of it also borders on hype, raising critical questions about how data is framed, interpreted, and deployed in different fields.

Since around 2011, the concept of Big Data has started to gain more attention than most other areas of computer science (Ward & Barker, 2013). This interest has extended beyond academia into the public and media world. Big Data comes not only with storage and management issues, but it also includes socio-technical issues such as ethics and privacy. The term Big Data originated from across multiple disciplines. Thus, Ward &

Barker (2013) discuss how it can lead to contradictory interpretations as the early definitions vary widely. Further, they point out that Big Data is mainly linked to two core ideas: *data storage* and *data analysis*. These concepts already have their history in computer science and information technologies, but the question of how Big Data is fundamentally different from traditional techniques lies in its name. Ward & Barker (2013) suggest that the challenge of the quantification of Big Data contributes to the difficulty in the universally accepted definition for Big Data.

Currently, an almost incomprehensible volume of data is generated daily, with current estimates indicating that approximately 402.74 million terabytes are created each day (Duarte, 2025). Here, "created" refers to data that is newly generated, captured, copied, or consumed. This daily output contributes to a projected annual total of 147 zettabytes in 2024, which is expected to rise to 181 zettabytes by 2025. The growth trajectory has been exponential, with global data creation increasing 74-fold since 2010, when only 2 zettabytes were produced. Notably, it is estimated that 90% of the world's data has been generated in just the past two years. A significant portion of this data comes from video content of all global internet traffic. Social media platforms heavily contribute to this figure, as video sharing becomes increasingly central to user activity. Supporting this surge in data creation, the United States alone houses over 2,700 data centers to manage the growing demand for storage and computational power.

3.1.1 Definitions and Characteristics of Big Data

One of the earliest uses of the term Big Data in literature belongs to Cox and Ellsworth (1997), which refer to the large datasets in terms of their storage, management, and visualization attributes. Diebold (2012) contributes to the origins of the term Big Data by mentioning that before 2000, some academic and non-academic references to "Big Data" existed but were generally superficial, lacking a deep understanding of the phenomenon. Academics recognized the challenges of large-scale data but rarely used the term itself. Notably, in the mid-1990s, Silicon Graphics Inc. (SGI) actively engaged with the Big Data concept, led by John Mashey. Mashey produced a 1998 slide deck titled "Big Data and the Next Wave of InfraStress," explicitly acknowledging Big Data's significance. SGI also used the term in advertising from 1996 onward in Black Enterprise, Info World, and CIO magazines. This shows that Mashey and SGI were early pioneers, promoting Big Data as both a technical seminar theme and marketing concept, well ahead of broader adoption in academia and industry, thus crediting them with early informed use and conceptualization of Big Data. Diebold continues that after the early industry use of "Big Data" by Mashey and colleagues, there was at least one significant pre-2000 academic reference to the term within computer science. Unlike industry pioneers, this reference came from the academic community. Weiss and Indurkhya (1998) not only used the term "Big Data" but also showed a clear understanding of the phenomenon.

Another notable definition of Big Data is made by Alan Laney (2001) as an organizational report. Even though the report itself did not mention the term ‘Big Data’ directly at that time, it is seen as one of the key definitions currently. Laney (2001) implicitly refers to Big Data through the concepts of massive data volumes, high velocity, and diverse data formats, emphasizing the challenges and strategies for managing such data. While it does not provide a formal, precise definition of "big data," it describes the phenomenon as an explosion of data (due to e-commerce, mergers, partnerships, etc.) characterized by increasing **volume**, **velocity**, and **variety** that exceed traditional data management capabilities. Volume refers to the amount of data generated and stored, which increases with the depth and breadth of data collected from transactions and interactions, velocity represents the speed at which data is created, captured, and needs to be processed, especially in real-time or near-real-time scenarios like e-commerce interactions and lastly variety describes the different formats, structures, and semantics of data, leading to challenges in integrating and managing incompatible data sources (Laney, 2001). Even though the first aspect often associated with the term 'big data' is its size, Gandomi and Haider (2015) remark that additional characteristics have gained recognition in recent discussions around the mid-2000s. As an organizational report, Oracle’s (2019) definition of Big Data is also defined by the three V's: Volume, Velocity, and Variety, encompassing large, fast, and diverse data from sources like social media and sensors. This complexity challenges traditional methods, requiring technologies such as Hadoop and NoSQL for effective management and analytics. They mention the challenges posed by Big Data as the need for organizations to shift from familiar data integration tools to new technologies, impacting productivity and increasing costs due to the required new skills.

NIST Big Data Public Working Group’s (2019) report defines Big Data as datasets that require a scalable architecture for efficient storage, manipulation, and analysis due to their extensive size and complexity. Specifically, Big Data is characterized by four main attributes, often referred to as the ‘Vs’. Fourth ‘V’, **veracity**, introduced by IBM to address the uncertain and imprecise facets of data (Gandomi & Haider, 2015), which the original three Vs (volume, velocity, and variety) did not fully capture. Additionally, Gandomi and Haider (2015) mention that the concept of big data's "Three V's", Volume, Variety, and Velocity, lacks universal benchmarks, as definitions vary based on a firm's size, sector, and location and evolve over time. These dimensions are interrelated; a change in one often affects others.

Big Data is generated from a vast array of sources, which SAS introduced **variability (and complexity)** as another key dimension, referring to the inconsistent flow of data over time. Rather than maintaining a steady pace, big data often experiences fluctuations, with intermittent spikes and drops. **Complexity**, as another key dimension, involves connecting, aligning, cleaning, and transforming diverse sets of data, and originally, Oracle introduced **value** or **low-value density** (Gandomi & Haider, 2015). But also, in IBM's (2024) web page about Big Data, another characteristic or 'V', is mentioned: **Value**. Value in big data refers to the practical benefits it offers, such as improving operations and uncovering new opportunities, by using analytics, AI, and machine learning to turn raw data into useful insights.

Big data's wide-ranging applications across different settings, disciplines, and uses put the concept at risk of becoming vague and overly broad, essentially *everything and nothing* (Simsek et al., 2019). Popular definitions that focus on characteristics such as volume and variety have introduced uncertainty about what qualifies as big data. Therefore, Simsek et al. (2019) define Big Data as a classification that refers to the processes of generating, organizing, storing, retrieving, analyzing, and visualizing large, diverse datasets. Its scale and variety introduce new methodological challenges in data handling, epistemological questions about knowledge and interpretation, and politico-ethical issues concerning privacy, transparency, and fairness. Thus, big data involves complex technical, philosophical, and ethical considerations beyond just volume.

Favaretto et al. (2020) conducted a study to explore how researchers in psychology and sociology interpret the concept of Big Data, assessing whether existing definitions are seen as sufficient and examining the possibility of establishing a standardized, discipline-specific definition. They found that although no single agreed-upon definition of Big Data appeared, there were important overlaps. Most researchers describe Big Data as huge amounts of digital data generated from technological devices that require specific algorithmic or computational methods to address research questions. Many researchers found the traditional "Vs" definitions (such as volume, variety, velocity) insufficient and preferred more practical, usage-based definitions that focus on data collection and processing aspects. They conclude that Big Data is a dynamic, interdisciplinary phenomenon rather than a single, fixed definition. This perspective may help address some of the challenges that come with Big Data's conceptual vagueness and its role as a hype term in academia.

Although big data has been defined in various ways, there is no single, specific definition that encompasses all aspects. Many definitions focus on what big data does, its applications and capabilities, while relatively few concentrate on what big data fundamentally is (Bhadani & Jothimani, 2017). They further add that as technology advances, especially in storage and processing capabilities, the data considered *big* today may no longer be seen as such in the future. Therefore, the definition of Big Data is not fixed but evolves alongside technical progress. Beyond just defining Big Data, it is crucial to understand how to make the best use of this data. Extracting valuable information from Big Data to support informed decision-making is where its real significance lies. In other words, the value of Big Data is realized not just by managing its volume and complexity but by leveraging it to generate meaningful insights that drive better decisions.

3.1.2 Evolution of Big Data

Scholars discuss that Big Data caught many off guard. Gandomi and Haider (2015) contribute this argument, mentioning that normally, new technologies would first be introduced through technical and academic literature, gradually spreading to broader knowledge platforms like books. However, the swift advancement of big data technologies and their quick adoption by both public and private sectors allowed limited time for thorough academic discussion and conceptual development. They caution about dangers like; false or misleading relationships that appear in big data simply due to the vast amount of information and numerous variables analyzed. These misleading correlations have not been thoroughly examined or addressed in prior research. Most discussions tend to focus on identifying correlations but often overlook the deeper and more complex issue of causation, like understanding whether and how one factor causes changes in another.

The evolution of big data began in the mid-1990s with manageable datasets and traditional methods, according to Badman and Kosinski in their (2024) IBM report. They further explain that the internet’s growth and connected devices caused an explosion of diverse data sources. Early innovations like Hadoop introduced an open-source framework that stores data across many servers and allows processing in parallel, making it easier to manage large datasets cost-effectively. Apache Spark, another key technology, improves on this by performing data analysis directly in memory, which speeds up processing significantly. Data lakes serve as large storage repositories that keep raw data in its native format, accommodating both structured and unstructured data. Data lakes are ideal for high-volume and varied data, supporting advanced tasks like AI and machine learning. Cloud computing further democratized access by providing scalable, affordable storage and processing resources. Together, these technological shifts transformed big data into a dynamic ecosystem that turns massive, complex datasets into actionable insights, driving innovation across industries.

Gitelman (2013) traces the evolution of big data as a complex and historically rooted phenomenon that reflects broader shifts in how knowledge is produced, classified, and understood. Further emphasizes that what is often called "raw data" is inherently processed and "cooked", challenging the common assumption of data transparency and purity. The emergence of big data is linked to a historical move from traditional knowledge systems, such as paper archives and classification schemes, to digital and computational systems deeply embedded in the infrastructures of disciplines and institutions.

Furthermore, Gitelman (2013) suggests that the rapid growth and scale of data in the contemporary era, characterized as the "Age of Big Data", are marked by an accelerated history comparable to past technological revolutions, like the rise of industrial aquaculture or the development of cybernetics. These shifts are tied to the increasing complexity of data, Gitelman (2013) adds, which often surpasses human understanding, especially in scientific fields like physics and climate science. Overall, Big Data is portrayed not merely as an abundance of information but as a deeply embedded social and epistemic practice, shaped by historical, cultural, and infrastructural factors that influence how data is classified, interpreted, and used. As another important point, Gitelman (2013) emphasizes that accumulating more data is insufficient for effectively managing the planet and society. Instead, it calls for a humanistic, interpretive approach to analyzing data. This approach helps us understand the forms data takes and imagine new possible futures, preventing us from being overwhelmed by the excessive amounts of data available.

Wiggins and Jones (2023) follow Big Data to its roots and point out that extensive data collection activities characterized both the pre-World War II and post-war periods. During the war, this primarily involved military data, whereas in the following decades, the focus shifted toward the accumulation of personal data by various institutions. By the 1970s, the accumulation of data had begun to have far-reaching effects on individuals, impacts that extended well beyond concerns of privacy alone. More than privacy, this matter brought forward essential concerns about who holds the authority to make decisions based on data, how those decisions are made, and who has the right or ability to challenge them (Wiggins & Jones, 2023). In the 1980s and 1990s, privacy concerns were largely set aside. It is safe to state that privacy concerns have emerged intermittently over the decades, fading and resurfacing at different times, with a notable reemergence in the 2010s as discussions increasingly linked privacy to broader issues of justice.

When discussing the emergence of the first commercialized digital computers or machines, Wiggins and Jones (2023) highlight a significant detail: regardless of possible ethical concerns, much of the computing innovation during that period stemmed from the deeply interconnected efforts of commercial and military sectors, an indicator of the state-led capitalist model common to the Cold War. Due to the situation that 'Big Data needs big infrastructure', during this period, especially through military channels, the U.S. government played a major role, covering over half of computer research and development costs up to the late 1950s, and actively participated in the development process. After these technologies became commercialized, they enabled the transfer of data from formats like punched cards and prepared the way for new methods of statistical analysis, as well as the collection of new types of data (Wiggins & Jones, 2023).

Technological shifts in data storage and processing did not occur entirely naturally; the digitization of information required significant effort from various actors. As Wiggins and Jones (2023) emphasize, over time, digital computers not only became faster at performing calculations but, more importantly, expanded their ability to collect, process, and store data. While early efforts focused on digitizing existing information, these developments eventually reached a point where data could be captured and processed in near real time. This advancement enabled major changes in how data was used, particularly in the operation of large administrative systems such as airlines and welfare agencies.

Critically and maybe most importantly, Wiggins and Jones (2023) add, the shift that was often overlooked involves changes in institutional logics, specifically, the underlying decisions about what types of work and knowledge are valued, and how these should be transformed or preserved**.** This suggests that, beyond technical developments, there were deeper, often understated, shifts in how institutions operate and prioritize knowledge, which are just as significant but less visible than the technological costs. Wiggins and Jones (2023) also place a slight reminder about the consequences of excessive data conversion and storage. They underscore that data is not simply abstract information, but inherently material, dependent on dense physical infrastructures for its storage and security. This materiality brings with it substantial technological and logistical challenges. Additionally, the process of maintaining and managing vast amounts of data demands significant human labor, much of which remains invisible or undervalued. Together, these factors serve as a sobering reminder that the large-scale digitization and storage of data entail hidden costs, both technical and social, that are often overlooked in discussions about data's potential.

The adoption of digital systems and electronic data processing was neither smooth nor inevitable. Each shift required deliberate effort and advocacy, as there was significant resistance from both employees and management. Even in the 1980s and beyond, integrating technologies like computer modeling into industries such as finance involved institutional restructuring**.** Overall, these developments faced persistent resistance, driven by the complexity of human systems and concerns about the social implications of automation (Wiggins & Jones, 2023).

According to the paper by Simsek et al. (2019), the evolution of big data reflects a significant shift in how research is conducted and understood across disciplines. Initially, big data was characterized by its vast volume, variety, velocity, and veracity, which enabled new ways of processing and analyzing information beyond traditional methods. This transition is described as shifting from theory to data-driven discoveries, from causality to patterns, and from assessing a theory to insights born from data (Elragal & Klischewski, 2017; Kitchin, 2014a; Mayer-Schönberger & Cukier, 2014).

While early big data efforts prioritized raw data collection and processing, contemporary approaches recognize the importance of integrating theory to guide analysis and interpretation (Simsek et al., 2019). The authors emphasize that this progression allows management scholars to challenge existing theories and expand their understanding of managerial phenomena through innovative methodologies. However, they also caution that this evolution demands careful attention to research design, transparency, and the balance between data-driven discovery and theory-driven reasoning.

3.2 Management of Big Data

Big Data is operationalized through processes that begin with its sourcing and extend into its computational handling. A wide range of input channels contributes to data collection, including digital platforms, smart devices, and embedded sensors. These sources reflect the growing integration of technology into everyday systems and highlight the ongoing transition from analogue to digitized environments. As data generation increases, more sophisticated methods are required to store, manage, and make sense of this information.

The evolution of data science captures how the field has developed in response to these growing demands. Early limitations in data interpretation have given way to more robust, scalable, and performance-driven approaches. Rather than remaining a purely statistical discipline, data science now brings together knowledge from mathematics, computer science, social sciences, and engineering. This interdisciplinary foundation supports more effective strategies for analyzing real-world data and addresses the complexities of working with large, fast-moving datasets.

Analytical tools and processing frameworks serve as the backbone of Big Data's practical application. Traditional software and storage methods are no longer sufficient, prompting the adoption of distributed systems, in-memory computing, and specialized algorithms. Languages and platforms such as Python, Spark, and cloud-based services allow for the processing of diverse and high-volume data in real time. These developments not only enable better decision-making but also present new challenges in data representation, interpretation, and technical infrastructure.

3.2.1 Sources and Collection of Big Data

Big Data management encompasses several key topics, including collection, analytics, and storage (Badman & Kosinski, 2024). Data collection involves capturing large volumes of information from various sources, such as transactions, social media, and IoT devices. Storage solutions include data lakes, data warehouses, and data lakehouses, which store raw, structured, semi-structured, and unstructured data in scalable environments. Data analytics employs tools like machine learning, data mining, and statistical methods to identify patterns, trends, and insights within the data. Effective management ensures data quality, integrity, and accessibility, enabling organizations to derive valuable insights and support decision-making processes.

Bhadani and Jothimani (2017) explain the various sources contributing to the generation of big data, emphasizing how technological advancements have significantly increased the volume of data produced across different sectors. As industries convert physical records and analog information into digital formats, the amount of data available for analysis grows substantially. Advances in technology continue to accelerate the generation of this data. Additionally, the implementation of "smart" instrumentation contributes to big data generation. This finer granularity in data measurement creates a wealth of detailed data points, reflecting the broader trend that smarter, automated devices are producing large quantities of data in real time. Together, these factors, industrial digitization, technological advancements, and smart instrumentation, are key drivers in the rapid expansion of big data sources which can be seen at (Table 3.1).

The Internet of Things (IoT) has become a significant new source of data generation as an addition to social media data (Bhadani & Jothimani, 2017). IoT refers to the network of interconnected devices embedded with sensors, software, and other technologies that enable them to collect and exchange data over the internet. This data is valuable for real-time monitoring, analysis, and decision-making aimed at enhancing operational efficiency, resource management, and quality of life. Thus, IoT serves as a vital and growing contributor to big data, complementing data from traditional sources like social media by providing rich, real-world information from physical infrastructure and environments.

Table 3.1 Sources of the Big Data (based on Bhadani and Jothimani, 2017)

Sector / Domain	Types of Data Generated	Key Purposes / Use Cases
Astronomy	Star movement data, satellite signals, space weather signals	Space exploration, climate modeling, satellite tracking
Financial	Market transactions, news feeds, trading algorithm logs	Real-time trading decisions, fraud detection, financial forecasting
Healthcare	Electronic Health Records (EHR), wearables data, genomic sequences	Personalized healthcare, outbreak monitoring, predictive diagnostics
Internet of Things (IoT)	Edge sensor streams, device signals, context-aware data	Smart infrastructure, energy efficiency, predictive maintenance
Life Sciences	DNA sequences, protein structure data, lab experiment logs	Drug development, genetic research, treatment optimization
Media & Entertainment	Streaming metrics, user engagement logs, AR/VR interaction data	Audience analysis, content personalization, media strategy
Social Media	Posts, videos, sentiment tags, interaction histories	Consumer behavior analysis, trend prediction, social dynamics research
Telecommunications	Call Detail Records (CDR), app usage, geolocation data	Network optimization, churn prediction, service personalization
Transportation & Logistics	GPS, RFID data, fleet signals, transaction logs	Route planning, delivery optimization, supply chain efficiency
Video Surveillance	CCTV/IPTV feeds, facial recognition, movement detection data	Security monitoring, behavior analysis, access control
Smart Devices	Voice assistant logs, device usage patterns, AR/VR input	Adaptive interface design, smart home analytics, behavior prediction
Metaverse & Virtual Worlds	Avatar movement data, interaction signals, 3D environment logs	Virtual experience design, user engagement metrics, digital twin optimization
Agritech	Soil sensors, drone images, satellite crop data	Precision agriculture, crop yield prediction, climate risk analysis
Autonomous Systems	LIDAR scans, vehicle telemetry, object detection sequences	Navigation, autonomous decision-making, collision prevention

3.1.2 Evolution of Data Science

Until the 1980s, though storing large volumes of data became manageable, analyzing that data for meaningful insights and economic value remained a major challenge (Wiggins & Jones, 2023). During this time, there were few effective tools to extract understanding or profit from the growing data sets. From the late 1980s onward, the focus in computing shifted toward optimizing measurable tasks, such as handwriting recognition, using real-world, often messy data. This shift prioritized practical performance over abstract goals like artificial intelligence, encouraging the development of scalable, high-performing algorithms. As a result, the field moved toward performance-based competitions, which helped organize the research community but also narrowed its focus away from earlier, more conceptual ambitions.

Data science does not have a single, clear definition. Instead of defining it precisely, Wiggins and Jones (2023) focus on how the term has been debated over time. As data science gained importance in industry and academia, with new jobs, funding, and university programs, people began to argue about what it really means. These debates often compare data science to related fields like statistics, machine learning, and data mining. According to the authors, the disagreements are not just about technical differences but about who has the authority to work with data and who gets access to resources and influence. Thus, data science is described as a field shaped by social, institutional, and power-related dynamics, rather than just a fixed set of tools or methods. Data science is sometimes presented in an overly ambitious way, as a "master discipline" with the power to reshape science, business, and government. It is portrayed not just as a technical field, but as one that reorganizes knowledge and power in today's institutions.

Elaborating further, Wiggins and Jones (2023) emphasize that data science has diverse and mixed roots. It comes from advanced mathematics and academic environments, but also from practical work in marketing, politics, and engineering. This "impure" blend reflects a broader trend: the rise of automated decision-making systems powered by large-scale data infrastructures. They also highlight that data science emerged from the intersection of statistics applied to real-world data, machine learning, and data-driven business practices. A key part of its evolution includes the role of statisticians who challenged traditional academic approaches and pushed the field toward greater engagement with real-world data and practical problems.

Between the 1960s and 1990s was a transformative period in the development of data analysis as a discipline, largely inspired by statisticians like John Tukey and his colleagues at Bell Labs (Wiggins & Jones, 2023). Approaches to statistics and data have evolved from rigid, mathematically focused theory toward a more flexible, computationally empowered, application-driven discipline focused on comprehending the increasingly massive and complex data generated by the modern world. Wiggins and Jones (2023) outline the early 1990s as the emergence and growth of data mining as a distinct field in response to the rapidly expanding volumes of data generated by businesses, government, and scientific research in the late 1980s and early 1990s. In this era, data mining has matured as a field focused on the scalable extraction of patterns and knowledge from massive databases, instrumental in enabling industries and scientific domains to open into their enormous, otherwise underused, data resources.

The late 1990s saw critical breakthroughs in scaling search and data mining for the web's enormous and unstructured data. Stanford's MIDAS group, through Sergey Brin and Larry Page, developed foundational algorithms that powered Google and revolutionized how people accessed and used web information on the go. The first decade of the 2000s witnessed the evolution from traditional data mining to big data analytics powered by the explosion of web-generated data, advances in distributed computing, and the rise of the data scientist, setting the stage for modern data science practices and technologies (Wiggins & Jones, 2023).

Providing a critical perspective on AI in data science, Wiggins and Jones (2023) explain the direction artificial intelligence has taken by highlighting a significant shift in values. They point out that modern AI, especially through machine learning and data science, has become narrowly focused on optimizing measurable outcomes rather than pursuing the broader, more philosophical goal of understanding or replicating human intelligence. This shift, from curiosity-driven research to performance-based, industrial-scale data processing, is seen as contributing to today's ethical and political challenges surrounding AI. They also argue that the commercialization and industrialization of data have moved AI far away from the original vision of early pioneers like John McCarthy, who aimed to build systems that mimic human reasoning. Instead, current AI systems prioritize learning from large data sets to achieve profitable and efficient results, often neglecting deeper concerns about fairness, transparency, and meaning. Thus, the authors suggest that the evolution of AI has come at the cost of its foundational values.

Bhadani and Jothimani (2017) contribute to the discussions in data science by pointing out that the emergence of it has been led by technological advancements, combined with the challenges. Data science integrates knowledge from multiple discipline, which is depicted in (Figure 3.1), to effectively address the complexities of big data analytics. This interdisciplinary approach helps in developing comprehensive methods and models that not only analyze data but also consider human behavior, economic factors, and social dynamics.

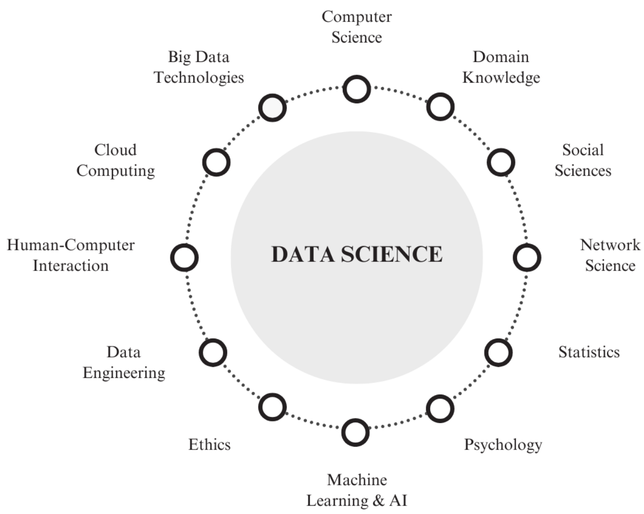

Figure 3.1 Key Disciplines and Domains Contributing to Data Science (based on Bhadani and Jothimani, 2017)

3.1.3 Big Data Analytics, Processing, and Analyzing Tools

IBM's (2024) report, called "What is Big Data?", mentions that Big Data and traditional data differ significantly in several key aspects, including the types of data, the volume they encompass, and the analytical tools necessary for their analysis. While traditional data consists of structured information stored in relational databases, it is organized into clearly defined tables that promote easy questioning using standard tools such as SQL. This format is well-suited for statistical analysis and works effectively with smaller, predictable datasets. Conversely, big data involves vast amounts of information across various formats, including structured, semi-structured, and unstructured data. Due to this diversity and scale, analyzing big data requires the employment of sophisticated techniques such as machine learning, data mining, and data visualization to uncover valuable insights. Additionally, managing and processing such enormous datasets efficiently requires the use of distributed processing systems, which enable scalable and effective handling of complex data at large volumes.

Kitchin (2014b) contrasts traditional data, scarce, static, well-structured, and collected for specific questions, with Big Data, which is massive in volume, created in real-time, diverse, and aims to be exhaustive. Traditional analysis methods suited for small, clean datasets cannot handle Big Data's complexity, velocity, and variety. Big Data requires new analytical techniques like machine learning to reveal patterns from vast, dynamic, and relational datasets. This shift challenges conventional research methods, prompting a need for new epistemologies that accommodate Big Data's scale and complexity. Further, he concludes the comparison by mentioning that Big Data analytics introduces a fundamentally new way of understanding the world, shifting from the traditional method of testing theories with selected data to uncovering insights that emerge directly from the data itself.

Based also on earlier discussions, big data analytics faces significant limitations and challenges. One of the primary concerns for researchers is ensuring security and privacy, as handling massive amounts of sensitive data raises risks regarding unauthorized access and misuse of personal information. Big data analytics has had a widespread impact across virtually all sectors of the economy and society, transforming how organizations operate and make decisions. Sectors such as telecommunications, retail, and finance, which have leveraged big data technologies by adopting earlier and gained competitive advantages and improved their services (Bhadani & Jothimani, 2017).

(Table 3.2) illustrates the broad application of big data analytics across diverse sectors, each leveraging data to innovate, optimize, and personalize their services and operations.

Table 3.2 Application Areas of Big Data Analytics (created by author)

Sector	Definition	Role of Big Data Analytics
Healthcare	Uses EMRs, clinical trials, genomics, and wearables.	Enhances diagnosis, personalized medicine, outbreak detection, treatment planning, and cost reduction. Integrates AI for robotic surgery and faster diagnostics.
Telecommunications	Involves communication metadata, customer data, and network logs.	Optimizes network performance, predictive maintenance, fraud detection, and customer experience.
Financial Services	Includes banking, investment, and fintech data from markets, transactions, and behavior.	Supports algorithmic trading, risk modeling, fraud detection, regulatory compliance, and customer segmentation.
Retail & E-commerce	Analyzes online/offline purchases, behavior data, and sensor inputs.	Enables recommendation systems, dynamic pricing, inventory forecasting, and personalized marketing.
Public Safety & Law Enforcement	Uses video surveillance, emergency calls, and social media.	Facilitates predictive policing, real-time response, crime mapping, and facial recognition.
Marketing & Advertising	Involves consumer behavior data, digital tracking, and campaign metrics.	Enables targeted ads, A/B testing, customer segmentation, real-time bidding, and sentiment analysis.
Product Design & Development	Uses market data, user feedback, and trend analytics.	Aids in identifying user needs, predicting market trends, and optimizing product features through iterative design.
Energy & Utilities	Uses data from smart meters, weather, and grid systems.	Optimizes energy use, predicts demand, prevents outages, manages renewables, and detects theft.
Insurance	Analyzes claims, customer profiles, and IoT sensor data.	Improves risk assessment, pricing personalization, claims automation, and fraud detection.
Education	Uses learning management system data and student performance metrics.	Enables adaptive learning, dropout prediction, performance analysis, and curriculum development.
Agriculture	Involves satellite, drone, soil, and weather sensor data.	Enables precision farming, crop health monitoring, yield prediction, and efficient resource use.
Manufacturing & Industry 4.0	Uses machine logs, production data, and IoT sensors.	Drives predictive maintenance, process automation, defect detection, and production planning.
Media & Entertainment	Analyzes viewer behavior, social media, and consumption patterns.	Powers content recommendation, audience analysis, and targeted advertising.
Transportation & Logistics	Includes GPS tracking, vehicle telemetry, and route planning data.	Optimizes logistics, reduces delays, improves fleet efficiency, and supports autonomous systems.
Urban Planning & Smart Cities	Uses environmental, infrastructure, and citizen feedback data.	Supports traffic control, energy management, pollution monitoring, and participatory governance.
Space & Climate Sciences	Uses satellite imagery, climate models, and scientific observations.	Enhances climate forecasting, space exploration, earth observation, and environmental monitoring.

Bhadani and Jothimani (2017) explain that organizations gain competitive advantage by analyzing vast data for informed decisions. Traditionally, Relational Database Management Systems (RDBMS) and SQL managed structured data well for daily operations. However, with the surge in data volume and especially unstructured data like Facebook reviews and tweets, these traditional methods struggle. Existing storage and query techniques cannot efficiently process such large, diverse datasets. This challenge necessitates new big data technologies designed to handle the scale and variety of modern data sources more effectively (Bhadani & Jothimani, 2017).

Gandomi and Haider (2015) explain that big data alone is not valuable unless it is effectively leveraged to support decision-making. To extract meaningful insights from the vast volumes of fast-moving and diverse data, organizations need efficient processes. This overall process of gaining insights from big data consists of five stages, grouped into two main sub-processes: data management and analytics which is illustrated in (Figure 3.2).

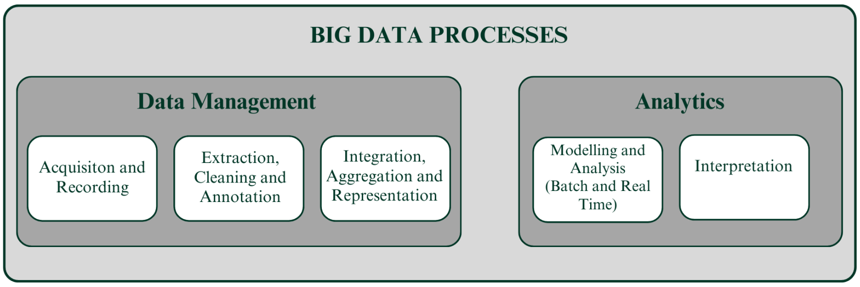

Figure 3.2 Process of Extracting Insights from Big Data (based on Gandomi and Haider, 2015)

Storing and retrieving massive structured and unstructured data quickly is challenging with traditional methods. The challenges in handling and processing vast amounts of data using traditional storage techniques led to the emergence of advanced tools (Bhadani & Jothimani, 2017). They further explain, to handle the scale of big data, this requires specialized algorithms that leverage modern computing techniques, including distributed data storage (storing data across multiple machines), in-memory computation (processing data directly in a computer's memory for speed), and cluster computing (using multiple interconnected computers to perform parallel processing).

Previously, these large-scale computations were done using grid computing (Bhadani & Jothimani, 2017), a system where users volunteer their computing resources, but more recently, cloud computing has become the dominant approach. Cloud computing offers scalable, on-demand resources over the internet, making it more flexible and accessible than traditional grid computing for big data processing.

Bhadani and Jothimani (2017) emphasize how the combination of relatively affordable high computational power and data science techniques is transforming the way big data is utilized. Advances in computing technology have made it possible to process and analyze massive datasets efficiently without excessive costs. This availability enables researchers to fully leverage Big Data. Traditionally, data was mostly stored for record-keeping or basic reporting. Now, data is actively explored using intelligent systems, such as machine learning and artificial intelligence, to uncover patterns, correlations, and insights that were not anticipated when the data was initially gathered.

Tools that collect data can encompass various types such as *smart watches, cameras, mobile devices and applications.* These applications may generate this data in different formats (Bhadani & Jothimani, 2017). This variety and volume of data, often unstructured and complex, pose challenges for traditional data processing methods.

Researchers are developing new techniques that improve the way unstructured data is represented and organized. These innovations make it possible to discover useful information that might have been overlooked before. In the context of big data, such progress is essential for unlocking valuable knowledge and supporting better decision-making. To manage the challenges posed by big data, new programming languages, computational frameworks, and data storage technologies have been developed. These tools provide specialized capabilities to manage big data's unique characteristics. Together, these innovations from an ecosystem that supports data ingestion and storage to pressing and analysis. Modern big data environments rely on a wide array of programming languages, processing frameworks, and cloud-native services to store, manage, and analyze large datasets efficiently. (Table 3.3) provides an overview of the modern data ecosystem.

R is still widely used for statistics and data science. However, in big data contexts, R is less common than Python or Spark due to its performance limitations. Python is still the most popular language for data science and big data analytics. For example, Apache Hive/ Pig is outdated in many modern systems.

Table 3.3 Categorization of Programming Tools, Frameworks, and Platforms in Modern Data Science (created by author)

Category	Tool/Language	Key Features	Key Features
Programming Language	Python	Most popular, scalable, many libraries (Ray, PySpark)	General-purpose, Big Data & Machine Learning
	R	Statistical analysis, data visualization (ggplot2), sparklyr	Research, exploratory analysis
	Scala	High-performance, used in Spark core	Backend, Spark development
Processing Framework	Apache Spark	In-memory, scalable, supports SQL, ML, streaming, Delta Lake	Batch & stream processing
	Apache Flink	Real-time stream processing, stateful apps	Event-driven systems
	Ray	Distributed Machine Learning/AI, supports Python	Machine Learning workloads, model training
Storage & Format	Delta Lake / Iceberg	ACID (key properties of database transactions) on data lakes	Lakehouse architectures
	Apache Hadoop (HDFS)	Distributed storage	Legacy systems, data lakes
	MongoDB, Cassandra	NoSQL, document-based or column-family databases	Unstructured / Semi-Structured Data
Query & Warehousing	Snowflake, BigQuery	Scalable cloud-based SQL engines	Cloud analytics, Business Intelligence, data warehousing
	Apache Hive (legacy)	SQL on Hadoop (now rarely used)	Legacy Hadoop systems
Visualization & BI	Tableau, Power BI	Interactive dashboards and analytics	Business Intelligence
	Grafana, Superset	Open-source visualization tools	Monitoring, analytics dashboards
Cloud/Platform Services	AWS EMR, SageMaker	Managed big data and Machine Learning platforms	Cloud-native data science
	Databricks	Spark-based collaborative analytics & Machine Learning	Lakehouse, team data projects
Orchestration	Apache Airflow	Workflow scheduling & DAGs (represent the order and dependencies of tasks)	ETL (Extract, Transform, Load), data pipeline management

3.3 Discussions of Big Data

Big data encompasses more than simply large quantities of information; it represents a complex ecosystem of technologies, methodologies, and processes designed to capture, store, manage, and analyze extensive volumes of diverse (Badman & Kosinski, 2024). The benefits of Big Data can be understood as a progression driven by advancements in technology and analytical capabilities. Initially, organizations relied on traditional data management systems designed for structured data. As digital technologies and data sources expanded, big data introduced new opportunities, allowing organizations to analyze large, diverse datasets in real-time. The development of distributed processing frameworks enabled more efficient handling of massive datasets. These technological improvements facilitated several critical benefits, such as improved decision-making, enhanced customer experience, increased operational efficiency, responsive product development, and dynamic pricing strategies. Over time, the integration of these innovations has transformed big data into a vital asset for organizations, enabling them to innovate, optimize operations, and gain a competitive edge across various sectors (Badman & Kosinski, 2024).

Frické (2015) explores whether Big Data enables the discovery or evaluation of universal scientific theories. Data-driven science enables the passive data collection, but it also may cause flawed statistical manipulation. The advent of computers, networks, and sensors has led to a vast amount of recorded data. Frické (2015) comments that even some commentators favor the idea that the more data there is, the more discoveries can be made, and the topic needs to be approached cautiously. Even though computers play an essential and growing role in science by rapidly generating and testing numerous hypotheses, it remains a guess-and-test method, but data-driven science is not meant to be this. All the sources that contribute to this data abundance do not provide explanations, develop theories, or solve scientific problems. Even though there is the *logic of scientific justification,* it is not supported *by logic of scientific discovery.*

The concern about present data-driven Big Data is just it might be inductivist (Frické, 2015). In this context, inductivism refers to the philosophical approach that suggests scientific knowledge can be obtained by gathering a large number of observations and deriving general theories or laws directly from these data patterns. However, this approach is flawed because mere data collection doesn't ensure valid theories; instead, scientific progress requires critical testing and falsification, not just induction from data (Popper, 1963, as cited in Frické, 2015). Big Data can work together with experimental science, but tends to favor passive observation (Frické, 2015), such as conducting surveys and making observations, which involves collecting data without active intervention.

Even though Big Data makes data collection affordable and sometimes reduces the associated risks, it also raises important ethical concerns, such as privacy (Frické, 2015). Still, it has the potential to provide stronger evidence to support scientific theories. In conclusion, data-driven science relies excessively on post hoc analysis, meaning that it seeks patterns not pre-specified, which carries the risks of false correlations that have occurred by chance rather than true effects (Frické, 2015). This might support untested hypotheses, resulting in the compromising of the validity and reliability of findings. Calling the data-driven science as *fourth paradigm* might be an illusion, as Frické points out, and adds that true science requires problems, ideas, and theories. What data-driven Big Data and current science need more theories and, arguably, less data.

Could Big Data and advanced analytics could render traditional scientific theories outdated? Popularized by Chris Anderson (2008b), this view claims that with massive data availability, empirical data can "speak for itself" through correlations and pattern detection, eliminating the need for prior hypotheses or theoretical models. It emphasizes knowledge production based on observable data rather than theory-driven science. However, this perspective is critiqued for oversimplifying science's epistemological complexity and ignoring the importance of theory for causation, context, and explanation. While provocative, it faces opposition from proponents of data-driven science that integrate theory with empirical analysis (Kitchin, 2014a).

Conversely, Kitchin (2014a) contemplates that data-driven science is a new scientific approach that combines traditional scientific methods with Big Data analytics. It integrates abductive, inductive, and deductive reasoning to generate and test hypotheses. Guided by existing theory and domain expertise, data generation and analysis are strategically directed rather than unguided mining. This approach uses machine learning and visualization techniques to handle large, complex datasets, fostering interdisciplinary research and building holistic models of complex systems. Unlike pure empiricism, data-driven science recognizes that data do not speak for themselves but require theoretical framing to produce meaningful insights. It represents a reconfigured scientific method adapted to the challenges and opportunities of Big Data.

Kitchin (2014a) also emphasizes that Big Data and new analytics are disruptive innovations reshaping research across disciplines. However, in the humanities and social sciences, they are unlikely to fully replace traditional methods due to diverse philosophical underpinnings and data limitations. Instead, Big Data expands available data and techniques, demanding urgent critical reflection on epistemological implications and the development of situated, reflexive approaches to knowledge.

Badman and Kosinski (2024) highlight Big Data challenges, including maintaining data quality, vast, diverse, and fast-streaming sources, which complicates accuracy and consistency. Scalability demands continuous expansion of storage and processing, even with cloud solutions. Privacy and security are critical due to strict regulations. Additionally, integrating varied data types, structured, semi-structured, and unstructured, requires complex technical solutions. Overall, these challenges highlight the complexities involved in effectively managing and utilizing big data for insights and decision-making.

Chapter 4

Architectural Design Process in Transformation

Architectural design is undergoing a significant transformation driven by the expanding influence of digital technologies, computational thinking, and data integration. What was once a process grounded in manual drafting, individual intuition, and relatively linear progression is now increasingly shaped by interconnected digital tools, collaborative networks, and dynamic feedback systems. This shift marks a departure from conventional methods of representation and decision-making toward new forms of reasoning and experimentation, where data and technology do not simply assist but actively participate in shaping architectural outcomes.

The evolution begins with the gradual incorporation of computer-aided design (CAD) and the later adoption of Building Information Modeling (BIM), which introduced more comprehensive, data-rich ways of conceiving, simulating, and communicating architectural ideas. As these systems matured, they began to embed environmental, material, and structural data directly into design models, allowing for predictive simulations and more informed design iterations. The emergence of general-purpose technologies like cloud computing further accelerated this shift, enabling real-time collaboration and expansive access to design data across disciplinary boundaries.

Design frameworks and creative methodologies are also being redefined. Creativity is no longer viewed solely as the generation of novel and meaningful ideas by individuals but is increasingly understood as a process distributed across teams, technologies, and datasets. Insights from psychology, neuroscience, artificial intelligence, and design studies contribute to a more nuanced understanding of how creativity manifests in architectural contexts. These interdisciplinary perspectives help to explain how design can balance aesthetic, functional, environmental, and social considerations through both intuitive and computational means.

Design thinking, parametric design, and performative approaches exemplify how contemporary architecture integrates user feedback, environmental analysis, and iterative prototyping. Representational practices evolve alongside this integration, as architects move between sketches, physical models, digital environments, and immersive simulations to externalize and evaluate ideas. These representational tools are no longer static; they are part of dynamic workflows where real-time data informs design adjustments and performance evaluations. Scanning technologies, environmental sensors, and responsive systems enable more accurate, localized, and context-aware solutions, enriching the feedback loop between environment, user, and built form.

Within this landscape, data becomes more than an input and operates as a central design medium. It supports not only analysis but also form generation, behavioral modeling, and decision-making processes. The role of the architect is consequently transformed. Architects are required to navigate complex datasets, computational platforms, and ethical challenges. Their expertise now includes managing the integrity and fidelity of information across design phases, coordinating interdisciplinary collaboration, and maintaining design intent in increasingly automated and data-saturated workflows.

Models such as Integrated Design Process (IDP), research-by-design, and actor-network-informed frameworks emphasize the importance of collaborative and non-linear methodologies. These models respond to the inherent uncertainty and complexity of design, highlighting the need for reflection, iteration, and adaptability. The introduction of tacit knowledge, intuition, and human-centered insights into structured, data-driven systems ensures that architecture remains grounded in lived experience and cultural context while embracing innovation.

As new technologies, including AI, machine learning, and big data analytics, enter the architectural domain, the boundaries between designer, tool, and user continue to blur. Data is generated by and for buildings, users, and environments, often autonomously. This flow of information supports the emergence of smart systems, responsive environments, and adaptable infrastructures. However, this transformation also raises critical questions about agency, authorship, responsibility, and ethics in design practice.

Ultimately, the contemporary architectural design process is characterized by merging, of disciplines, methods, data types, and representational forms. It reflects an ongoing negotiation between creativity and computation, tradition and innovation, structure and flexibility. As architecture adapts to the demands and possibilities of a digital and data-rich world, design becomes an evolving process of inquiry, synthesis, and transformation, responsive not only to the challenges of the present but also to the uncertainties of the future.

4.1 Foundations and Historical Evolution of the Design Process

Architectural design has been deeply shaped by evolving technological contexts and disciplinary traditions. The historical trajectory of digital transformation marks a significant shift in how architecture is conceived, communicated, and constructed. From the early adoption of tools like CAD to the more integrative systems such as BIM and cloud computing, design practices have transitioned from analog representations to data-rich environments. This shift is not only technical but also epistemological, altering how professionals engage with design information, predict outcomes, and collaborate within increasingly networked and interdisciplinary workflows.

As technology redefines tools and processes, it also reshapes how creativity is understood and operationalized in architectural design. Multiple disciplinary perspectives, from psychology and neuroscience to artificial intelligence, contribute to the development of new design frameworks and representational systems. The integration of digital and computational tools introduces algorithmic and performance-driven approaches that extend traditional notions of intuition and aesthetic judgment. Architectural creativity today is closely tied to iterative modeling, feedback loops, and the capacity to visualize and simulate complex conditions. Representational media, both traditional and digital, continue to play a vital role by enabling designers to externalize, evaluate, and refine their ideas across different phases and levels of abstraction.

The integration of data into architectural workflows brings further transformation. Data informs decisions, reveals patterns, and supports optimization, turning it into a generative and analytical force in design. It enhances collaboration between disciplines and allows for more precise, context-aware, and performance-driven outcomes. As architects work with increasingly complex and varied data sources, ranging from environmental sensors to behavioral analytics, their role expands to include data interpretation and system coordination. Ultimately, the merging of data and design fosters new forms of practice that are more adaptive, informed, and responsive to the challenges and opportunities of contemporary environments.

4.1.1 Digital Transformation in Architectural Design

It is clear that digital technology affects design, yet Cantamessa et al. (2020a) also highlight that the transformation we are witnessing is just beginning, and much deeper change is still underway. The move from traditional to digital design representations started decades ago with tools like computer-aided design (CAD) and later Product Lifecycle Management (PLM) systems, but these were only initial steps. The current shift is broader and more fundamental. This shift requires moving beyond the surface-level adoption of tools and toward a deeper understanding of how digital design redefines professional practices.

Furthermore, simply Cantamessa et al. state that " digital has changed, is changing and will change the world of design" is somewhat superficial and does not help in understanding the true implications. Such broad statements lack specificity and do not clarify what working in digital design requires for professionals, researchers, or policymakers. In other words, recognizing digital's impact is one thing, but articulating the practical meaning and challenges of digital design in everyday work and academic study requires deeper analysis. To better grasp the scope of this transformation, it is useful to look at the historical engagement of architecture with digital tools and technologies.

Architecture has long experimented with, and adapted technologies originally created for other fields, Bernstein (2018) contributes, such as engineering drafting tools evolving into CAD and gaming industry software repurposed for managing complex geometry. The transition to Building Information Modeling (BIM) marked a major turning point, shifting architecture from traditional 2D drawings to comprehensive 3D digital simulations. This paradigm shift occurred about 30 years ago, mirroring similar changes previously experienced in manufacturing and product design. This move toward BIM represents not just a technological update, but a foundational change in how design is conceived and communicated.

Initially, BIM was primarily used to enhance existing workflows by creating more accurate technical drawings. Over time, BIM has evolved into a "general-purpose technology" because it now supports all phases of building design, construction, and operation as Bernstein (2018) emphasizes. When combined with cloud computing, another broad-reaching technology, BIM enables profound changes in how design is conceived, developed, and implemented, fundamentally altering the nature of architectural practice. These developments signal a broader shift from isolated tools to interconnected digital ecosystems.

These technological advances are reducing architects' reliance on intuition and judgment by enabling predictive strategies regarding a building's final appearance, performance, and behavior when completed. As a result, the roles of architects, their methodologies, and their relationships within the delivery system of construction are undergoing significant transformation. Ultimately, architecture is beginning to "catch up" with other professions that are fully engaged in this second wave of digital innovation, shaping a new future for the discipline. One key aspect of this future is the changing role of data, not just as an output of design tools, but as a central design material itself.

Speed and Oberlander (2016) discuss a fundamental shift in computing from focusing on physical computers as devices toward emphasizing the continuous flow and processing of data. With the rise of cloud computing and pervasive internet connectivity, computing is no longer constrained to local machines. Instead, data is generated, processed, and acted upon across distributed networks involving people, machines, and things. This shift moves attention away from individual computers toward treating data as the central asset and medium of interaction, leading to new challenges and opportunities in human-data interaction.

Designers have historically incorporated data into their processes, often without explicit awareness that what they were using constituted "data." According to the discussion by Melendez et al. (2020), even before the digital era, designers relied on collected information, such as measurements, observations, or spatial relationships, to inform their decisions. However, with today's technologies, the role of data is becoming far more active, interactive, and embedded in real-time feedback systems.

Quinones-Gomez (2021) discusses how today’s advanced technologies play a crucial role in enhancing creativity, especially by fostering collaboration between humans and machines. Gomez explains that modern designed objects are intelligent devices with embedded computer technology, often connected to the Internet. This connectivity makes objects and people identifiable, locatable, addressable, and controllable, helping users complete tasks efficiently. These devices generate vast amounts of data, transforming how people process information, behave, and socialize. Access to relevant, real-time data from these products enables designers to gain a holistic, accurate understanding of the usage environment and user experience. This insight supports improved, user-centered design decisions that better address real-world needs and contexts.

4.1.2 Design Frameworks and Representations

The architectural design process is deeply intertwined with how creativity is conceptualized, structured, and represented. As digital tools, data, and computational techniques increasingly shape design thinking, it becomes essential to revisit foundational perspectives on creativity and understand how they inform emerging frameworks and representational practices. This chapter begins by exploring how creativity has been defined and studied across various disciplines, setting the conceptual groundwork for examining how design methodologies evolve and are expressed through representational tools in both traditional and digital contexts.

The literature on creativity, its definitions, and methodologies are too broad. The key definitions mostly stem from psychology. In psychology, it is commonly defined as the ability to produce work that is both novel and appropriate (Sternberg & Lubart, 1999). Guilford (1950) emphasized creative thinking as involving fluency, flexibility, and originality. Philosophical approaches tend to focus on the nature and value of originality, with creativity often described as the generation of ideas that are both new and meaningful (Cropley, 2015), and, as Schopenhauer suggested, linked to the capacity for independent and insightful thought. Art and design disciplines typically frame creativity as the ability to generate ideas that are both aesthetically and conceptually significant (Runco & Jaeger, 2012). In business and management, creativity is understood as the foundation of innovation, defined as the production of novel and useful ideas within a group or organizational setting (Amabile, 2000; Bilton & Cummings, 2010). Neuroscience approaches examine creativity through the lens of brain activity, identifying it as a process emerging from the interaction of neural networks associated with imagination, attention, and cognitive control (Beaty et al., 2016). Additionally, computational fields such as artificial intelligence define creativity as the autonomous generation of outputs that are considered creative when compared to human standards (Boden, 2004). These diverse perspectives collectively highlight the central role of novelty and value in defining creativity, while also reflecting the unique priorities of each disciplinary context.

Architectural design process depictions are shaped by both design thinking methodologies and other long-standing architectural traditions. While design thinking which is popularized by disciplines such as product design and formalized through frameworks like empathize, define, ideate, prototype, and test (Brown, 2019; Cross, 2016) has influenced contemporary architectural workflows, especially in early conceptual phases and participatory or user-centered design (Bjögvinsson et al., 2012; Lindberg et al., 2011), it represents just one layer of influence. Architectural processes are historically rooted in frameworks such as the Vitruvian triad, which emphasize the integration of structure, function, and beauty (Vitruvius, 1960). Additionally, the studio-based model prevalent in architectural education fosters intuitive and iterative thinking through sketching, spatial modeling, and reflection-in-action (Kvan & Jia, 2005; Schön, 2017). In recent decades, the emergence of computational, parametric, and performance-based design approaches has further transformed process depictions, introducing algorithmic, data-driven, and environmental simulation workflows that prioritize optimization and material logic (Burry, 2011; Kolarevic, 2005; Oxman, 2008). These approaches reflect technical and performative considerations that move beyond human-centered ideation. The broader architectural tradition encompasses diverse, and hybrid influences that make its design process depictions far more complex and multifaceted.

In the review of performative computational architecture in design process, Ekici et al. (2019) emphasize that architectural design is complex due to multiple, often conflicting objectives like sustainability, cost, functionality, and structural needs. Each project is unique, shaped by specific problems, client expectations, regulations, and environmental context, involving numerous design parameters. Architects hold significant responsibility, creating environments for humans and other living beings, making thoughtful, consequence-aware decisions essential. This complexity necessitates iterative, multi-objective approaches to balance diverse factors and optimize building performance effectively.

The architectural design process is described as iterative (Lawson, 2006), typically divided into phases like consultancy, design, construction, and operation. Traditionally, architects complete schematic design before handing over to engineers, making engineers' roles reactive rather than initiative-taking. This division is criticized by Biskjaer et al. (2021) as a blind spot that hinders sustainable, integrated design. They advocate for a collaborative, interdisciplinary approach where architects and engineers work together continuously from the start. Such integration is essential for the successful implementation of the Integrated Design Process (IDP) and to address complex sustainability challenges in building design. The Integrated Design Process (IDP) in contemporary building design is a holistic, architecture-oriented method fostering continuous interdisciplinary collaboration between architects and engineers (Biskjaer et al., 2021). This promotes dialogue and integration of diverse expertise to achieve sustainable, innovative, and well-coordinated building solutions.

Todella (2023) mentions that the understanding of the architectural design process has evolved from viewing it as a linear and cause-effect sequence, primarily centered on the final product, to recognizing it as a complex, dynamic, and uncertain activity involving numerous actors and variables. Early models conceptualized design as a straightforward progression from idea to construction, neglecting the social negotiations, revisions, and interactions that shape real projects. Recent scholarship highlights the importance of “unpacking” these internal mechanisms, emphasizing practices informed by Actor-Network Theory (ANT), ethnographies, and problem-structuring method (Jacobs & Merriman, 2011; Latour & Yaneva, 2008; Yaneva, 2009). These approaches reveal that urban and architectural transformations involve negotiations, proposals, and revisions characterized by non-linear directions, with decision-making influenced by social, physical, and political constraints. Contemporary perspectives support especially analyzing and supporting these processes through continuous evaluation, making design mechanisms more transparent, participatory, and adaptable to change. Overall, modern theories stress that architectural design is a non-linear, relational, and reflective process that benefits significantly from analytical tools and methodologies tailored to its complexity.

Researchers working in the area of design education critically evaluate the design process by analyzing its phases, methodologies, and teaching strategies, often highlighting areas for improvement and the complexities involved. Several studies emphasize that the design process is inherently non-linear, involving iterative movements back and forth among different phases such as problem-seeking, analysis, generation, testing, and reflection (Rahbarianyazd & Nia, 2019; Soliman, 2017).

According to Soliman (2017), all design processes, including architectural design, follow a general schematic structure composed of distinct phases. However, design is not strictly a linear process where one phase fully completes before the next begins. Instead, even though one design phase is the foundation for the next, designers often move back and forth between different phases as illustrated in (Figure 4.1). This iterative movement allows them to continuously review, modify, and enhance design information and ideas, responding to feedback and new insights throughout the process, which helps improve the overall design solution.

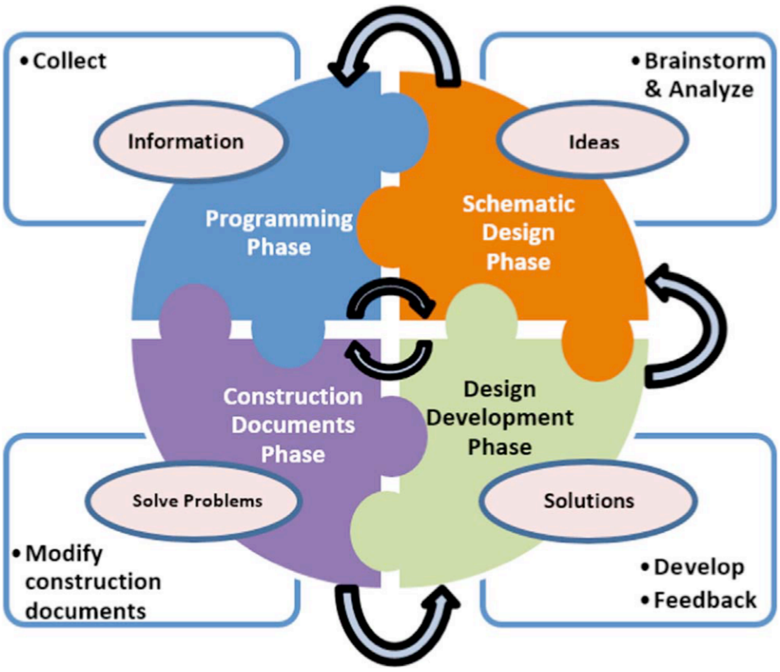

Figure 4.1 Phases of the Design Process (Soliman, 2017)

In the model, Soliman (2017) defines the architectural design process as consisting of four major phases: Programming Phase, Schematic Design Phase, Design Development Phase, and Construction Documents Phase. The process begins with the programming phase, where designers explore the context and users through tools like field visits, interviews, and observations. This sets the foundation for the schematic design phase, during which ideas are developed and analyzed using techniques such as brainstorming. Each phase builds upon the previous one, facilitating a structured approach to evolving the design from initial exploration to detailed documentation.

Rahbarianyazd and Nia (2019) illustrate the architectural design process, in (Figure 4.2), as consisting of three core stages: analysis, synthesis, and evaluation. Analysis involves understanding and breaking down the design problem. Synthesis is the creative integration of ideas to develop design solutions. Evaluation critically assesses these solutions for effectiveness and quality. This process encourages critical thinking and revision, enabling architects to iteratively refine their work from problem comprehension to final design expression.

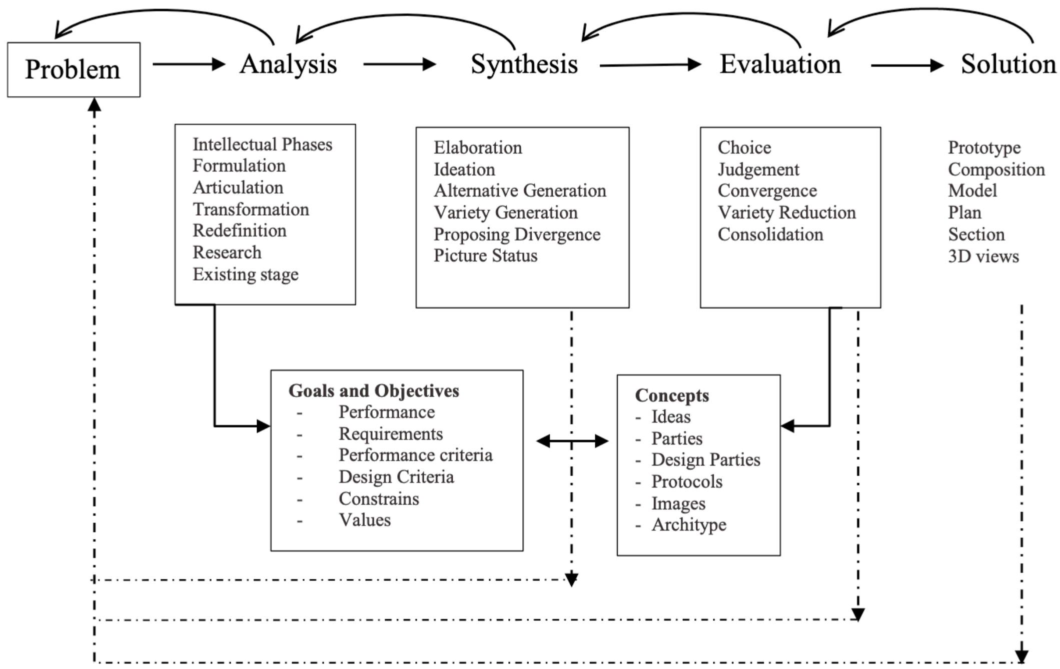

Figure 4.2 Design Process of Architectural Practice (Rahbarianyazd & Nia, 2019)

Uzunkaya and Kahvecioğlu (2022) draw attention to the nature of discipline's practice itself can serve as a research method. This highlights the fact that the main distinguishing feature in the research part and the features that give originality to the research process are the inherent characteristics of the architectural discipline. This inquiry may help us to customize our research while staying within the structure but being able to draw original outputs at the same time. While we are talking about a structure or a methodology, it does not necessarily mean that conclusions may come out as rather onedimensional. What makes the difference is whether we can include the essence of the area we are working on or not.

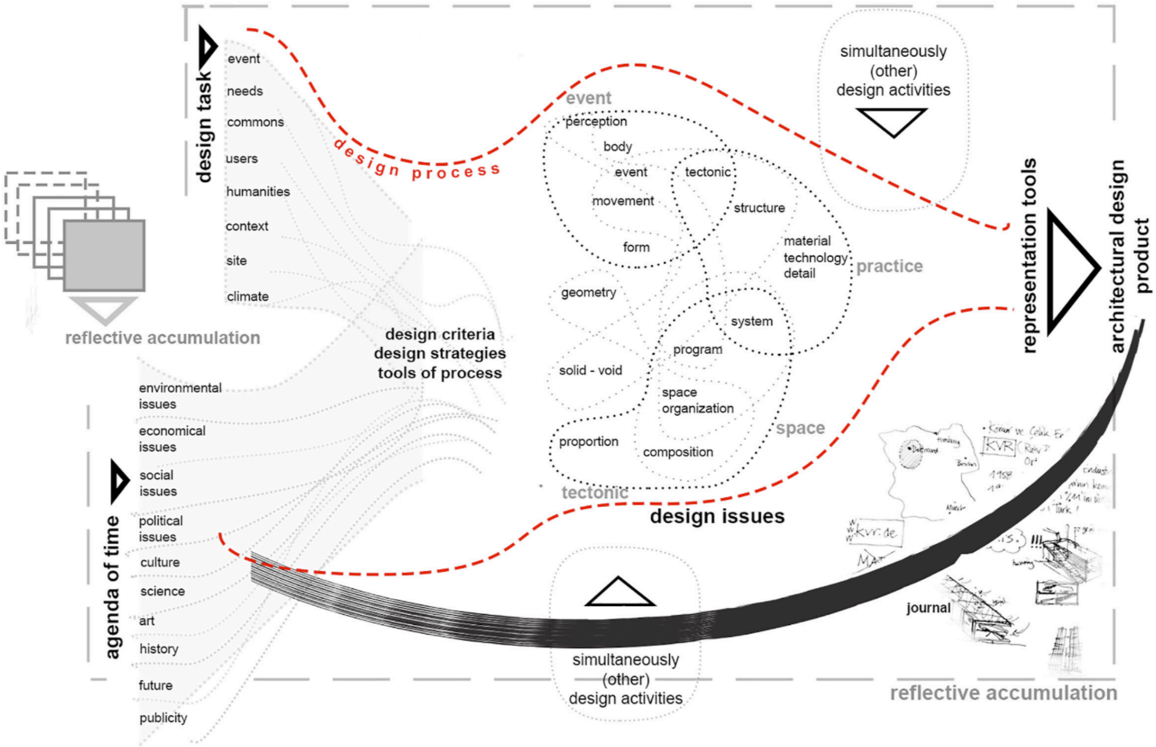

Figure 4.3 Subjects, Tools and Processes of Architectural Design (Uzunkaya & Paker Kahvecioğlu, 2022)

In their scholarly article, Uzunkaya and Paker Kahvecioğlu (2022) propose their conceptual framework in (Figure 4.3), regarding ‘research by design’ under the title of ‘architectural design research through reflection’. In their discourse, they propose an innovative sub-approach by defining the parameters of research by design, seeking to uncover tacit knowledge. In this effort, the reflections articulated by designers serve as the primary instrument for the extraction of this knowledge (Uzunkaya & Kahvecioğlu, 2022). A particularly intriguing and appropriate aspect for their investigation is the emphasis on tacit knowledge, which is embraced through the notion of the intrinsic essence of design; the individuality, which is referred to as originality, emerges from a foundation perceived as intangible. Consequently, by deconstructing the architectural design process (ADP) and integrating data methodologies along with big data strategies into the framework, it is anticipated that tacit knowledge may be enlightened.

Focusing on a humanistic perspective, Hettithanthri et al. (2023) points out that many scholars have studied the architectural design process, focusing on technical steps and procedures. However, there is a significant gap in addressing the involvement of actual users and the specific context throughout the design phases. This lack of user and context integration means that designs often fail to fully respond to human needs and environmental factors. Their study highlights the importance of incorporating user perspectives and contextual understanding to create more empathetic, realistic, and human-friendly architectural solutions.

Hettithanthri et al. further explain that the architectural design process begins with vague or underdeveloped emotions and intuitive feelings that are not yet fully formed or clearly expressed. Through the design process, these initial, often unclear ideas are shaped and refined. By applying a structured methodological framework, architects repeatedly revisit and improve their designs through iterative refinements, gradually transforming these early feelings and concepts into concrete, tangible design outcomes that can be realized in the real world.

Compared to similar design processes, Abdelhameed (2017) mentions that architectural design is more complex because it involves balancing diverse influential factors from various disciplines, including human needs, functionality, environment, structure, economy, and aesthetics. These factors can be observed in (Table 4.1). Architects face conflicting requirements and must make decisions by compromising among these factors. Due to this complexity and the nature of design problems, which often lack perfect solutions, architects adopt a bounded rationality approach, settling for "satisfactory" solutions that sufficiently meet essential needs and constraints. This process reflects the practical and multifaceted challenges unique to architectural decision-making.

Table 4.1 Factors Involved in Architectural Design (created by author)

Category	Factor	Disciplinary Domain	Description
Human-Centered	Human Needs	Psychology, Sociology	Includes comfort, accessibility, cultural relevance, and social behavior.
	User Experience	Cognitive Science, UX Design	How users interact with and perceive the built environment.
Functional	Functionality	Architecture, Engineering	Ensuring the building effectively serves its intended purpose.
	Flexibility & Adaptability	Design Strategy	Ability of space to accommodate change over time.

Environmental	Environmental Conditions	Environmental Science, Sustainability	Site conditions, climate responsiveness, ecological integration.
	Energy Efficiency	Building Science & Sustainability	Minimizing energy use through passive and active systems.
Technical	Structural Integrity	Structural Engineering	Ensuring stability and safety of construction.
	Technological Integration	Building Technology, IT	Incorporating smart systems and digital tools.
Economic	Models collective behavior to form spatial solutions.	Emergent circulation or clustering.	Concept Design, Form Generation
	Rule-based branching growth systems.	Tree-like pavilions or structures.	Concept Design
Aesthetic	Aesthetics	Art, Architecture	Visual and sensory appeal, form, materiality and expression
	Identity & Symbolism	Cultural Studies, Semiotics	Expression of meaning, context, and cultural narratives.
Regulatory	Codes & Regulations	Law, Urban Planning	Legal compliance, zoning and safety standarts
Contextual	Urban & Cultural Context	Urban Design & Anthropology	Harmony with surrounding environment, history and cultural fabric
Ethical / Social	Equity & Inclusivity	Social Sciences & Ethics	Designing for diverse users and promoting social justice.

Melendez et al. (2020) explores the evolution and significance of visualizing data within the design and architectural design process. They emphasize the importance of interpreting complex information through graphical representations, highlight methods pioneered by Edward Tufte for creating clear and meaningful visualizations. Further, they trace the historical development from early military applications, such as radar data visualization in the 1950s, to technological innovations like CAD, virtual reality, and immersive environments. This illustrates that effective data visualization transforms large, often invisible information into insights that inform innovative design solutions.

During the various phases of the design process, architects rely on representational media, such as drawings, models, sketches, or virtual environments, to conduct their design tasks (Abdelhameed, 2017). These media serve as tools to externalize and visualize their mental concepts, allowing them to explore, refine, and develop their design ideas more effectively. Representational environments help architects define the design problem by making abstract ideas tangible and manageable and also support evaluation by enabling architects to assess and modify design solutions as illustrated in (Table 4.2). Essentially, these media bridge the gap between internal imagination and practical design development.

Table 4.2 Media Types Used in both Traditional and Digital Workflows (created by author)

Representational Medium	Phase(s) Used	Function in Design Process
Sketches (Hand-drawn or Digital)	Conceptual Design, Early Stages	Externalize initial ideas, stimulate creativity, quick exploration of concepts.
Technical Drawings (Plans, Sections, Elevations)	Schematic to Construction Design	Communicate spatial, structural, and functional information; precise documentation.
Physical Models	Conceptual, Development, and Presentation	Tangible 3D representation to explore form, scale, and spatial relationships.
3D Digital Models (e.g., CAD/BIM)	Development, Technical Design	Detailed modeling, coordination with consultants, simulation, and analysis.
Diagrams	All Phases	Simplify and explain relationships, functions, or flows in the design.
Renderings (Static or Animated)	Design Development, Presentation	Visualize materiality, lighting, and atmosphere; used for client communication.
Virtual Reality (VR)	Design Development, Evaluation	Immersive environment for spatial experience and user interaction testing.
Augmented Reality (AR)	Design Evaluation, Construction Coordination	Overlay digital data on physical context; assist in on-site decision-making.
Simulation Tools	Design Development, Evaluation	Analyze performance (e.g., daylight, energy use, airflow) to inform design decisions.
Parametric/Algorithmic Models	Conceptual to Detailed Design	Generate and evaluate complex geometries or data-driven forms through computational rules.
Collaborative Platforms (e.g., BIM 360, Miro)	All Phases	Enable real-time design collaboration and information sharing among stakeholders.

Data collection in design involves the process by which architects and designers gather information to better understand environments, human behavior, and contextual factors to inform their design decisions, as explained by Melendez et al. (2020). This information can be obtained through a range of methods, from traditional analog approaches such as direct observation and surveys, to advanced technological means including satellites, sensors, and scanning technologies.

Melendez et al. (2020) further explains that public databases provide large volumes of existing data, such as weather files, which can aid in designing within specific climatic contexts globally. Locally, environmental sensors capture real-time data on parameters like light, sound, temperature, and movement, which can then be integrated into computational platforms. This integration allows for responsive and interactive systems where sensed data inform actuators (like motors, lights, valves), creating closed feedback loops that regulate environments dynamically.

Additionally, scanning technologies such as LIDAR produce detailed digital models (point clouds) of existing structures, landscapes, and urban environments, offering precise data for renovation, analysis, or design interventions. Photogrammetry is another method to digitize physical models or environments for similar purposes. Together, these approaches empower designers to work with accurate, relevant, and timely information, ultimately enhancing their ability to create spatial solutions grounded in real-world conditions (Melendez et al., 2020).

Abdelhameed (2017) emphasizes that cognitive processes like analysis and evaluation differ from other fields in architectural design due to architects' ability to adopt and adapt reference images from artificial and natural forms, aiding partial solutions or design directions. These partial solutions, termed "streams," come from the design problem context, external factors, subjective interpretations, media use, or personal biases. Key design phases include problem definition, concept construction, idea formation and exploration, and form development. Depiction of cognitive processes and influencing factors across architectural design phases can be observed in (Table 4.3).

Table 4.3 Cognitive Processes and Influencing Factors Across Architectural Design (created by author)

Design Phase	Primary Cognitive Processes	Influencing Factors / "Streams"	Role of Media / Reference Images
Problem Definition	Framing, Analyzing	Context of the design problem, external constraints	Reference images help define spatial or functional needs.
Concept Construction	Abstraction, Association	Natural and artificial forms, prior knowledge, subjective interpretation	Visual references inspire early design directions and metaphors.
Idea Formation & Exploration	Divergent Thinking, Synthesis	Personal biases, environmental cues, sketching media	Media enables exploration through drawing/modeling, adapting inspirations.
Form Development	Evaluation, Refinement, Decision-making	Technical requirements, performance expectations, iterative reinterpretation of prior "streams"	Models and digital tools assist in assessing, refining, and evolving form.

As a reflection of all the social, informational, and technological shifts, Bernstein (2018) elaborates that technology transforms the architectural process by reshaping representation, design reasoning, and collaboration within the building industry. This transformation is highlighted by several dimensions:

One of these dimensions is depicted by Bernstein (2018) as the change in representation. Traditionally, architectural designs were communicated through two-dimensional drawings, plans, elevations, and sections that represented in a more abstract way in the building. Technology, particularly the advent of Building Information Modeling (BIM), has shifted this paradigm from 2D drawings to rich, three-dimensional digital models that embed geometrical data alongside extensive metadata describing construction information, material properties, costs, and performance criteria. These models are not static images but dynamic repositories that can be manipulated directly, queried, and updated in real time, improving coordination and precision throughout the project lifecycle.

Design reasoning has also evolved from subjective intuition and manual calculations to data-driven, predictive processes enabled by analysis, simulation, and optimization tools. Bernstein (2018) adds that cloud computing and advancements in algorithms allow architects to simulate structural behavior, lighting, energy use, and even occupant behavior early in the design phase. These computational analyses provide quantitative feedback loops that help architects iterate designs more efficiently and optimize outcomes based on specific performance targets. Furthermore, emerging AI and machine learning tools promise adaptive and intelligent design assistance in the near future.

Additionally, digitally integrated design tools reshape the architect's role and how they collaborate with other stakeholders, engineers, builders, fabricators, and clients. With BIM and cloud platforms, project information is shared and managed collectively in a coordinated digital environment. This integrated approach facilitates synchronized workflows, reduces errors, and enables closer integration between design intent and construction execution. The traditional separation between architect (intent) and builder (execution) is shifting towards more interdependent delivery models, such as integrated project delivery, where architects assume roles that engage more directly in construction planning, procurement, and realization (Bernstein, 2018). This evolution requires new contractual frameworks, risk allocations, and compensation structures to reflect the shared value generated across disciplines.

Overall, these technological advances require the re-evaluation of design methodologies as depicted in (Figure 4.4). (Bernstein, 2018). Beyond simply making drawings or models, digital tools require architects to orchestrate vast arrays of data and computational processes, turning design into a systematic, data-informed activity. This shift extends architects' responsibilities into data management, quality control, and decision-making based on predictive analytics, fundamentally changing the value proposition of architectural practice in the digital era.

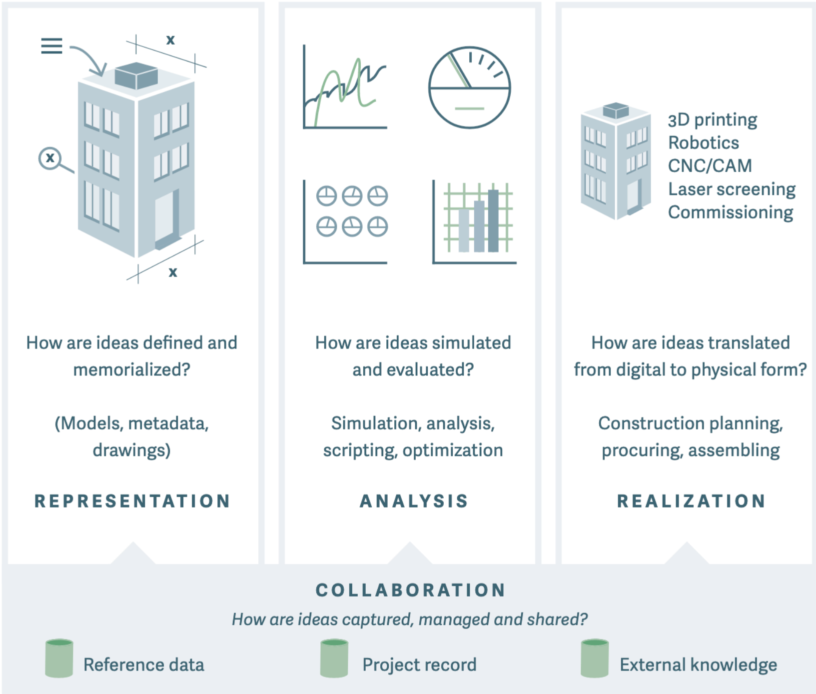

Figure 4.4 Re-evaluation of Design Methodologies Under the Influence of Technology (Bernstein, 2018)

Together, these typologies convey the integrated technological ecosystem architects engage with today: from creating and analyzing designs, through coordinating collaborative processes, to actualizing buildings via advanced construction technologies.

4.1.3 Integrating Data into the Architectural Design Workflow

Designers utilize data in their processes to improve decision-making and uncover invisible aspects of environments and behaviors. Melendez et al. (2020) highlight that data provides critical insights into complex patterns, environmental factors, and social dynamics, enabling a deeper understanding of spatial and contextual conditions. It supports iterative design by allowing validation and optimization of solutions over time. Data fosters interdisciplinary collaboration among architects, engineers, and scientists through a shared analytical framework. Additionally, by synthesizing data, designers can explore new patterns and generate innovative forms and systems. Ultimately, evidence-based data enriches design beyond intuition, resulting in more responsive, adaptive, and impactful outcomes that address real-world complexities and needs. The role of data in expanding both analytical depth and creative potential within architecture becomes evident through such applications.

Data is a fundamental platform in the design process, enabling informed decision-making and revealing aspects of the environment that are often unseen or intangible. Melendez et al. (2020) reinforces this perspective by mentioning how emerging digital technologies, such as computational design, robotics, and digital fabrication, expand the potential of data-driven processes in architecture and urban design. The increasing data integration across disciplines fosters hybridization and cross-fertilization among architects, urbanists, artists, and spatial practitioners, leading to more innovative and responsive design solutions. Melendez et al. (2020) also demonstrates how data, when combined with cutting-edge tools like robotics and physical computing, can generate new forms, systems, and environments, thus deepening each project's spatial, experiential, and performative dimensions. Together, authors underscore the transformative role of data as a generative resource that motivates collaboration and creative exploration within contemporary design practices. Cross-disciplinary integration signals a broader shift toward design cultures shaped by information exchange and computational strategies.

Focusing on the cross-disciplinary relationships between information science and architecture, Kalay (2004) highlights how computers have transformed architecture from traditional drawing techniques to detailed information systems. As information science becomes increasingly important in architectural design, the field now encompasses data management alongside visual design creation. Kalay (2004) melds information science concepts, such as data organization and presentation, with innovative design methods. He argues that architects must learn to organize and process data to create better buildings. The evolution of architectural practice now demands fluency in digital data systems as a foundational design skill.

Parametric design is also a new approach in which architects manipulate information to design flexible forms. Oxman (2017) suggests that in today's day and age, the role of an architect has moved away from drafting predefined form to acting like an information manager (organizing systems and rules). Carpo (2023) adds to the discussion by saying that architecture is deeply related to the processing of information. In the age of information science and Big Data, contemporary architecture is influenced by these concepts. Carpo (2023) also adds that ‘digital' in architecture is less related to screens and rather relates to the effects of information flows on buildings. As information science and architecture converge, the importance of value creation sustains itself but conventional design methodologies are adapting, challenging professionals with new mindframes. Professional roles in architecture continue to be redefined by growing entanglements with data, algorithms, and systems thinking.

Bernstein (2018) suggests, as design information passes through phases, from concept through bidding, building, to operation, the resolution of information increases while its reliability tends to drop due to handoffs between different teams and increasing scale and complexity as seen in (Figure 4.5). In the figure, interestingly, as project information passes through different phases and is handed over to different teams primarily via drawings, its reliability decreases at each phase transition. This drop in fidelity occurs because the scale, level of detail, and team composition change, causing knowledge and clarity to diminish despite accumulating information over time. Technology, notably digital models and integrated platforms, helps maintain the fidelity of design information across these transitions, mitigating losses in understanding and preserving design intent more effectively than paper-based communication. Digital modeling platforms play a crucial role in minimizing the degradation of design clarity across project phases.

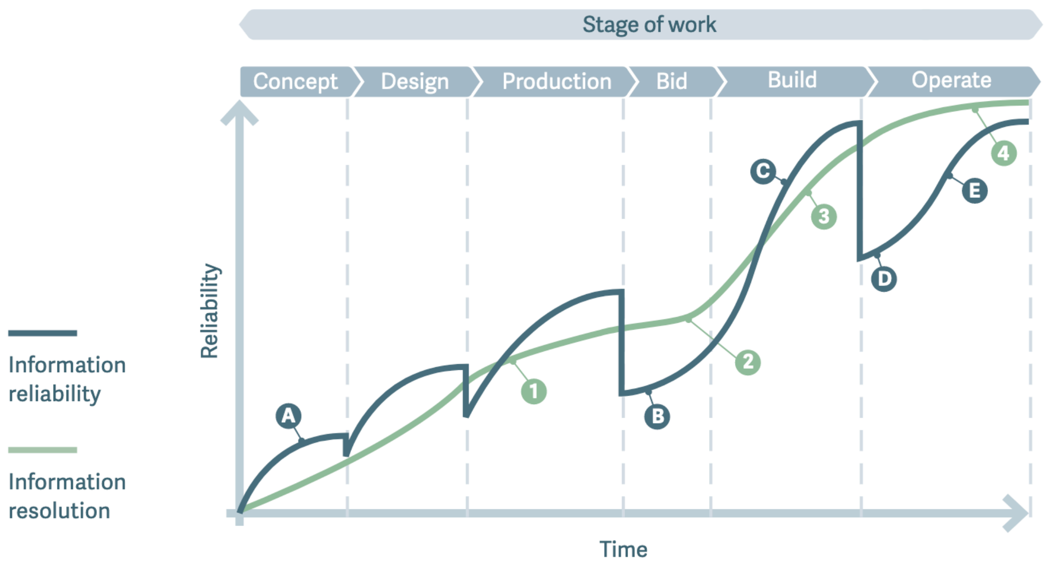

Figure 4.5 Re-evaluation of Design Methodologies Under the Influence of Technology (Bernstein, 2018)

Technological developments that eventually led to a change in the design process and tools and resulted in the accumulation of digital design data. Digital design data refers to the comprehensive collection of digital representations, models, and information related to the building's design process. This includes geometrical models (such as CAD and BIM models), analytical data (structural analyses, energy simulations), specifications, schedules, and detailed documentation generated or utilized during design development (Bernstein, 2018). The aggregation of digital content across software platforms forms an operational backbone for design and construction workflows.

These data assets are stored in various digital formats and software (like Rhino, Revit, ArchiCAD, Tekla) and can exist at different levels of resolution or detail depending on the stage of the design process (Bernstein, 2018). Because the existence of data is digital, coordination, sharing, and manipulation of data is far more convenient than traditional data. This leads to a reduction in the "execution gap", which means reducing inefficiencies and enabling the architect to better coordinate and control the construction process. Enhanced accessibility and coordination contribute to more efficient project delivery and fewer discrepancies between design and construction.

Speed and Oberlander (2016) introduce a framework to help designers understand and work with the increasing performativity of data, emphasizing three data values: raw measurements, commercial/social value, and moral/ethical value. They highlight the complexity as data intertwines with social, economic, moral, and ethical concerns. The growing field of Human-Data Interaction (HDI) is identified as a key challenge that designers must address by designing from, with, and by data. In their suggested framework, Speed and Oberlander (2016) distinguish three approaches to working with data in design. **Designing from data** uses traditional methods where designers analyze collected data to inform design decisions. **Designing with data** treats data as a dynamic, ongoing flow from networked artefacts, requiring designers to engage with continuous feedback and the social, ethical implications of human-data interactions. **Designing by data** envisions emerging scenarios where algorithms and machine learning autonomously generate new designs or products, making data itself a designer. This framework helps organize existing methods and anticipate new practices, guiding designers to adapt to data's increasing performativity and complexity within an interconnected human-machine system. While designing from data involves established methods, designing by data, where data and algorithms autonomously create designs, is still emerging. Designing with data, the middle ground, is a crucial area requiring urgent research to address the complexities of Human Data Interactions. The threefold typology introduces a conceptual structure that allows practitioners to engage more critically with machine-mediated design processes.

As architecture increasingly intersects with the digital world, data has become a central force reshaping not only tools and processes but also our understanding of space, infrastructure, and urban form. Melendez et al. (2020) highlight how the increasing use of data and shared information is fundamentally transforming design methods and the way designers approach their work. Although this transformation began years ago, it is now accelerating rapidly, creating numerous new possibilities and challenges that have not yet been fully understood or addressed. A new design character is emerging, defined not by material alone but by the informational systems that guide spatial production.

Authors explain that data itself is abstract and intangible; it exists as information without physical form. However, through processes such as evaluation, analysis, and decision-making, data gains meaning and exerts influence in various fields, including design. This means that although data may seem invisible or conceptual, like information stored in the "cloud," it has very real and tangible consequences.

For example, Melendez et al. (2020) point out that data centers, which store and process vast amounts of data in the cloud, have a significant physical presence and impact on urban environments. These facilities occupy space, consume energy, and alter city morphologies. Thus, while cloud data is often perceived as remote or virtual, its physical infrastructure manifests concretely in the built environment, making the abstract nature of data visible and influential on the shape and functions of cities. Architectural design, therefore, must increasingly account for data infrastructure as part of broader spatial and ecological systems.

Altogether, the integration of data into design reshapes the architectural profession by blending creativity, computation, and critical thinking. As data becomes an active agent in design, architects are required to rethink methods, tools, and responsibilities in light of emerging human-machine dynamics.

4.2 Big Data, AI, and Computational Design Approaches

The architectural design process is increasingly shaped by the interplay between big data, artificial intelligence, and computational approaches. Data operations and interpretation in design practice are becoming more central, as large datasets offer new possibilities for analyzing user behavior, predicting performance, and stimulating creative thinking. When used critically, data enhances design outcomes by informing decisions across all stages, from early ideation to final implementation, while simultaneously expanding the conceptual scope of what design can achieve. However, these benefits depend on the designer's ability to apply the right analytical lens, ensuring data is not just collected but meaningfully interpreted.

Data-driven approaches in the design process introduce new paradigms that move beyond intuition-based methods. These approaches rely on continuous feedback, generative algorithms, and platform-supported workflows that prioritize adaptability, user involvement, and evidence-based refinement. Computational tools such as parametric design, real-time simulations, and heuristic models are not only transforming design workflows but also requiring new forms of collaboration between architects, engineers, and data specialists. This evolution redefines the architect's role, positioning designers as coordinators of complex digital systems and decision-making environments.

Big data applications in design further extend these possibilities, especially during concept generation and early design phases. Large-scale datasets drawn from social behavior, environmental sensors, or urban infrastructure provide insights that guide more informed and responsive solutions. Big data supports creativity by identifying hidden patterns and generating new design directions through abductive reasoning and data mining techniques. At the same time, its effective use depends on ethical considerations, data quality, and the development of tools that respect human individuality and contextual nuance. When carefully integrated, big data opens new avenues for architectural exploration, bridging the gap between computational logic and creative intent.

4.2.1 Data Operations and Interpretation in Design Practice

Even though Big Data and data science have advanced rapidly and are used across many sectors, Quinones-Gomez (2021) highlights that their integration into design remains limited. Predictive analytics on large datasets forecast future scenarios, yet the designer's role and the impact of related technologies in design are unclear. Ongoing research shows that data is becoming a vital creativity tool, with larger information volumes stimulating creative thinking. Data science also offers new ways to analyze human behavior, giving valuable insights that help designers optimize timing and resources, leading to improved design outcomes.

In a similar way of interpretation, Ekici et al. (2019) highlights that although computer science techniques, such as optimization algorithms and computational methods, particularly evolutionary optimizations, offer significant benefits for architectural design, their integration into the architectural field remains inadequate. Specifically, many studies focus either on architectural concerns (e.g., building shape) or on engineering or computer science aspects (e.g., algorithmic comparisons) in isolation. This separation limits the potential advantages that could emerge from combining computer science investigations, such as algorithm development, data processing, and performance evaluation methods, with architectural design objectives and constraints. A more comprehensive integration would enable more effective and innovative solutions in performative computational architecture, enhancing design exploration and optimization outcomes.

Based on the theoretical approaches of Quinones-Gomez (2017), Speed & Oberlander (2016), and Zhu (2016), Quinones-Gomez (2021) depicts how data operates within the design process. According to (Figure 4.6), Quinones-Gomez explains a classification of data types used in the design process: quantitative data is considered "concrete data" because it consists of measurable, numerical information that can explicitly define, improve, or add value to a design through objective analysis. Qualitative data, on the other hand, is seen as "abstract data" since it involves subjective, interpretive information, such as user perceptions or theoretical insights, which helps designers develop new concepts by providing deeper understanding beyond numbers. Both types complement each other, supporting innovation and refinement in design.

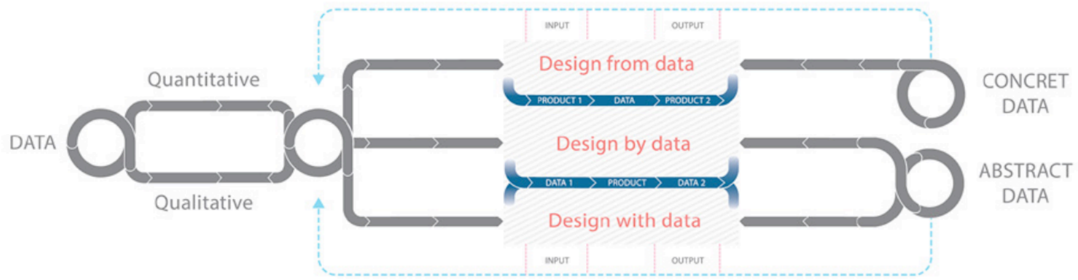

Figure 4.6 Operation of Data within the Design Process (J. C. Quinones-Gomez, 2021)

As an example of integrating data into the design process, Mortati et al. (2023) presents a model illustrating how big data and thick data are utilized differently across the four key design thinking activities in (Figure 4.7), identified by Gruber et al. (2015): observe and learn, synthesize and frame, vision and opportunity, solve and realize.

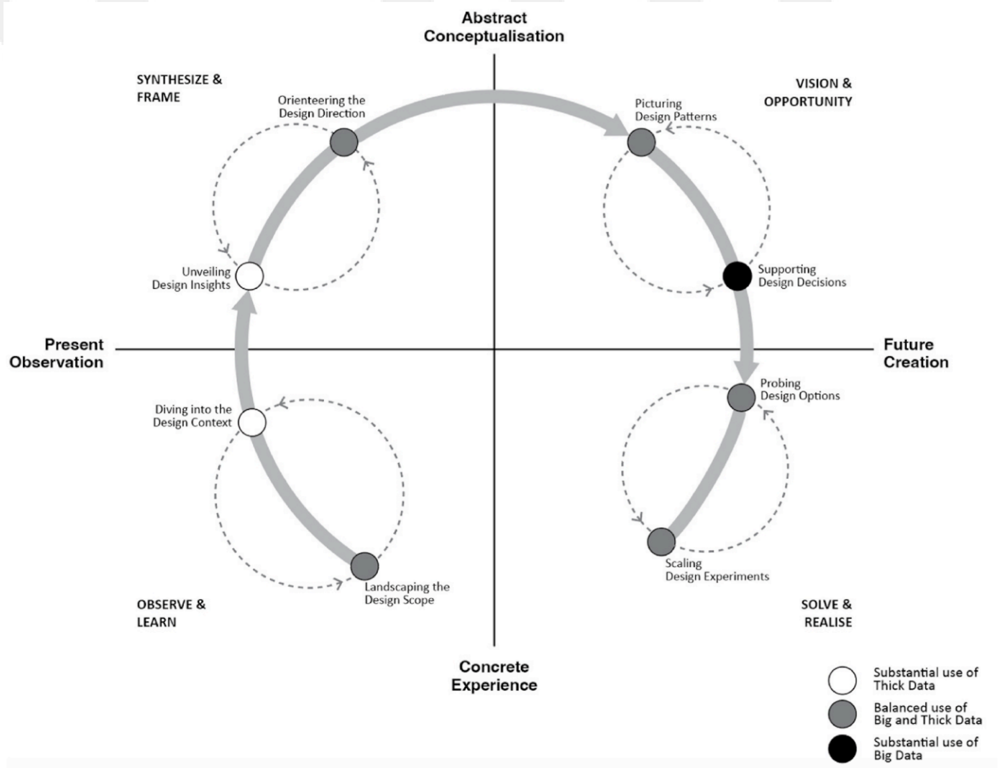

Figure 4.7 How Big and Thick Data Informs Design Thinking Projects (Mortati et al., 2023)

Mortati et al. (2023) conceptualizes a *dynamic and evolutionary* interplay between big and thick data, emphasizing that different types of data are relevant at different stages and practices within design thinking projects. It shows eight practices that utilize data variably depending on the activity, such as observation relies on thick data, validation leans on big data, pointing to a phenomenological view of data use where their role changes according to how they inform innovation activities.

On the subject of digital data and its effects in design process, Kun et al. (2018) conclude that merely having access to digital data is insufficient for effective design. Unlike flexible qualitative methods, digital data demands clear instructions, channeling creativity into focused hypothesis generation. Designers apply abductive sensemaking, using data not to prove hypotheses but as a generative tool to uncover hidden insights difficult to obtain through traditional methods. The process becomes iterative and question-driven, with continuous refinement of queries and targeted data gathering, supporting the co-evolution of problem and solution spaces. Digital data complements qualitative approaches, enriching design reasoning by revealing unseen phenomena. Overall, this leads to a data-informed yet creative and iterative design enquiry process, allowing for deeper understanding and more robust design concepts.

Melendez et al. (2020) emphasize that data, without thoughtful application or a critical perspective, are essentially meaningless. They state that the crucial aspect lies in the "lens" through which designers collect, filter, analyze, and apply data, essentially, the perspective or methodological approach used to interpret data. This lens influences how data are understood and integrated into the design process, shaping insights and guiding decisions.

For designers, the questions and opportunities that emerge from the way data are applied serve as both a source of inspiration and a means of productive inquiry. By critically engaging with data through appropriate lenses, designers can uncover new patterns, challenge assumptions, and develop innovative solutions. Thus, the meaningful use of data rests not just on the data itself, but on the interpretative frameworks employed, which can catalyze creativity and lead to impactful design outcomes (Melendez et al., 2020)

Being digital enables easy tracking of daily activities and interactions via digital interfaces, simplifying data collection. The main challenge is, Quinones-Gomez (2021) suggests, interpreting data meaningfully, as poor-quality or poorly analyzed data can cause errors. By embracing the design methodology that integrates data mining (Zhu, 2016), which systematically analyzes large datasets to find significant trends and relationships. Tools like clustering and decision trees help designers uncover hidden patterns, explore design variable relationships, and establish design rules, fostering creative breakthroughs through a data-driven design approach.

In the data-driven design model Quinones-Gomez (2021) suggests, he aims to uncover descriptive knowledge from structured data to understand past trends, and predictive knowledge from unstructured data to forecast future scenarios. Designers play a key role in transforming big data into design insights. Utilizing diverse consumer data sources enables companies to uncover hidden patterns, enhancing creativity and innovation in the design process.

Quinones-Gomez (2021) emphasizes the transformation of raw data into valuable knowledge through the application of data mining techniques and interpretation processes in (Figure 4.8). He identifies knowledge as the meaningful insights extracted from unstructured or semi-structured data, which are then integrated into the design process to inform creative decisions. The role of the designer is central in this context: designer function as the knowledge creator or integrator, interpreting the processed data to generate design insights, address problems, and inspire innovative solutions. Essentially, Gómez positions the designer as a key agent who transforms data-driven insights into design material, leveraging their perception and cognitive skills within the data transformation flow.

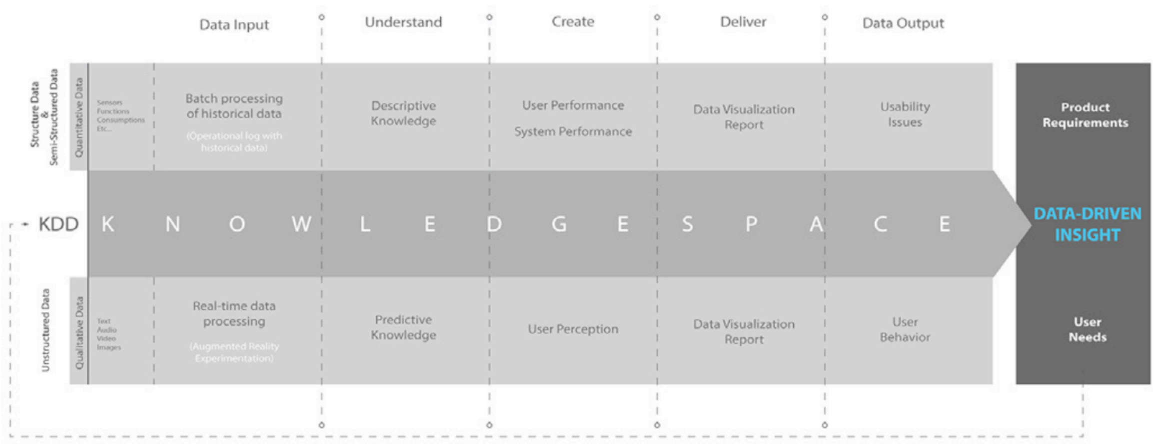

Figure 4.8 Data Transformation Flow into Design Material (J. C. Quinones-Gomez, 2021)

As the design discipline moves towards a more fluid environment, Bernstein (2018) states that, consequently, design intent is no longer a fixed set of graphical instructions but an evolving dataset that integrates analytical insights and can be continuously refined. Moreover, technology-enabled collaboration dissolves strict boundaries between architects and builders by integrating workflows (e.g., integrated project delivery models). This creates shared ownership of design decisions and execution strategies, shifting the architect’s responsibility beyond concept creation toward orchestration of complex data-driven processes that span design through construction and operation.

This permeable setting also brings out some critical questions. The reliability and clarity of information decline across phase boundaries due to transfer via drawings and evolving datasets. This can lead to increased ambiguity, misinterpretation, and potential errors, as knowledge becomes less certain when passed between teams. Additionally, the shifting responsibilities and shared ownership may obscure accountability, making it more challenging to ensure consistent design intent and coordination throughout the project lifecycle.

4.2.2 Data Driven Approaches in Design Process

Melendez et al. (2020) emphasizes that designers have long engaged in data collection as an inherent part of their spatial practices. By recording and collecting information, often through mapping and visualization techniques, they reveal aspects of a design or environment that are not immediately visible or obvious, described here as the “unseen”. Authors further elaborate that these data-driven processes enable designers to uncover hidden patterns, relationships, and phenomena that inform and enrich design decisions. Importantly, this application of data spans various scales, from small installations or localized projects to large architectural works and urban planning. Across these scales, data collection and use provide deeper insights into the complexities of spaces, environments, and human interactions, ultimately leading to more informed and meaningful design outcomes.

Quiñones-Gómez et al. (2024) describe how Data-Driven Design (DDD) has transformed traditional design practices by integrating big data analytics, AI, and machine learning across all design stages. This approach moves away from intuition-based decisions toward an evidence-based process that continuously refines products through empirical insights. It emphasizes a dynamic, user-centered methodology that leverages diverse data sources such as sensor data, user-generated content, and expert insights. These enable ongoing iteration, innovation, and real-time adaptation, improving product quality and user satisfaction.

Quiñones-Gómez et al. (2024) also state that DDD enhances creativity and strategic decision-making by incorporating heuristic tools like TRIZ, analogical frameworks, and computational simulations of human cognition. Real-time feedback from sensors and the Internet of Things (IoT) allows products to adapt rapidly to user needs. However, managing vast datasets and ensuring ethical standards, such as privacy and bias mitigation, pose ongoing challenges, requiring new skills and frameworks for responsible design.

Data-Driven Design (DDD) transforms the traditional design process by integrating data insights at each phase of the product development lifecycle, as depicted in (Figure 4.9). In the initial phase, the collection and interpretation of semantic data, guided by complex network analysis, inform early concept generation. During the design and development phase, predictive analytics and generative models such as GANs and VAEs, as explained by Quiñones-Gómez et al. (2024), are employed to propose innovative solutions based on historical and real-time data. In the prototyping phase, these models facilitate the creation of novel, user-aligned prototypes, reducing reliance on intuition. The testing and refinement stage benefits from real-time interaction metrics and dashboards, enabling continuous, data-informed iterations through the DDD feedback loop. Finally, during implementation and launch, data analytics monitor user engagement to validate and optimize the design, supporting evidence-based decision-making throughout the entire process. This approach, as outlined by the authors, ensures a balanced integration of creativity and empirical insights, fostering user-centered, innovative, and effective design outcomes across each phase.

	Step in Design Process	Description	Role of DDD
1	Identification of Need	Establishing the purpose and requirements for the new product.	DDD synthesises data to identify user needs and market gaps.
2	Research and Analysis	Gathering and analysing data relevant to the product's context and potential user base.	Predictive analytics forecast trends and user behaviours to shape the design direction.
3	Conceptualization	Generating ideas and possible solutions informed by the data collected.	DDD informs ideation with data-driven insights, ensuring concepts are user-centric and innovative.
4	Design and Development	Creating detailed designs and developing prototypes.	Real-time data allows for agile adaptation of the design to emerging insights.
5	Testing and Validation	Evaluating the prototypes using data-driven metrics to ensure they meet the established needs.	Targeted user experience studies enhance testing, ensuring the product meets actual user expectations.
6	Refinement	Refining the product based on feedback and additional data analysis.	Continuous data analysis informs iterative refinements, improving product quality and user satisfaction.
7	Finalization and Launch	Finalising the design for production and introducing the product to the market.	DDD supports decision-making for final adjustments, optimising the product's market fit.
8	Post-Launch Analysis and Iteration	Analysing user feedback and market performance post-launch.	Lifecycle analysis and user feedback inform product iterations, ensuring continuous improvement.

Figure 4.9 Detailed Description of Iterative Steps in the Design Process (Quiñones-Gómez et al., 2024)

As the sources which supports the DDD, including sensor data, user content, expert insights, and internal documents, are combined throughout the development process. Real-world usage data supports idea generation and validation, while AI-enabled design tools streamline workflows, fostering a harmonious blend of creativity and data analysis (2024). This interconnected system supports continuous refinement based on user interactions and measurable metrics.

Within the subject of "complete data-driven design paradigm", Cantamessa et al. (2020a) represents a comprehensive framework that integrates data and digital technologies into every aspect of the design process in (Figure 4.10). Building upon earlier discussions of technological impacts and shifts in development practices, this paradigm highlights the growing centrality of tools and methods.

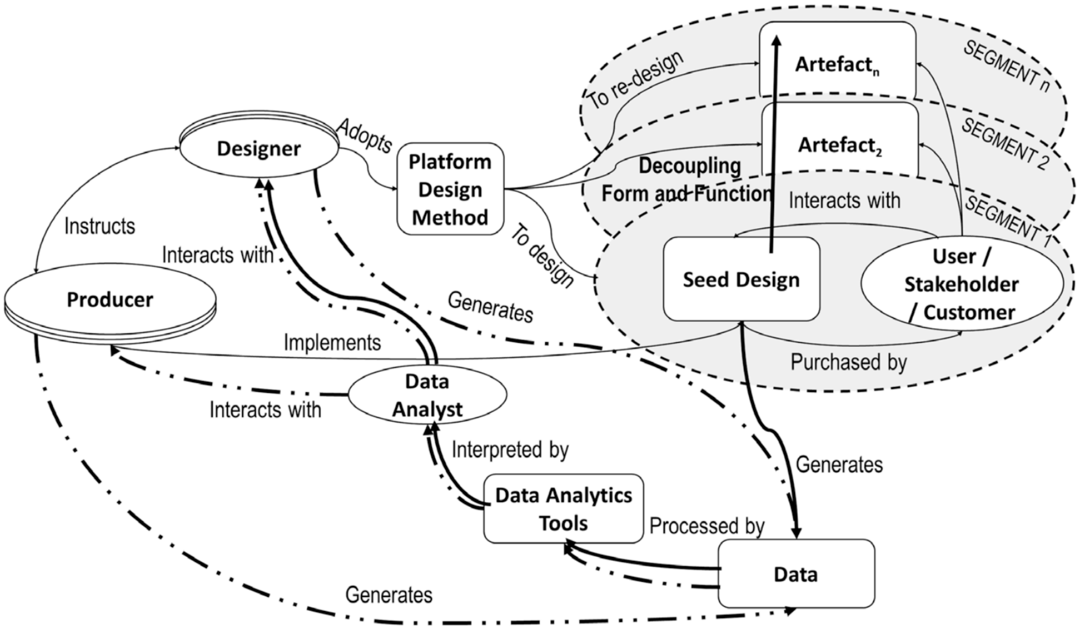

Figure 4.10 The New Paradigm of the Data-Driven Design Context (Cantamessa et al., 2020b)

In the (Figure 4.10), Cantamessa et al. (2020a) illustrates this transformation by showing how tools such as design analytics and platform design have gained increasing importance, effectively supporting designers in interpreting vast amounts of real-time data, exploring multiple design options, and iteratively refining solutions. Platform design here refers to a design approach centered on a core technological infrastructure, or platform, which supports building, extending, and iterating products and features. The figure emphasizes that these tools are no longer merely supplementary but have become central to driving design outcomes, enabling a shift from human-centric, intuition-based approaches to highly data-informed processes.

Overall, the data-driven design paradigm fosters a continuous, platform-driven development cycle where initial "seed designs" are iteratively improved through ongoing data analysis, predictive modeling, and prescriptive recommendations. Such an environment allows for rapid adaptation to changing requirements, customer feedback, and emerging trends, ultimately making the process more flexible, responsive, and effective. The paradigm also raises important research questions, as emphasized by Cantamessa et al. (2020a), about tool development, process management, and education, urges scholars and practitioners to rethink traditional approaches.

While Cantamessa et al. (2020a) focus on product development, the framework and conceptual paradigm shifts they present are highly relevant and translatable to the architectural design process, despite contextual differences between the two domains. Both architecture and product design are fundamentally creative, iterative practices that involve the integration of user needs, technological constraints, and functional objectives. However, the nature of the "product" differs; architectural artefacts are typically large-scale, long-lived, contextually embedded, and subject to a greater variety of regulatory, cultural, and environmental constraints.

Despite these differences, the shift from a linear to a data-informed iterative design approach, as depicted in (Figure 4.10) offers insightful parallels. In traditional architectural workflows, a complete set of drawings and specifications is often developed early and followed throughout construction with minimal change, similar to the “old paradigm” of product development. However, as in Cantamessa et al.’s new paradigm, architecture is increasingly embracing more agile and iterative methods, particularly through the use of Building Information Modeling (BIM), parametric tools, and real-time performance simulations.

Furthermore, the concept of a "seed design" is highly applicable to architectural practice. Early-stage conceptual models or prototypes, when informed by real-time user data, environmental analytics, or post-occupancy evaluations, can evolve dynamically, supporting more responsive and context-aware solutions. The emphasis on platform design methods and data analytics in Cantamessa et al.’s framework finds its architectural counterpart in computational design environments and generative systems that enable modularity, adaptation, and customization at scale.

What this framework suggests for architecture is the need for integrated workflows where architects not only design physical spaces but also engage with data analysts, digital platforms, and feedback loops from building performance and user behavior. Just as product designers must now collaborate with data scientists, architects may increasingly rely on interdisciplinary coordination to interpret data, adjust design parameters, and optimize for performance, sustainability, and user satisfaction.

Ultimately, Cantamessa et al.’s model underscores a broader epistemological shift: from deterministic, specification-driven approaches to dynamic, feedback-oriented processes. This shift calls for rethinking architectural education, practice, and tools, not merely incorporating digital technologies, but fundamentally reconfiguring roles, responsibilities, and design thinking around data-informed, collaborative frameworks.

On the optimization of platform-based design process in architecture, Hensel et al. (2025) illustrate the entire early-stage computational design workflow in (Figure 4.11) that they have developed, which is divided into two main parts: the **translational process** and the **generative process**.

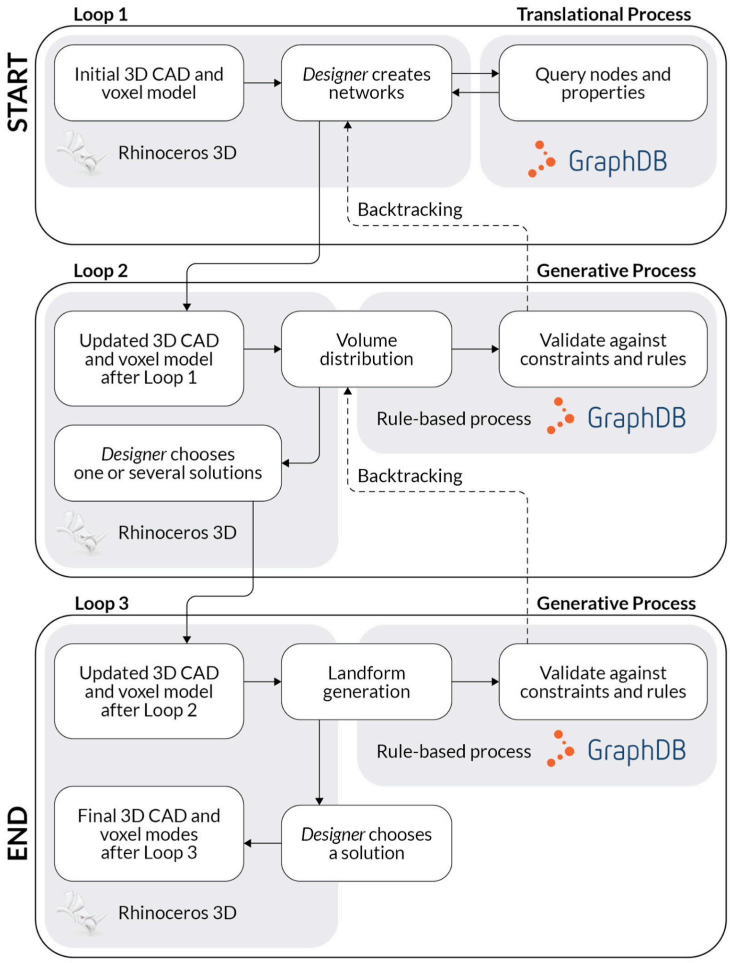

Figure 4.11 Early Design Workflow of the Ontology-Aided Generative Computational Design Process (Hensel et al., 2025)

The translational process, referred to as Loop 1, is the initial phase where the designer analyzes and spatially organizes architectural and ecological requirements derived from the design brief. Through a rule-based system, the spatial configuration of these nodes is evaluated for criteria such as proximity and solar access, providing computational feedback that guides the designer in refining the spatial network.

Following the translational process is the generative process, which includes Loops 2 and 3. Loop 2 focuses on generating various spatial organizations composed of distinct volumetric elements, architectural volumes, soil volumes, and biomass volumes, following generation criteria set by the designer. Loop 3 further refines these spatial organizations by producing diverse geometric articulations and surface geometries, which can be interactively adjusted by the designer in Rhinoceros 3D. The generated solutions are assessed using key performance indicators such as geodiversity and ecological functionality, enabling informed decision-making.

In essence, Hensel et al. (2025) visualizes the cyclic, integrated process where the designer and computational tools collaboratively generate, evaluate, and refine early-stage ecological building envelope designs, bridging conceptual understanding with formalized, rule-based modeling.

Another important approach, The Data-Driven Architecture Framework, introduced by Qabshoqa (2017) and depicted in (Figure 4.12) is a comprehensive conceptual model that outlines the different levels of data intervention in architectural design and operations. This framework was developed through Grounded Theory analysis and serves to clarify how data can be progressively integrated into the architectural process, from initial design decisions to fully autonomous building systems.

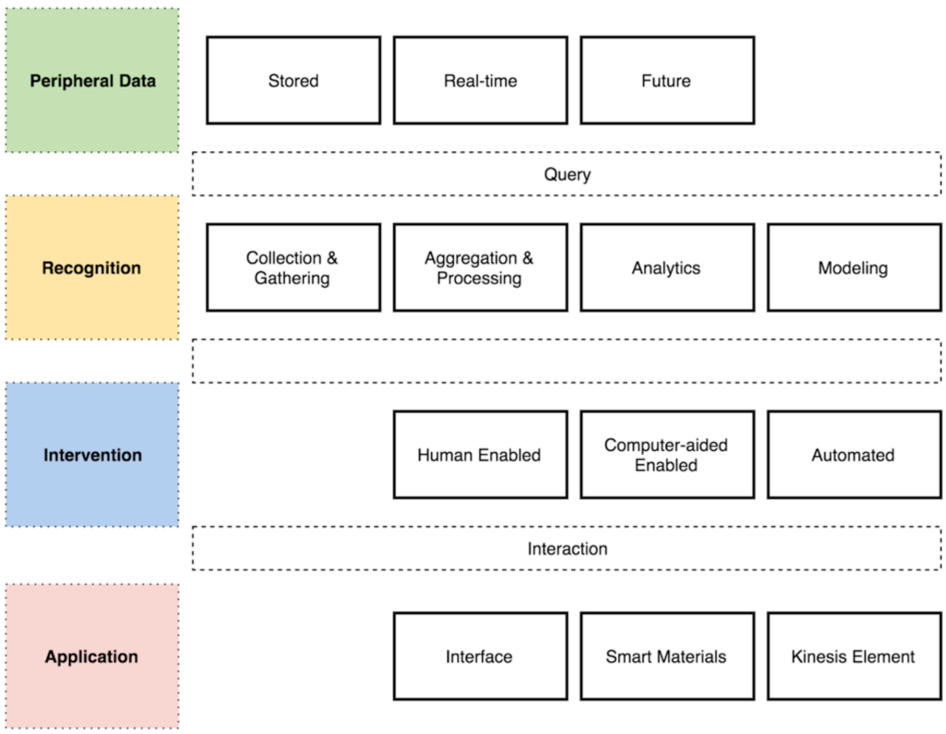

Figure 4.12 Data-Driven Architecture Framework (Qabshoqa, 2017)

At its core, in the framework Qabshoqa (2017) identifies multiple levels of data involvement, starting from Level 0, where design decisions are made solely by architects and designers without the use of data analytics. It then progresses to Level 1, where parametric design tools and predictive simulations assist architects in making informed decisions. Level 2 introduces a data ecosystem within buildings, where real-time data from occupants and operations are collected and analyzed, although human supervision and intervention are still essential. In level 3 and 4, Qabshoqa (2017) reflects increasing degrees of automation and autonomy. Level 3 buildings adjust their configurations automatically using smart materials and architectural elements while allowing occupant intervention. At Level 4, buildings become fully autonomous data-driven entities that communicate with each other as part of a larger urban ecosystem, optimizing performance collectively with minimal to no human intervention.

In the framework, Peripheral Data refers to background data supporting the system. Recognition is how the system detects and understands this data. Intervention describes who or what acts on the data, humans, computer-aided processes, or fully automated machines. Application is how insights are used, such as through interfaces, smart materials, or responsive building elements.

Qabshoqa (2017) elaborates that as architectural data levels progress from 0 to 5, these components evolve: Recognition moves from human-based to machine-driven; Intervention shifts from human-only to fully automated; and Application integrates more directly into building systems. Lower levels rely on human understanding and control of data, while higher levels enable autonomous buildings that sense, decide, and adjust themselves dynamically, creating smarter, self-regulating environments.

The framework highlights the transformative potential of Big Data and the Internet of Things (IoT) in reshaping architectural practices and building operations. By clearly defining these progressive stages, this model helps architects, designers, and engineers understand and implement data-centric approaches to create smarter, more adaptive built environments, ultimately leading toward fully autonomous, self-optimizing architecture.

When the applications of data-driven approaches in architecture are observed, Ahmeti et al. (2025) adopt a data-driven approach that combines ontology-based data management (OBDM) with computational design workflows. Their method focuses on integrating heterogeneous datasets from various domains, including ecological data (such as species distribution and functional traits), geographical data (like maps and site-specific environmental conditions), and architectural data (from CAD and voxel-based models). These datasets are collected from both public sources and internally developed project databases. The authors use ontologies, formal systems for organizing knowledge, to connect and align this diverse data in a consistent and meaningful way. This integrated data is stored in a shared digital framework that can be searched and updated using user-defined rules and queries. Designers can then interact with this system through visual design tools like Rhino and Grasshopper to evaluate whether their design decisions meet ecological requirements, such as whether a location receives enough sunlight for a specific plant or maintains the proper distance between species. Overall, the approach demonstrates how structured data, and expert knowledge can be combined to guide architectural design that supports urban biodiversity.

Presenting several innovative data-driven approaches related to architecture and urban environments, Hensel and Bier (2022) discuss that rapid changes caused by urbanization, such as ecosystem degradation, resource depletion, and social challenges, are complex and interconnected, making them difficult to address through isolated disciplinary or domain-specific approaches. Therefore, there is a need to move beyond narrow, specialized methods toward broader strategies that integrate knowledge and tools across multiple disciplines, cover various related fields, and consider different spatial and temporal scales.

This comprehensive approach relies on several key components, Hensel and Bier (2022) explain: acquiring diverse types of data, sharing this data openly among stakeholders, and integrating it to form cohesive datasets. Moreover, data-driven modeling techniques are essential to analyze and simulate complex interactions within urban environments. Together, these methods provide a powerful means to understand and manage the multifaceted sustainability challenges posed by rapid urban growth, enabling more effective, holistic, and adaptive responses.

By following this explanation, Nicholson et al. (2022) highlight the importance of knowledge exchange for addressing urban complexity through an integrated ecological perspective, Sunguroğlu Hensel et al. (2022) developed an interdisciplinary approach and computational framework to link large-scale territorial planning with detailed architectural design through multi-domain and trans-scalar modeling. Their research has two main focuses: first, understanding environments by uncovering, restoring, and adapting knowledge about land across various conditions and contexts; second, designing environments using computational workflows that integrate data-driven planning and design processes. Van Ameijde and Leung (2022) monitor people’s location, movement, and activity to analyze behaviors for evidence-based urban planning. The paper explores experiments developing user-driven generative design and new computational tools for site analysis and monitoring, enabling data-driven urban place studies.

Attempting a more systematic approach, Van Geest et al. (2021) embrace the domain-driven architecture design methodology that combines thorough domain analysis with the use of architecture viewpoints. This approach involves first conducting a detailed analysis of the, in their case, a smart warehouse domain to identify key features and concerns, which are represented using feature diagrams and domain models. Subsequently, they employ architecture viewpoints, following frameworks such as the Views and Beyond approach, to model the reference architecture from different perspectives, ensuring a comprehensive and adaptable design. The method emphasizes integrating domain knowledge systematically into architectural models, validated through industrial case studies, to support flexible and effective smart warehouse development.

In response to the ongoing digital evolution and its transformative impact on architectural practice, Melendez et al. (2020) explain that the built environment is a complex, dynamic system full of flows and interactions that can be understood as a vast interface of data and information. Traditionally, architects were limited by their ability to manually calculate and analyze these intricate systems. With the advent of computational tools, architects can now model, simulate, and understand complex networks and data flows within architectural and urban contexts.

As a reflection of these changes, Melendez et al. (2020) point that digital tools have evolved significantly from early computer-aided drafting programs like AutoCAD, to sophisticated 3D modeling software such as SketchUp, Rhino, Revit, 3D Max, Maya, and ArchiCAD. These platforms have further developed to include generative and parametric design capabilities (e.g., Grasshopper integrated with Rhino), allowing architects to script and manipulate design parameters algorithmically. This evolution underscores a broader shift towards data-informed and computation-enhanced approaches in contemporary architectural design.

4.2.3 Big Data Applications in Design Process

Offering a general view on Big Data in design and creative processes, Quinones-Gomez (2021) explains that both creative and design processes share key stages: problem definition, idea generation, and idea evaluation in (Figure 4.13). However, they represent different realities. Creativity is a cognitive process focused on formulating ideas, while design is a practical work process producing product or process proposals. Research indicates that integrating big data into the creative process increases novelty and variety by enabling higher levels of abstraction in conceptual thinking, thus expanding the creative potential. Even though integrating big data in design is growing, transforming idea generation and knowledge discovery, Quinones-Gomez (2021) emphasizes that creating value depends on decisions supported by data analysis, which requires specialized data science skills to process complex datasets and extract insights that enhance and inform the design process.

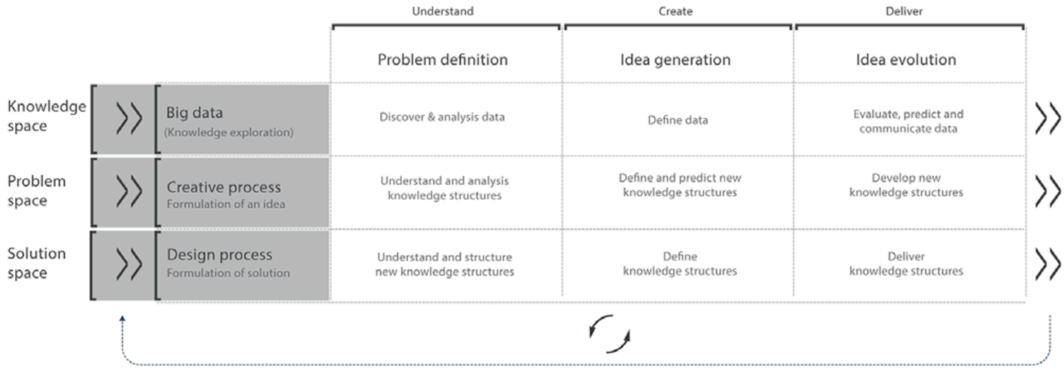

Figure 4.13 Big Data and Creative Process and Design Process Framework (J. C. Quinones-Gomez, 2021)

Focusing on the Big Data in ideation phase, Quinones-Gomez (2019) contributes the subject by stating that after an initial intuitive phase, data inspires designers to generate ideas by enabling abductive perception, forming new meaningful connections. This process increases creativity, fostering novelty and variety, and ultimately raising the chances of designing innovative concepts through data-driven generative design.

Although there is a lack of direct studies linking creativity and big data, Quinones-Gomez (2021) points out that the findings indicate that combining these fields will create a new context in which big data acts as a catalyst for creativity. In this new scenario, which Gomez visualized in (Figure 4.14), Big Data will stimulate and enhance the creative process by providing richer information, deeper insights, and new inspiration to designers, ultimately supporting idea generation and innovation.

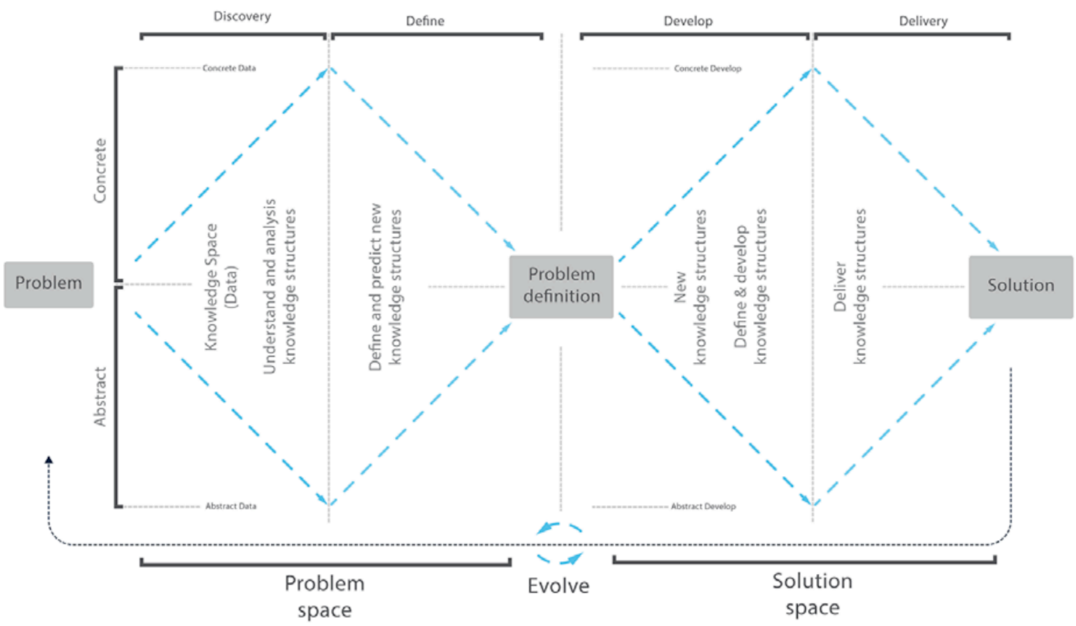

Figure 4.14 Data-Driven Design Model Process (J. C. Quinones-Gomez, 2021)

On the application of Big Data in design process, Quinones-Gomez (2018) highlights five ways big data support the creative process. First, it enhances inspiration by providing well-classified information for idea generation. Second, the Data-Driven Design Model (DDDM) enables iterative feedback, refining creative solutions. Third, combining creativity and science helps designers intelligently evaluate innovative outcomes. Fourth, big data's volume, velocity, and variety advance agile design, simplifying the process with dynamic insights. Finally, integrating big data encourages innovation across multidisciplinary fields by supporting the development of new solutions. Together, these aspects make the creative process more informed, iterative, evaluative, agile, and widely applicable, fostering greater creativity and innovation.

Quiñones-Gómez et al. (2024) highlight that big data analytics provide predictive insights that guide strategic decisions. Generative models like GANs and VAEs leverage historical data to propose innovative design solutions, which are refined through iterative cycles aligned with performance metrics. These tools enable designers to explore creative solutions grounded in empirical evidence, fostering both innovation and feasibility.

Big data has become a comprehensive tool that contributes to architecture across a wide range of domains, extending far beyond its technological roots. From design decision-making and sustainability analysis to the study of user behavior and large-scale urban planning, data-driven approaches are reshaping architectural practice. At various stages of the architectural process, big data enables more informed design, optimized workflows, and deeper insights into user needs. This transformation is not only influencing the formal aspects of architecture but also redefining its functional and environmental dimensions:

In a very fundamental approach, Hois et al. (2009) discuss how architectural design can benefit from combining different types of information, like spatial measurements, conceptual ideas, and functional details, all organized through modular ontologies. Modular ontologies are a way of organizing knowledge into distinct, interconnected modules or parts, each capturing a specific aspect or perspective of a domain. This approach allows for flexible combination, extension, and refinement of knowledge, keeping different conceptual areas separated yet able to interact coherently.

They further elaborate, in the context of architectural design, that modular ontologies are used to formalize various aspects of building environments, such as spatial constraints, functional requirements, and conceptual considerations. They facilitate the formal representation of heterogeneous information, like geometric data, functional roles, and spatial relationships, and help analyze whether a design satisfies specific criteria, especially in ambient intelligence environments.

While developed prior to the flow of big data in architectural practice, particularly the use of modular ontologies to represent functional, spatial, and conceptual knowledge remains highly relevant for structuring and interpreting the complex datasets characteristic of contemporary architectural workflows. In the context of big data and data analytics, where information from sensors, BIM models, user behavior, and environmental feedback must be analyzed coherently, ontologies provide a way to give data clear meaning and relationships so it can be better understood by humans and machines. The modular aspect of the approach of Hois et al. (2009) discuss, aligns with current practices in scalable data architecture, allowing for specialized yet interconnected data domains such as thermal comfort, spatial accessibility, or user interaction to be developed independently and synthesized contextually. IFC is a particular example of modular ontologies, which is used for enhancing interoperability in BIM by making building elements machine-readable and allows architects and systems to query and analyze building components semantically.

Evins (2013) evaluates the subject of data management in computational optimization for sustainable building design by highlighting the challenges associated with handling complex, multi-objective datasets and emphasizing the importance of effective data analysis and visualization methods. Specifically, he discusses the difficulty in interpreting outputs from multi-objective optimizations, which often involve complex interactions in high-dimensional design and objective spaces. He emphasizes that advances in data handling tools are fundamental for translating large simulation datasets into actionable insights in sustainable building design.

Meekings and Schnabel (2016) highlight that Big Data tends to average out or ignore individual differences and special cases. This means it can overlook unique or rare characteristics that make a person or situation special, which are often crucial in areas like architecture, where individuality matters. Instead, they advocate focusing on personal big data, which preserves unique characteristics and provides deeper insights.

They further emphasize that real-time big data, particularly from social media and other digital sources, plays a crucial role in creating more responsive and adaptive architectural environments. This dynamic feedback loop allows for a more personalized and engaging experience, especially in settings with diverse and changing populations like public buildings, transport hubs, or hotels.

However, this approach includes concerns related to privacy, consent, and data security. Real-time data collection can intrude on individuals' privacy by continuously monitoring and analyzing their behaviors, preferences, and movements without explicit permission. Additionally, there is a risk that such data might be misused or exploited, leading to ethical issues and potential harm to individuals' rights. Ultimately, these challenges highlight the importance of establishing ethical guidelines and safeguards to protect individuals while leveraging big data in architectural practices.

Ekici et al. (2019) reviews the performative computational by examining how computational methods like swarm intelligence and evolutionary algorithms are used to optimize architecture for performance. PCA is a way of designing buildings using computer tools that help explore different shapes and solutions. It focuses on creating designs that not only look good but also perform well in criteria like energy use, lighting, and comfort. It combines ideas from architecture and computer algorithms for better, more efficient building designs:

Jin and Dong (2020) employed qualitative data analysis techniques, specifically using software like NVivo for coding and extracting meaningful patterns from a large dataset of award-winning digital designs. The researchers collected extensive data from 998 design examples sourced via web crawling technology, which included images, descriptions, and categories. They systematically grouped similar designs based on shared functions and features to identify design heuristics.

In the design process, this big data approach enabled the identification of common innovative strategies and features across diverse digital products and services. By analyzing a large, representative sample of successful designs, the researchers derived generalizable heuristics that assist designers during early conceptual stages. Essentially, the data acts as a rich knowledge base, informing the creation of heuristics that foster diversity and novelty in conceptual design and helping to overcome issues like design fixation and supporting more innovative outcomes.

On the inspirational aspect of Big Data in architecture, Trapold and Saldana (2023) suggest that Big Data and AI tools can be used primarily as sources of *inspiration* instead of directly generating physical or finalized architectural designs. In other words, rather than expecting big data to produce a complete building or plan, its value lies in offering designers new ideas, themes, visual atmospheres, or conceptual insights drawn from vast and diverse datasets. Authors offer a methodology where data collected from social media platforms is aggregated and visualized through SOM (Self-Organized Maps) to identify patterns, atmospheres, and design cues. Further, refining the data sets by relying on human input and focusing on relevant information.

By using big data purely as a stimulus for creativity, Trapold and Saldana (2023) emphasize that human intuition and judgment are central to design. AI and big data serve as complementary resources that inform and spark innovation rather than replace the nuanced decision-making that architects bring to their work. This balance helps avoid the risk of over-reliance on automated design tools, which may overlook important qualitative aspects and context-specific subtleties.

Exploring the AI in the conceptual design phase, Castro Pena et al. (2021) explains that conceptual architectural design is inherently complex because it relies on both the architect's prior knowledge and creative thinking to develop new design ideas. Unlike later stages in design where requirements are clearly defined, at the conceptual phase, the design goals and constraints are still vague and evolving. Therefore, applying artificial intelligence (AI) in this context should not aim to find a definitive solution within a fixed set of parameters (a defined search space), because the problem itself is not fully specified yet.

Instead, AI should support an exploratory process Castro Pena et al. (2021) claims, helping to investigate and better understand both the requirements (which may change over time) and the range of potential design solutions that could fulfill those needs. This exploratory use of AI aligns more closely with how architects iterate and develop concepts during early design, fostering creativity and discovery rather than optimization of a finalized problem.

Castro Pena et al. (2021) highlights an increasing use of data-driven AI methods in architectural design, particularly in shape optimization and performance evaluation, with evolutionary computing and deep learning playing key roles. However, acquiring comprehensive, standardized, and domain-specific datasets remains challenging, limiting AI's full potential in architecture. AI must integrate both quantitative factors, such as structural performance and energy efficiency, and qualitative aspects, like aesthetic appeal and user preferences, but the lack of explicit evaluation criteria and standardized data hinders this integration. Additionally, Castro Pena et al. (2021) adds that many AI approaches lack rigorous independent validation due to limited or inconsistent data, affecting scientific rigor and reproducibility. Overall, improving data availability and quality is essential to advancing the reliability and practical application of AI-driven architectural design tools.

Castro Pena et al. (2021) also outlines key data analytics techniques used in AI for architectural conceptual design. Evolutionary Computation (EC), particularly genetic algorithms. (GA) is widely employed to explore and optimize complex design solutions by balancing multiple criteria (Ekici et al., 2019). Artificial Neural Networks (ANN), including deep learning and GANs, aid in modeling and predicting architectural forms and performance due to their adaptive learning capabilities. Fractal algorithms, inspired by natural patterns, generate complex shapes using recursive rules, supporting nature-inspired designs. Swarm Intelligence and Particle Swarm Optimization simulate collective insect behaviors to optimize spatial arrangements. Cellular Automata (CA) models spatial features via rule-based local interactions and were prevalent in earlier studies, but less so recently. Additionally, Interactive Evolutionary Computing integrates human feedback in optimization, combining computational and qualitative inputs. These methods collectively advance AI-driven architectural design (As et al., 2018; Caldas & Norford, 2003; Evins, 2013; Takagi, 2001; Vehlken, 2014)

Elaborating on the concept of big data in architecture, Bernstein (2018) frames the subject as the collection of vast and diverse datasets, often referred to as 'data lakes', gathered by industry players from multiple sources. These sources include design-related information such as digital models and analytical results, as well as real-world data encompassing construction processes and sensor readings from occupied buildings (e.g., temperature, energy use, and system performance.

Following the data aggregation phase, Bernstein (2018) adds that machine learning algorithms then process and analyze these extensive datasets to identify patterns, correlations, and insights that are not immediately apparent through traditional analysis. This exhaustive examination enables the extraction of valuable knowledge about building performance, construction efficiency, and operational outcomes. Ultimately, these insights inform improved design decisions, predictive analytics, and optimized building management, transforming raw data into actionable intelligence that enhances all phases of the building lifecycle.

As architecture continues to evolve within a technologically accelerated context, data has emerged as a pivotal resource shaping the design process. From performance optimization to form generation, and from environmental simulation to machine learning integration, a wide array of computational techniques now supports, augment, or transform how architects conceive, develop, and evaluate built environments.

A comprehensive overview of the major data-driven techniques currently used in architectural design is depicted in (Table 4.4). Table classifies each method by its computational category, such as performative simulation, generative algorithms, machine learning, behavioral modeling, and human, AI collaboration, offering a detailed description, common applications, relevant design stages, and the types of data they engage with. The aim is to clarify the functional scope and methodological diversity of these tools, enabling a deeper understanding of how data and digital computation are reconfiguring design thinking and practice.

Crucially, many of these techniques are deeply connected to **big data** and **data analytics**, either by relying on large-scale datasets for training, simulation, or pattern recognition, or by generating significant volumes of data as part of iterative design-feedback loops. Techniques such as predictive modeling, digital twins, and unsupervised learning draw directly from big data frameworks, while others like agent-based modeling or performance feedback systems benefit from analytics to drive real-time or evidence-based decision-making.

Table 4.4 Major Data-Driven Techniques Currently Used in Architectural Design (created by author)

Category	Technique	Description	Application	Design Stage	Data Type
Performative, Simulation-Driven Design	Performative Computational Architecture	Embeds performance metrics directly into design generation.	Designing forms responsive to daylight or airflow.	Concept Design, Optimization	Environmental Performance
	Environmental Simulation Integration	Uses tools like CFD, daylight, or energy models in the design loop.	Evaluating thermal comfort in early design.	Design Development, Evaluation	Environmental Material, Climatic
	Multi-Objective Optimization	Balances multiple goals using evolutionary search.	Optimizing cost, structure, and daylight.	Design Development, Optimization	Performance, Geometric
	Topological Optimization	Removes unnecessary structure based on load analysis.	Structurally optimized beams, slabs, or shells.	Detailed Design	Structural, Geometric
	Physics-Informed Neural Networks (PINNs)	Neural nets trained with physical constraints.	Predicting material behavior without simulation.	Design Evaluation	Structural, Material, Simulated
	Digital Twin Systems	Virtual models updated with real-time sensor data.	Adaptive HVAC and lighting systems.	Post-Occupancy, Operations	Sensor, Behavioral, Environmental
	Predictive Modeling & Regression	Forecasts future building performance.	Pre-construction performance assessment.	Feasibility, Evaluation	Historical, Environmental
Generative & Form-Finding Techniques	Evolutionary Algorithms	Fitness-based optimization to generate optimal forms.	Skyscraper shapes balancing views and shadow.	Concept Design, Form Generation	Performance, Geometric
	Swarm Intelligence / PSO	Models collective behavior to form spatial solutions.	Emergent circulation or clustering.	Concept Design, Form Generation	Spatial, Contextual
	L-Systems	Rule-based branching growth systems.	Tree-like pavilions or structures.	Concept Design	Geometric, Structural Logic
	Shape Grammars	Symbolic rules for form variation.	Modular façades, housing layouts.	Early Design, Schematic	Geometric, Typological
	Cellular Automata	Evolving forms via local interaction rules.	Urban growth and block patterns.	Site Planning, Urban Design	Spatial, Topological
	Morphogenetic Design	Growth based on biological or physical processes.	Biomimetic shells and facades.	Form Generation	Environment Structural
	Parametric & Algorithmic Design	Geometry controlled by parameters.	Adaptive shading or responsive structures.	All Stages	Geometric, Environmental
	Design Space Exploration (DSE)	Automates exploration of multiple design alternatives.	High-volume option generation and evaluation.	Form Generation, Optimization	Geometric, Performance
Machine Learning & AI-Based Techniques	Generative Adversarial Networks (GANs)	Learns and synthesizes new images or forms.	Façade design and style transfer.	Concept Design	Visual, Historical
	Artificial Neural Networks (ANNs)	Predicts complex outputs from various inputs.	Energy performance prediction.	Evaluation, Optimization	Historical, Performance
	Deep Learning (Image-to-Design)	Converts visual data into spatial configurations.	Site zoning from sketches or satellite imagery.	Early Design, Planning	Visual, Spatial
	Reinforcement Learning (RL)	Learns design decisions through iterative feedback.	Adaptive layout and routing strategies.	Planning, Optimization	Spatial, Behavioral
	Supervised / Unsupervised Learning	Classifies and clusters large design datasets.	Urban form classification or user patterns.	Site Analysis, Typology	Behavioral, Geometric
	Bayesian Optimization	Efficient search over complex solution spaces.	Form and performance optimization.	Optimization	Performance, Simulated
Behavioral & Emergent Modeling	Agent-Based Modeling	Simulates interaction of autonomous agents in space.	Pedestrian movement in public architecture.	Space Planning, Simulation	Behavioral, Spatial
	Cellular Automata	Models form evolution using local rules.	Organic urban expansion or density modeling.	Urban Design, Early Planning	Spatial, Site-Specific
	Swarm Intelligence	Uses group behavior to self-organize form.	Spatial clustering or aggregation logic.	Concept Design	Spatial, Topological

Data Analysis & Interpretation	Data Mining & Clustering	Finds patterns and groups in large datasets.	User behavior analysis, urban typologies.	Site Analysis, Post-Occupancy	Behavioral, Historical
	Predictive Modeling & Regression	Statistical prediction of outcomes.	Estimating energy or cost outcomes.	Feasibility, Evaluation	Historical, Environmental
	Space Syntax Analysis	Measures spatial configuration and connectivity.	Flow optimization in buildings and cities.	Schematic Design, Planning	Spatial, Geometric
	Performance Feedback Loops	Real-time analysis integrated into design iteration.	Automated façade adjustment for solar data.	Form Development, Optimization	Performance, Environmental
	Dimensionality Reduction (PCA)	Summarizes complex simulation data.	Simplifying design variable exploration.	Data Analysis	Simulated, Performance
Human–AI Collaborative Systems	Hybrid Human–AI Co-Design	Human-guided selection of AI-generated options.	Adaptive massing studies or planning blocks.	Concept Design, Schematic	Mixed (Geometric, Behavioral)
	Interactive Evolutionary Design	Designers evolve generative results based on preference.	Aesthetic-led form evolution.	Concept Design	Visual, Geometric
	Active Learning Systems	AI systems that improve through user feedback.	Adaptive recommendation systems for layouts.	Design Development	Behavioral, Feedback-Based

4.3 Systematic Impacts and Future Directions for Design Processes

Digital technologies are continuing to redefine the foundations of architectural infrastructure, implementation strategies, and future professional outlooks. As tools like BIM, data analytics, machine learning, and cloud-based collaboration platforms are increasingly adopted, architectural processes become more fluid, interconnected, and reliant on real-time information. These tools enable designers to coordinate complex projects with greater precision, simulate outcomes with predictive accuracy, and integrate various forms of input across disciplines. The shift reflects a movement away from static documentation and isolated phases toward more integrated, data-enriched design and delivery models.

This transformation invites reflection on deeper questions related to authorship, accountability, and disciplinary identity. As automation and algorithmic systems begin to shape decisions once made through human judgment, architects must adapt by developing new skills and ways of thinking that embrace computational logic while retaining the core values of design. The architect’s role expands from form-making to include responsibilities in data curation, system coordination, and collaborative management within digital ecosystems.

With these advances, critical challenges and considerations arise. The growing reliance on data and intelligent systems raises ethical concerns around transparency, privacy, and bias. Questions about control over data, algorithmic decision-making, and the inclusivity of design processes become central. Navigating this new narrative requires not only technical competence but also critical awareness, ethical sensitivity, and a willingness to engage with the broader societal implications of digital architecture. As the discipline continues to evolve, balancing technological potential with human-centered priorities remains an essential task.

4.3.1 Design Evolution Through Digitalization

In the current work environment, with digital tools housing vast amounts of design information, architects must develop competencies in managing, controlling, and verifying data quality. Bernstein (2018) assumes roles akin to information curators and integrators, coordinating multidisciplinary data streams and ensuring coherence with project goals. Consequently, architectural methodology is reconceived around iterative, predictive, and collaborative processes, elevating the profession's value proposition as technology transforms architectural practice.

Critically, Cantamessa et al. (2020a) explores the specific impacts of digitalization on product development and design processes. Digital technologies and the data revolution are not only transforming products themselves but also fundamentally reshaping how design development is conducted, as illustrated in (Figure 4.15), which builds upon Cantamessa et al.'s (2020a) relational model, this shift moves development from a traditional, linear, and fixed approach toward more agile and iterative methods. Instead of aiming to produce a complete and reliable set of design specifications upfront, development now emphasizes adaptable, ongoing refinement driven by real-time data.

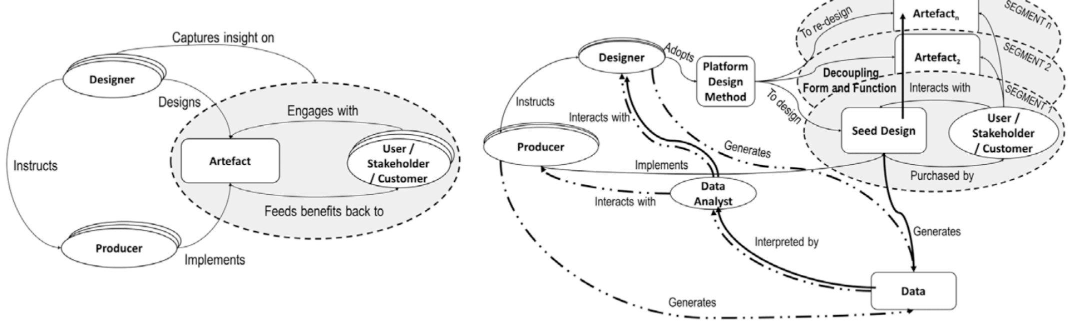

Figure 4.15 The Shift from old paradigm to the New One (Cantamessa et al., 2020b)

In (Figure 4.15), the authors illustrate the shift in product development paradigms brought about by digitalization. It compares the traditional "old" development process with the new, digital-era process. In the old paradigm, product development was a linear, sequential process. Teams worked toward establishing a complete set of specifications upfront, and all subsequent activities adhered to these fixed plans. Any design changes occurring late in the process proved costly and difficult to implement, limiting flexibility and responsiveness.

In contrast, Cantamessa et al. (2020a) further depict that the new paradigm introduces a non-linear, more flexible approach. Instead of finalizing all specifications from the start, development begins with a “seed design”, a basic initial concept or prototype. This seed design is then iteratively improved, extended, and refined through multiple cycles over time. Decisions about the product’s form, behavior, and functions may be deferred or postponed to later stages, allowing the design to evolve dynamically. This process is platform-driven, with digital platforms enabling continuous updates, collaboration, and integration of real-time data.

Overall, within this transformation, numerous operational, managerial, and organizational challenges arise as Cantamessa et al. (2020a) emphasize. For example, iterative validation steps must be managed through close collaboration among designers, marketing teams, and data analysts, requiring new coordination mechanisms. Additionally, existing legal and certification frameworks face difficulties adapting to specifications that evolve throughout the development process.

Perhaps the most critical outcome of this transformative design era is determining what is changing and what remains as a core value, preserving its foundational roots. While information technologies have created significant disruption across academic, professional, and educational fields and redefined everything from job descriptions to design tools, the fundamental methodology of design remains intact. As Rittel and Webber (1973) describe, the design process will always rely on heuristic reasoning, as building design is inherently a "wicked problem": complex and lacking definitive solutions. Bernstein (2018) asserts that although the core methodology of design remains unchanged, what evolves are the inputs that shape design decisions and the validation processes used to justify them. These new inputs emerge from three primary sources: advances in modeling, algorithmic generation, and analytical evaluation.

Building on the design’s foundational value, Bernstein (2018) exemplifies that future modeling tools might enable buildings to be represented at varying levels of detail, from simple proxies to detailed, performance-based models that include construction data, costs, and installation procedures. By linking these models to algorithms through scripting tools like Grasshopper or Dynamo, architects can quickly generate numerous design alternatives, enhancing exploration. Additionally, analytical tools will provide real-time feedback on structural, environmental, and economic impacts during the design process, supported by enriched simulations and Big Data from construction and operations, broadening the scope of design evaluation.

As a connected discipline in the design area, another presumption by Bernstein (2018) is that architectural education will continue emphasizing formal design but will increasingly train designers to integrate form-generation and form-validation tools. This process demands heightened synthetic thinking and judgment to navigate the expanded design alternatives effectively. Further, Bernstein emphasizes, architects must avoid generating excessive “useless alternatives” computationally, a risk cautioned against by architect Cesar Pelli, because it undermines a validated and productive design process.

Qabshoqa (2017) also emphasizes a significant limitation in current valuation methods used in architecture: the lack of comprehensive integration of data collection, analysis, and transformation throughout both the valuation and design processes. Typically, the values considered are predefined early in the design stage, and although efforts are made to measure these values during the process, there is no continuous assessment of the actual value achieved, nor an evaluation of its impact on the building or its occupants.

Also, Qabshoqa (2017) identifies two main issues that contribute to this gap. First, these valuation methods are primarily designed for management purposes, highlighting values separately from the design development and the final constructed outcome. As a result, the valuation process does not actively inform or integrate into the evolving design. Second, the methods are externally imposed on the design operation rather than being an inherent part of it. Consequently, values do not naturally emerge or get realized within the design process itself or through the involvement of stakeholders, creating a disconnect between value assessment and design practice.

This disconnect stems largely from insufficient data and the absence of continuous feedback mechanisms throughout the project lifecycle, as Qabshoqa (2017) emphasizes. However, the adoption of digital technologies such as Big Data and the Internet of Things promises to address these shortcomings. By enabling enhanced data collection, integration, and real-time feedback, these technologies can foster continuous value assessment embedded within the design and construction processes. This approach could effectively bridge the current isolation between valuation methods and design operations, paving the way for more informed and value-driven architectural practice.

Qabshoqa (2017) further elaborates and emphasizes the pivotal role of digital data in architecture, introducing "Digital Value" as a new external value that enhances social, economic, and psychological benefits in design. As the author depicts in (Figure 4.16), the proposed Digital Valuation Method integrates data-driven insights to create and measure value throughout the architectural process. This shift requires architects to develop specialized digital skills, leading to the emergence of the "Digital Architect" role., The Digital Architect designs and implements digital systems that ensure Digital Value generates other value types during projects. This evolution transforms architectural practice by embedding data at its core, enabling more informed decisions and richer value creation within the built environment.

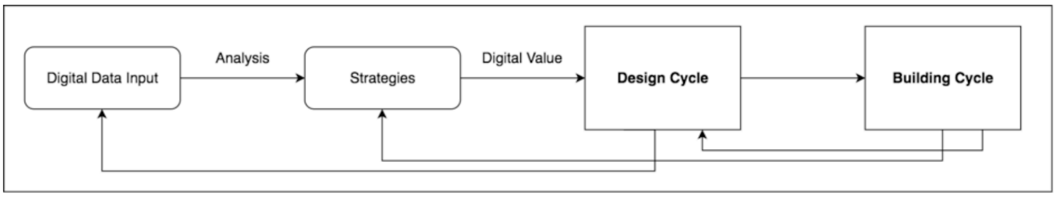

Figure 4.16 Digital Valuation Method (Qabshoqa, 2017)

The evolving nature of architectural design increasingly reflects a synthesis between intuition and data, between qualitative aspirations and quantitative validation. In this context, Biskjaer et al. (2021) highlight the critical relationship between “softer” values such as user experience, cultural meaning, and aesthetics, and “harder” values like energy performance, structural efficiency, and cost in (Figure 4.17). As design processes become more interdisciplinary and data-informed, the integration of these values becomes a central concern. The growing use of big data and computational tools supports this evolution by enabling architects to evaluate both types of criteria more systematically. This convergence marks a shift in design thinking: from isolated, discipline-specific priorities toward more holistic, value-balanced frameworks that align with the complex demands of contemporary building environments.

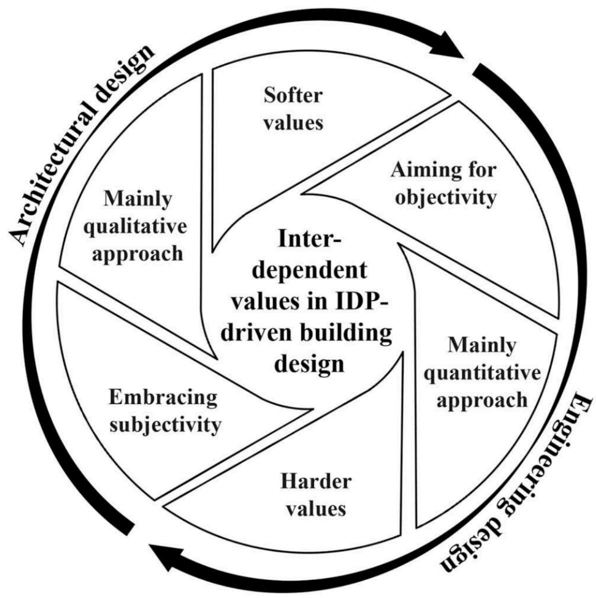

Figure 4.17 Interdependent Values in Integrated Design Process-Driven Building Design (Biskjaer et al., 2021)

4.3.2 Future Perspectives on Data, Technology, and Design Methodologies

Looking ahead, the integration of data and digital technologies not only reshape design methodologies but also produces a redefinition of core architectural concepts such as materiality. In the age of digital and virtual technologies, Picon (2011) critically examines the evolving relationship between materiality and representation in architecture. Although written over a decade ago, Picon’s analysis remains highly relevant, as many of the questions he raised continue to resonate in today's rapidly advancing digital landscape. Key discussion posed by Picon is whether the digitalization of architectural design truly represents a radical break from traditional understandings of materiality or if it is part of a continuing evolution in which materiality remains central, though its definition is shifting. He argues that virtual representations do not stand apart from physical realities but are integrated within a continuum of architectural experience that involves both tangible and intangible dimensions and further explores how digital tools and computer simulations may expand and redefine how architects conceive and manipulate materials, allowing for new forms, properties, and interactions that were previously unimaginable.

This perspective highlights reconsideration of architectural representation’s role, emphasizing its importance in carrying the experiential, sensory, and material qualities of built environments to both designers and users. Ultimately, Picon’s exploration invites architects to rethink materiality as a dynamic, multifaceted concept that adapts alongside technological innovation, preserving its enduring significance even in virtual or digitally mediated realms.

Reflecting on a more idealized, future scenario, Bernstein (2018) examines how the fundamental nature of the project delivery would be transformed when digital technologies such as BIM, collaboration infrastructure, big data analytics, and machine learning become integral tools in architecture and construction. In that case, he presumes that project delivery centers on creating and exchanging digital information that directly guides physical construction. Further, this digital information is managed by different players according to their roles, risk capacity, and agency. In such a context, high-resolution BIM-organized data is integrated across design and construction, allowing all participants to access, update, and leverage it regularly. The resulting data infrastructure, ideally, exhibits key characteristics: **accessibility**, information is indexed and searchable by all team members, **transparency**, each contributor’s work and progress is visible to others, **immediacy**, data is delivered to the precise point of use, **accuracy** information is coordinated and reliable through BIM principles, and **predictability** simulation and analysis allow testing and forecasting future conditions. This, concludes Bernstein, fundamentally transforms project delivery into a collaborative, data-driven, and reliable process.

For the next decade of the architectural design and processes**,** Bernstein (2018) presumes that architectural technology will have matured significantly, transforming design and construction practices. Digital modeling will offer high-resolution representations of materials and methods, while computational analytics will enable rapid evaluation of design decisions across cost, energy, occupant behavior, and lifecycle performance. Algorithms will generate numerous design alternatives, expanding creative possibilities. Vast data lakes from design, construction, and building operations will be leveraged through machine learning to inform evidence-based design. Integrated collaboration platforms will enhance transparency and coordination among project teams. Construction will increasingly shift toward industrialized methods with off-site fabrication and automated assembly, improving efficiency and precision. These developments will necessitate architects to master managing complex data, leading multidisciplinary teams, and merging design intent with construction execution, fundamentally evolving architectural practice and competencies by 2030.

Bernstein (2018) illustrates the interconnected relationship between technological advancements, the development of professional competencies, and pedagogical strategies in architecture in (Figure 4.18). He emphasizes how emerging digital tools and methods drive the evolution of necessary skills for architects and highlight the imperative for educational approaches to adapt accordingly. This framework underscores the cyclical and dynamic nature of practice, where technological progress informs required competencies, which in turn shape pedagogical reforms to prepare future architects for a rapidly changing profession.

	Technology/Process								
	Modeling	*Prediction*	*Generation*	*Big data*	*Collaboration*	*Industrialized building*	**Obligation**	**Competency**	**Pedagogy**
1 Design validation	●	●	●	●			Demonstrate outcomes of design	Integrate analytical results with design strategy	Deploy analysis is service of heuristics
2 Information integration and coordination	●	●		●	●	●	Coordinate all digital project information to service project	Organize and prioritize collaboration approaches	Manage team dynamics
3 Design-to-construction continuity	●	●	●			●	Bridge design information to assist in construction	Understand construction implications of design decisions	Design to build
4 Data aggregation and analysis		●	●	●			Leverage big data sources to support design and construction outcomes	Structure, collect, manage data and systems	Understand building data epistemology

Figure 4.18 Relationship between Technology, Competencies and Pedagogy (Bernstein, 2018)

Quiñones-Gómez et al. (2024) call for future research to improve Data-Driven Design methodologies, better integrate data into design workflows, and uphold privacy and ethical standards. They evaluate DDD as a human-centered, transformative approach that combines data insights with creativity and ethical judgment to advance innovative, responsible design in a rapidly evolving digital landscape.

While data insights inform design choices, Quiñones-Gómez et al. (2024) emphasize that creative intuition remains essential. Designers now need hybrid skills, combining traditional design expertise with data analysis, to navigate complex environments, and shift from interpreting constraints to actively defining conditions using dashboards, visualizations, and predictive models. This integration results in responsible, innovative, and user-centered outcomes.

Overall, Quiñones-Gómez et al. (2024) underscore those ethical issues, such as data privacy, ownership, bias, and transparency, are central to DDD. As reliance on personal, sensor, and usage data grows, establishing comprehensive privacy frameworks is vital. Managing large, diverse datasets also demands advanced skills, highlighting the importance of balancing automated analysis with human judgment to ensure responsible innovation.

As the role of data and digital tools continues to grow in architecture, the way design is practiced and understood is also changing. Instead of following traditional step-by-step workflows, design today is moving toward more flexible, iterative, and data-informed processes. This shift does not mean leaving behind established design values like creativity, spatial thinking, or sensitivity to context. Rather, these values are now being supported by new forms of analysis, modeling, and feedback that come from working with digital systems. In this new setting, architects are not just drawing or modeling, they are organizing complex information, making decisions based on data, and working closely with both machines and people.

Approaches like Data-Driven Design show how tools that were once optional have now become essential parts of the design process. But this also brings new responsibilities. As more personal and sensor-based data is used in design, ethical concerns like privacy, bias, and control need to be addressed more carefully. At the same time, relying too much on automated systems can risk overlooking the human side of design. Intuition, interpretation, and lived experience are still central to meaningful architectural work and must remain part of future design methods.

These changes also affect how architects are trained. It is no longer enough to just be skilled in drawing or software. Today's architects need to understand how data works, how to collaborate with others across disciplines, and how to manage uncertainty and complexity. As design becomes more connected to large-scale systems, from smart buildings to urban infrastructures, methodologies will need to stay flexible and responsive to different contexts, environments, and users.

In the end, the future of design is not about choosing between digital and traditional methods, but about finding ways to connect them meaningfully. Data and technology can open new directions for architectural thinking, but only if used carefully and with purpose. A balanced approach, one that values both innovation and critical reflection, can help shape a design practice that is better prepared to meet the complex needs of today's world, without losing sight of its human and creative foundations.

Chapter 5

Conclusions and Future Prospects

5.1 Conclusions

This study has investigated the multifaceted role of Big Data in the architectural design processes, focusing on its conceptual, technical, and epistemological dimensions. From the foundational understanding of data as an abstract construct to its current applications, this research has emphasized that Big Data is no longer a secondary concern. Instead, it is a transformative force reshaping how design problems are defined, analyzed, and represented.

The integration of data analytics into architectural workflows introduces new challenges and possibilities. Architects are now expected to interpret large-scale data inputs, navigate unfamiliar computational tools, and collaborate across disciplines, skills that exceed the traditional boundaries of design education. Yet, as this thesis demonstrates, these emerging demands also offer opportunities for redefining the role of the architect, not as a passive recipient of information, but as a translator of data into spatial experience.

To clarify and reflect on the structure of the exploration, the thesis revisits the five key research questions in (Table 5.1) below, their associated objectives, and the underlying evaluations that guided the research. Questions were designed not only to uncover the technical roles of big data but also to examine its broader conceptual and disciplinary relevance within architectural design. They aim to explore how data transforms the foundation of design knowledge, reshapes methodologies, and introduces new dynamics in authorship, creativity, and decision-making. By foregrounding these issues, the thesis moves beyond a tool-oriented analysis to engage critically with the shifting paradigms that data introduces into architectural discourse.

The findings demonstrate that Big Data is increasingly being embedded in early-stage design, performance analysis, user behavior modeling, and adaptive design frameworks. Simultaneously, limitations in tool accessibility, skill development, and conceptual clarity remain persistent barriers. As such, the thesis argues that any productive engagement with data must go beyond software implementation, it must be rooted in critical reflection and interdisciplinary literacy.

Table 5.1 Research Questions, Objectives and Evaluations (created by author)

Research Questions	Objectives	Evaluations
How is big data currently used in architectural design?	To identify key applications of big data in design workflows.	Big data is already applied in simulation, environmental modeling, urban analysis, and performance-based design, allowing for more data-rich and predictive workflows.
What challenges arise in integrating data analytics in design?	To explore both technical and conceptual barriers.	Key challenges include lack of data literacy among architects, limited interoperability between tools, and concerns around authorship and creative autonomy.
What potential does data hold for shaping future design logic?	To reflect on new directions in data-driven architecture.	Big data fosters anticipatory, adaptive, and systemic thinking in design, supporting real-time responses to evolving spatial and environmental conditions.
How does data influence the ways architects represent and communicate design ideas?	To analyze how data-driven workflows affect design language, tools, and drawing conventions.	Data-driven methods shift traditional representation toward algorithmic drawings, interactive models, and responsive visualization systems.
In what ways does data reshape architectural knowledge and authorship?	To explore the epistemological and disciplinary implications of integrating data in design.	The architect's role is evolving from only creator to strategist and interpreter of data, operating within hybrid human-machine design environments.

These insights align closely with the research questions that structure the inquiry, each of which is supported by specific objectives and evaluations. To further clarify the structure and progression of this research, the questions outlined in (Table 5.1) provide a conceptual and methodological framework through which the thesis unfolds. Each question is connected to specific research objectives and is addressed through different chapters of the thesis, offering a layered and contextual engagement across scales and domains.

One of the central concerns of the thesis has been to understand how big data is currently utilized in architectural design workflows. With the objective of identifying key applications, Chapter 4 presents examples such as simulation, performance modeling, and urban analysis, which reflect the growing reliance on predictive and data-rich processes in contemporary practice. These cases demonstrate that big data is already integrated into architectural workflows, although often unevenly, with varying degrees of critical reflection and epistemological grounding.

The second line of inquiry addresses the technical and conceptual challenges encountered when incorporating data analytics into design. This includes issues like limited interoperability between tools, lack of data literacy among practitioners, and concerns about authorship and creative autonomy. Chapter 2 introduces these as epistemological tensions, while Chapter 4 expands them in the context of evolving workflows. The objective here is to explore both the practical and cognitive barriers that hinder the full integration of data in architectural processes, highlighting the need for interdisciplinary competence and educational reform.

The thesis also investigates the potential of big data to shape future design logic. Chapter 4 illustrates how generative algorithms, adaptive systems, and real-time feedback loops foster a more systemic and anticipatory approach to design. This corresponds to the objective of envisioning new trajectories in data-driven architecture, where design becomes increasingly iterative, context-responsive, and process-based. The evaluation suggests that big data encourages a shift in mindset, from static form-making to dynamic environmental engagement.

Another key question focuses on the influence of data on architectural representation and communication. With the objective of analyzing how data alters design language and visualization tools, Chapters 3 and 4 examine emerging representational formats, including algorithmic drawings, interactive models, and responsive simulations. These modes move beyond conventional static representations, introducing new ways to engage with spatial information and communicate design intent across disciplines and audiences.

Finally, the thesis explores how the incorporation of data reshapes architectural knowledge and authorship. Rooted in the discussions of Chapter 2 and developed in Chapter 4, this question addresses the evolving role of the architect, from a singular author to a hybrid operator who negotiates between human judgment and algorithmic input. The evaluation highlights the emergence of hybrid authorship within human-machine design environments and reflects on the broader epistemological consequences of this shift.

Through this chapter-specific and objective-driven engagement, the thesis demonstrates how each question is not isolated but interrelated, offering overlapping perspectives on how big data is reconfiguring architectural thought and practice. Together, they construct a comprehensive inquiry into the contemporary role of data in design, bridging theory and application, and grounding the study's broader conclusions.

These interconnected responses to the research questions, distributed across the thesis chapters, provide the groundwork for drawing broader disciplinary reflections. By synthesizing these perspectives, the thesis identifies a continuous thematic pattern across the discussions: data is most productive when employed not merely as a tool of optimization but as a stimulant for reflective design thinking. The evolving design culture requires architects to actively negotiate between computational logic and creative intuition, fostering an expanded skill set that includes technological proficiency, data literacy, and ethical responsibility. This expanded role also involves reconsidering traditional notions of authorship, as design decisions increasingly emerge from the interplay between human judgment and algorithmic processes.

The broad contribution of this thesis lies in advocating for a conscious and balanced engagement with data in design processes. Data can enrich architectural thinking by enabling performance-based decisions, enhancing iterative exploration, and supporting more informed spatial interventions. Yet, without critical reflection, the adoption of data-driven methodologies risks negatively impacting the cultural, historical, and contextual sensitivities that define architectural identity.

In conclusion, this research suggests that the future of architectural design resides in hybridized workflows, where human creativity and computational analysis co-exist. Architects must navigate this evolving landscape by adopting dual competencies: the ability to decode and apply data intelligently, while preserving the narrative, experiential, and ethical dimensions of design practice. This balanced perspective is crucial in ensuring that the integration of data strengthens rather than diminishes architectural quality, enabling designers to contribute meaningfully to the built environment in the digital age.

5.2 Societal Impact and Contribution to Global Sustainability

The findings of this research are situated within the broader context of sustainable development and digital urban futures. Architectural design, as a spatial and cultural practice, holds significant potential in responding to global challenges such as urbanization, climate crisis, and resource scarcity. Big data analytics, when ethically and inclusively applied, enables designers and decision-makers to model energy usage, optimize environmental performance, assess walkability, and improve access to public services.

This thesis aligns particularly with UN Sustainable Development Goal 11: Sustainable Cities and Communities, by exploring how data-informed design can support livable, inclusive, and resilient urban environments. Furthermore, by emphasizing the role of architecture as both a technical and cultural system, the study highlights the importance of human-centered approaches to data. Data is never neutral; it carries assumptions, priorities, and consequences. Therefore, the value of Big Data in architecture must be measured not only by efficiency or automation but by its ability to foster equity, accessibility, and long-term adaptability.

Designers have the opportunity, and responsibility, to ensure that data-driven systems serve public interest rather than reinforcing bias or exclusion. From predictive models of environmental behavior to real-time participatory platforms, Big Data offers a set of tools that, if critically engaged, can advance more just and sustainable design practices.

5.3 Future Prospects

As architecture becomes increasingly entangled with computation, the future of design will likely be shaped by a continued combination between data science, artificial intelligence, and spatial thinking. This direction demands a reformulation of both education and practice. In design schools, curriculum must incorporate data literacy, coding basics, critical technology studies, and interdisciplinary collaboration. Within the profession, firms will need to build hybrid teams capable of interpreting, visualizing, and acting upon complex datasets.

One of the most promising directions lies in the integration of "big data" with "thick data", that is, the combination of large-scale quantitative inputs with qualitative understandings of people, culture, and context. This hybrid approach is critical to ensuring that data-driven design does not become reductionist or overly technical. The future of architecture may well depend on how successfully it can merge measurable insight with human experience.

Additionally, developments in AI-based design tools and machine learning algorithms present new roles for architects. Rather than being solely authors of form, designers may increasingly function as curators of data, orchestrators of systems, and narrators of intelligent environments. This repositioning offers an opportunity to reclaim creative agency while embracing the strategic potential of computational logic.

Future research may focus on:

- Designing open-source architectural datasets to ensure equitable access to design knowledge.
- Creating evaluation frameworks for ethical data use in urban environments.
- Developing participatory tools that incorporate citizen-generated data into the design process.
- Investigating how architectural authorship evolves under algorithmic co-creation.

In conclusion, the integration of Big Data into architecture is not a passing trend, it is a structural transformation. Whether in the studio, the city, or the cloud, data is now a core actor in shaping design logic, spatial agency, and architectural knowledge itself. The challenge for the next generation of architects is not just how to use data, but how to think with it, critically, ethically, and creatively.

BIBLIOGRAPHY

Abdelhameed, W. (2017). Creativity in the Initial Phases of Architectural Design. *Open House International*, *42*(1), 29–34. https://doi.org/10.1108/OHI-01-2017-B0005

Ahmeti, A., Hensel, D. S., Pruski, C., Tyc, J., & Hensel, M. (2025). Enabling Biodiversity-Informed Architecture Through Ontology-Driven Data Integration. *Applied Sciences*, *15*(10), Article 10. https://doi.org/10.3390/app15105311

Amabile, T. M. (2000). A Model of Creativity and Innovation in Organizations. *Research in Organizational Behavior*, *22*. https://www.hbs.edu/faculty/Pages/item.aspx?num=7441

Amoore, L. (2020). *Cloud ethics: Algorithms and the attributes of ourselves and others*. Duke University Press. https://doi.org/10.1515/9781478009276

Anderson, C. (2008a). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/

Anderson, C. (2008b). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/

As, I., Pal, S., & Basu, P. (2018). Artificial intelligence in architecture: Generating conceptual design via deep learning. *International Journal of Architectural Computing*, *16*(4), 306–327. https://doi.org/10.1177/1478077118800982

Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, *54*(15), 2787–2805. https://doi.org/10.1016/j.comnet.2010.05.010

Avanço, K., Giglia, E., & Gingold, A. (2021). *OPERAS-P Deliverable D6.3: Report on innovative approach to SSH FAIR data and publications*. https://zenodo.org/records/4817525

Badman, A., & Kosinski, M. (2024). *What is Big Data? | IBM*. https://www.ibm.com/think/topics/big-data

Baskarada, S., & Koronios, A. (2013). Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. *Australasian Journal of Information Systems*, *18*. https://doi.org/10.3127/ajis.v18i1.748

Beaty, R. E., Benedek, M., Silvia, P. J., & Schacter, D. L. (2016). Creative Cognition and Brain Network Dynamics. *Trends in Cognitive Sciences*, *20*(2), 87–95. https://doi.org/10.1016/j.tics.2015.10.004

Bellinger G., Castro D., & Mills A. (2004). *Data, Information, Knowledge, and Wisdom*. https://resources.saylor.org/wwwresources/archived/site/wp-content/uploads/2012/11/BUS206-4.1.1-Data-Information-Knowledge-and-Wisdom.pdf

Bernstein, P. G. (2018). *Architecture, design, data: Practice competency in the era of computation*. Birkhauser.

Bhadani, A., & Jothimani, D. (2017). *Big Data: Challenges, Opportunities and Realities* (No. arXiv:1705.04928). arXiv. https://doi.org/10.48550/arXiv.1705.04928

Bilton, C., & Cummings, S. (2010). *Creative strategy: Reconnecting business and innovation*. John Wiley & Sons. https://doi.org/10.1002/9781119208280

Biskjaer, M. M., Kamari, A., Jensen, S. R., & Kirkegaard, P. H. (2021). Exploring Blind Spots in Collaborative Value Creation in Building Design: A Creativity Perspective. *CoDesign: International Journal of CoCreation in Design and the Arts*, *17*(4), 374–391. https://doi.org/10.1080/15710882.2019.1654521

Bjögvinsson, E., Ehn, P., & Hillgren, P.-A. (2012). Design Things and Design Thinking: Contemporary Participatory Design Challenges. *Design Issues*, *28*(3), 101–116. https://doi.org/10.1162/DESI_a_00165

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd ed.). Routledge. https://doi.org/10.4324/9780203508527

Bordignon, F., & Maisonobe, M. (2022). Researchers and their data: A study based on the use of the word data in scholarly articles. *Quantitative Science Studies*, *3*(4), 1156–1178. https://doi.org/10.1162/qss_a_00220

Brown, T. (2019). *Change by Design, Revised and Updated: How Design Thinking Transforms Organizations and Inspires Innovation*.

Burry, M. (2011). *Scripting cultures: Architectural design and programming*. J. Wiley & Sons.

Caldas, L. G., & Norford, L. K. (2003). Genetic algorithms for optimization of building envelopes and the design and control of HVAC systems. *Journal of Solar Energy Engineering, Transactions of the ASME*, *125*(3), 343–351. Scopus. https://doi.org/10.1115/1.1591803

Cantamessa, M., Montagna, F., Altavilla, S., & Casagrande-Seretti, A. (2020a). Data-driven design: The new challenges of digitalization on product design and development. *Design Science*, *6*, e27. https://doi.org/10.1017/dsj.2020.25

Cantamessa, M., Montagna, F., Altavilla, S., & Casagrande-Seretti, A. (2020b). Data-driven design: The new challenges of digitalization on product design and development. *Design Science*, *6*, e27. https://doi.org/10.1017/dsj.2020.25

Carpo, M. (2023). *Beyond digital: Design and automation at the end of modernity*. The MIT Press.

Castro Pena, M. L., Carballal, A., Rodríguez-Fernández, N., Santos, I., & Romero, J. (2021). Artificial intelligence applied to conceptual design. A review of its use in architecture. *Automation in Construction*, *124*, 103550. https://doi.org/10.1016/j.autcon.2021.103550

Chaffey, D., & Wood, S. (2005). *Business information management: Improving performance using information systems*. Financial Times Prentice Hall. http://www.ub.unibas.ch/tox/HBZ/HT014199141/PDF

Chen, Y. (2021). Application of Big Data and Artificial Intelligence in Architectural Design. *2021 4th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA)*, 274–277. https://doi.org/10.1109/ISTTCA53489.2021.9654515

Choo, C. W. (2005). *The Knowing Organization*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195176780.001.0001

Clark. (2004). *Talent & Performance*. http://www.nwlink.com/~donclark/performance/understanding.html

Cleveland, H. (1982). Information As a Resource. *Futurist*, *16*(6), 34–39.

Cox, M., & Ellsworth, D. (1997). *Managing big data for scientific visualization*.

Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

Cropley, A. (2015). *Creativity in Education and Learning: A Guide for Teachers and Educators*. Routledge. https://doi.org/10.4324/9780203826270

Cross, N. (2006). *Designerly Ways of Knowing*. Springer-Verlag. https://doi.org/10.1007/1-84628-301-9

Cross, N. (2016). *Design thinking: Understanding how designers think and work* (Reprinted). Bloomsbury Academic.

Diebold, F. X. (2012). *A Personal Perspective on the Origin(s) and Development of "Big Data": The Phenomenon, the Term, and the Discipline, Second Version* (SSRN Scholarly Paper No. 2202843). Social Science Research Network. https://doi.org/10.2139/ssrn.2202843

Duarte, F. (2025). *Amount of Data Created Daily (2025)*. Exploding Topics. https://explodingtopics.com/blog/data-generated-per-day

Ekici, B., Cubukcuoglu, C., Turrin, M., & Sariyildiz, I. S. (2019). Performative computational architecture using swarm and evolutionary optimisation: A review. *Building and Environment*, *147*, 356–371. https://doi.org/10.1016/j.buildenv.2018.10.023

Elragal, A., & Klischewski, R. (2017). Theory-driven or process-driven prediction? Epistemological challenges of big data analytics. *Journal of Big Data*, *4*. https://doi.org/10.1186/s40537-017-0079-2

Esterbrook, S. (2012). *What is Climate Informatics?* https://www.easterbrook.ca/steve/2012/09/what-is-climate-informatics/

Evins, R. (2013). A review of computational optimisation methods applied to sustainable building design. *Renewable and Sustainable Energy Reviews*, *22*, 230–245. https://doi.org/10.1016/j.rser.2013.02.004

Favaretto, M., Clercq, E. D., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLOS ONE*, *15*(2), e0228987. https://doi.org/10.1371/journal.pone.0228987

Floridi, L. (2010). *Information: A very short introduction*. Oxford University press.

Floridi, L. (2011). The logical unsolvability of the Gettier problem. In L. Floridi (Ed.), *The Philosophy of Information* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199232383.003.0009

Floridi, L. (2014). *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press.

Frické, M. (2015). Big data and its epistemology. *Journal of the Association for Information Science and Technology*, *66*(4), 651–661. https://doi.org/10.1002/asi.23212

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*(2), 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Geertz, C. (1973). Thick Description: Toward an Interpretive Theory of Culture. In *The Interpretation of Cultures* (pp. 311–323). Basic Books.

Gitelman, L. (Ed.). (2013). *"Raw data" is an oxymoron*. The MIT Press.

Gray, J., Gerlitz, C., & Bounegru, L. (2018). Data infrastructure literacy. *Big Data & Society*, *5*(2), 2053951718786316. https://doi.org/10.1177/2053951718786316

Groat, L. N., & Wang, D. (2013). *Architectural Research Methods* (2 nd). John Wiley & Sons.

Gruber, M., de Leon, N., George, G., & Thompson, P. (2015). Managing by Design. *Academy of Management Journal*, *58*(1), 1–7. https://doi.org/10.5465/amj.2015.4001

Guilford, J. P. (1950). Creativity. *American Psychologist*, *5*(9), 444–454. https://doi.org/10.1037/h0063487

Harrison, A. K. (2013). Thick Description. In J. R. McGee & R. L. Warms (Eds.), *Theory in Social and Cultural Anthropology: An Encyclopedia* (pp. 860–861). SAGE Publications, Inc. https://doi.org/10.4135/9781452276311.n281

Hensel, M., & Bier, H. (2022). Interdisciplinary Data-integrated Approaches. *SPOOL*, *9*, 3–4. https://doi.org/10.47982/spool.2022.1.00

Hensel, M., Tyc, J., Ahmeti, A., & Sunguroglu Hensel, D. (2025). An Ontology-Aided Generative Computational Design Process for Ecological Building Envelopes: Development of a Conceptual Framework for the Early Design Stage. *Land*, *14*, 840. https://doi.org/10.3390/land14040840

Hettithanthri, U., Hansen, P., & Munasinghe, H. (2023). Exploring the architectural design process assisted in conventional design studio: A systematic literature review. *International Journal of Technology and Design Education*, *33*(5), 1835–1859. https://doi.org/10.1007/s10798-022-09792-9

Hjørland, B. (1998). Theory and metatheory of information science: A new interpretation. *Journal of Documentation*, *54*. https://doi.org/10.1108/EUM0000000007183

Hois, J., Bhatt, M., & Kutz, O. (2009). Modular Ontologies for Architectural Design. In *Frontiers in Artificial Intelligence and Applications* (Vol. 198, p. 77). https://doi.org/10.3233/978-1-60750-047-6-66

Holl, S., Pallasmaa, J., & Pérez-Gómez, A. (2008). *Questions of perception: Phenomenology of architecture*. William Stout Pub.

Hume, D. (1979). *Enquiries concerning human understanding and concerning the principles of morals* (L. A. Selby-Bigge & P. H. Nidditch, Eds.; Repr. from the posthumous ed. of 1777. 3. ed. repr). Clarendon Pr.

Jacobs, J., & Merriman, P. (2011). Practising architectures. *Social & Cultural Geography - SOC CULT GEOGR*, *12*, 211–222. https://doi.org/10.1080/14649365.2011.565884

Jang, S., Roh, H., & Lee, G. (2025). Generative AI in architectural design: Application, data, and evaluation methods. *Automation in Construction*, *174*, 106174. https://doi.org/10.1016/j.autcon.2025.106174

Jin, X., & Dong, H. (2020). New Design Heuristics In The Digital Era. *Proceedings of the Design Society: DESIGN Conference*, *1*, 607–616. https://doi.org/10.1017/dsd.2020.321

Kalay, Y. E. (2004). *Architecture's new media: Principles, theories, and methods of computer-aided design*. MIT Press.

Kitchin, R. (2014a). Big Data, New Epistemologies and Paradigm Shift. *Big Data & Society*, *1*, 1–12. https://doi.org/10.1177/2053951714528481

Kitchin, R. (2014b). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

Kolarevic, B. (2005). *Architecture in the digital age: Design and manufacturing*. Taylor & Francis.

Kun, P., Mulder, I., & Kortuem, G. (2018). *Design Enquiry Through Data: Appropriating a Data Science Workflow for the Design Process*. https://doi.org/10.14236/ewic/HCI2018.32

Kvan, T., & Jia, Y. (2005). Students' learning styles and their correlation with performance in architectural design studio. *Design Studies*, *26*(1), 19–34. https://doi.org/10.1016/j.destud.2004.06.004

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Latour, B., & Yaneva, A. (2008). «Give Me a Gun and I Will Make All Buildings Move»: An ANT's View of Architecture. *Ardeth*, *1*. https://doi.org/10.17454/ARDETH01.08

Lawrence, K. (2013). *Developing Leaders in a VUCA Environment* [Organizational Report]. UNC Executive Development. https://www.emergingrnleader.com/wp-content/uploads/2013/02/developing-leaders-in-a-vuca-environment.pdf

Lawson, B. (2006). How Designers Think – The Design Process Demystified. *University Press, Cambridge*.

Liew, A. (2013). DIKIW: Data, Information, Knowledge, Intelligence, Wisdom and their Interrelationships. *Business Management Dynamics*.

Lindberg, T., Meinel, C., & Wagner, R. (2011). Design Thinking: A Fruitful Concept for IT Development? In C. Meinel, L. Leifer, & H. Plattner (Eds.), *Design Thinking* (pp. 3–18). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-13757-0_1

Loukissas, Y. A., & Bowker, G. C. (2019). *All data are local: Thinking critically in a data-driven society*. The MIT Press.

Mayer-Schönberger, V., & Cukier, K. (2014). *Big data: A revolution that will transform how we live, work, and think* (First Mariner Books edition). Mariner Books, Houghton Mifflin Harcourt.

McElroy, M. (2003). The new knowledge management: Complexity, learning, and sustainable innovation. *Inf. Res.*, *8*.

Meekings, S., & Schnabel, M. A. (2016). Big Data on Individuals in the Architectural Design Process: Combining Individual's Data with the Architects Toolset. *International Journal of Parallel Emergent and Distributed Systems*, *31*, 13–20. https://doi.org/10.1080/17445760.2017.1390091

Melendez, F., Diniz, N., & Signore, M. D. (Eds.). (2020). *Data, Matter, Design: Strategies in Computational Design*. Routledge. https://doi.org/10.4324/9780367369156

Merriam-Webster. (2025, March 24). Merriam-Webster.Com Dictionary. https://www.merriam-webster.com/dictionary/data

Micheli, P., Wilner, S. J. S., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing Design Thinking: Conceptual Review, Synthesis, and Research Agenda. *Journal of Product Innovation Management*, *36*(2), 124–148. https://doi.org/10.1111/jpim.12466

Mortati, M., Magistretti, S., Cautela, C., & Dell'Era, C. (2023). Data in design: How big data and thick data inform design thinking projects. *Technovation*, *122*, 102688. https://doi.org/10.1016/j.technovation.2022.102688

Néroulidis, A., Pouyet, T., Tournon, S., Rousset, M., Callieri, M., Manuel, A., Abergel, V., Malavergne, O., Cao, I., Roussel, R., Granier, X., Rodier, X., & De Luca, L. (2024). A digital platform for the centralization and long-term preservation of multidisciplinary scientific data belonging to the Notre Dame de Paris scientific action. *Journal of Cultural Heritage*, *65*, 210–220. https://doi.org/10.1016/j.culher.2023.09.016

Nicholson, S., Tomasi, M., Belleri, D., & Ratti, C. (2022). 'Greening' the Cities: How Data Can Drive Interdisciplinary Connections to Foster Ecological Solutions. *SPOOL*, *9*(1), 5–18. https://doi.org/10.47982/spool.2022.1.01

NIST Big Data Public Working Group, Definitions and Taxonomies Subgroup. (2019). *NIST Big Data Interoperability Framework: Volume 1, definitions version 3* (No. NIST SP 1500-1r2; p. NIST SP 1500-1r2). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.1500-1r2

Oracle Corporation. (2019). *Big Data and Enterprise Data, Bridging Two Worlds with Oracle Data Integration*. Oracle.

Oxman, R. (2006). Theory and design in the first digital age. *Design Studies*, *27*(3), 229–265. https://doi.org/10.1016/j.destud.2005.11.002

Oxman, R. (2008). Performance-Based Design: Current Practices and Research Issues. *International Journal of Architectural Computing*, *6*(1), 1–17. https://doi.org/10.1260/147807708784640090

Oxman, R. (2017). Thinking difference: Theories and models of parametric design thinking. *Design Studies*, *52*, 4–39. https://doi.org/10.1016/j.destud.2017.06.001

Pallasmaa, J. (1996). The Geometry of Feeling: A look at the phenomenology of architecture. In K. Nesbitt (Ed.), *Theorizing a New Agenda for Architecture, An Anthology of Architectural Theory* (pp. 447–453). Princeton Architectural Press.

Panagoulia, E. (2023). *Data Governance and Control For Performance Based Design. An Aspect System for Data Reliability And Completeness*. https://hdl.handle.net/1853/75585

Peters, M., Jandric, P., & Green, B. (2024). The DIKW Model in the Age of Artificial Intelligence. *Postdigital Science and Education*. https://doi.org/10.1007/s42438-024-00462-8

Picon, A. (2011). Architecture and the Virtual Towards a new Materiality? *Praxis: Journal of Writing + Building*.

Qabshoqa, M. (2017). *Conceptual Operational Model of Architecture: An approach for capturing values in architectural practices based on Big Data capabilities* [Doctoral Thesis]. University of Liverpool.

Quinones-Gomez, J. C. (2017). The growing influence of design data in the design process through a methodological development. *R. Valušytė, A. Biamonti, & C. Cautela (Eds.) D*, *4*, 178–187.

Quiñones-Gómez, J. C. (2019). Moving Away from the Basic, Adopting a New Approach to the Creative Process. In F. Cavas-Martínez, B. Eynard, F. J. Fernández Cañavate, D. G. Fernández-Pacheco, P. Morer, & V. Nigrelli (Eds.), *Advances on Mechanics, Design Engineering and Manufacturing II* (pp. 670–679). Springer International Publishing. https://doi.org/10.1007/978-3-030-12346-8_65

Quinones-Gomez, J. C. (2021). Creativity Forward: A Framework that Integrates Data Analysis Techniques to Foster Creativity within the Creative Process in User Experience Contexts. *Creativity Studies*, *14*(1), 51–73. https://doi.org/10.3846/cs.2021.12933

Quiñones-Gómez, J. C., Mor, E., & Chacón, J. (2024). Data-Driven Design in the Design Process: A Systematic Literature Review on Challenges and Opportunities. *International Journal of Human-Computer Interaction*, *41*, 1–26. https://doi.org/10.1080/10447318.2024.2318060

Quinones-Gomez, J. C. Q. (2018). Supporting the creative process from data. *CERN IdeaSquare Journal of Experimental Innovation*, *2*(2), Article 2. https://doi.org/10.23726/cij.2018.736

Rahbarianyazd, R., & Nia, H. A. (2019). International Journal of Cognitive Research in Science, Engineering and Education (IJCRSEE). *International Journal of Cognitive Research in Science, Engineering and Education*, *10*(2), Article 3. https://doi.org/10.5937/IJCRSEE1903061R

Rescher, N. (2003). Epistemology: An Introduction to the Theory of Knowledge. In *Epistemology: An Introduction to the Theory of Knowledge*.

Riswadkar, M. R. (2006). *Data: Definition, Types, Nature, Properties and Scope* [Educational Report]. Indira Gandhi National Open University. https://egyankosh.ac.in/handle/123456789/10935

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, *4*(2), 155–169. https://doi.org/10.1007/BF01405730

Rosenberg, D. (2013). Data before the Fact. In L. Gitelman (Ed.), *“Raw Data” Is an Oxymoron* (pp. 15–40). The MIT Press. https://doi.org/10.7551/mitpress/9302.003.0003

Rowley, J. (2006). Where is the wisdom that we have lost in knowledge? *Journal of Documentation*, *62*(2), 251–270. https://doi.org/10.1108/00220410610653320

Rowley, J. (2007). The wisdom hierarchy: Representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. https://doi.org/10.1177/0165551506070706

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal*, *24*(1), 92–96. https://doi.org/10.1080/10400419.2012.650092

Russell L. Ackoff. (1989). *From Data to Wisdom*. https://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf

Schöch, C. (2013). » Big? Smart? Clean? Messy? Data in the Humanities Journal of Digital Humanities. *Journal of Digital Humanities*, *2*(3). https://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/

Schön, D. A. (2017). *The Reflective Practitioner* (0 ed.). Routledge. https://doi.org/10.4324/9781315237473

Simsek, Z., Vaara, E., Paruchuri, S., Nadkarni, S., & Shaw, J. (2019). New Ways of Seeing Big Data. *Academy of Management Journal*, *62*, 971–978. https://doi.org/10.5465/amj.2019.4004

Snowden, D. (2002). Complex acts of knowing: Paradox and descriptive self‐awareness. *Journal of Knowledge Management*, *6*(2), 100–111. https://doi.org/10.1108/13673270210424639

Snowden, David J. & Boone, Mary E. (2007). *A Leader's Framework for Decision Making*. https://hbr.org/2007/11/a-leaders-framework-for-decision-making

Soliman, A. (2017). Appropriate teaching and learning strategies for the architectural design process in pedagogic design studios. *Frontiers of Architectural Research*, *6*. https://doi.org/10.1016/j.foar.2017.03.002

Sönmez, N. O. (2018). A review of the use of examples for automating architectural design tasks. *Computer-Aided Design*, *96*, 13–30. https://doi.org/10.1016/j.cad.2017.10.005

Speed, C., & Oberlander, J. (2016, June 27). *Designing from, with and by Data: Introducing the ablative framework*. DRS2016: Future–Focused Thinking. https://doi.org/10.21606/drs.2016.433

Sternberg, R. J., & Lubart, T. I. (1999). The concept of creativity: Prospects and paradigms. In *Handbook of creativity* (pp. 3–15). Cambridge University Press.

Sun, Z., Pambel, F., & Wu, Z. (2022). *The Elements of Intelligent Business Analytics: Principles, Techniques, and Tools* (pp. 1–20). https://doi.org/10.4018/978-1-7998-9016-4.ch001

Sunguroğlu Hensel, D., Tyc, J., & Hensel, M. (2022). Data-driven design for Architecture and Environment Integration. *SPOOL*, *9*(2), 19–34. https://doi.org/10.47982/spool.2022.1.02

Takagi, H. (2001). Interactive evolutionary computation: Fusion of the capabilities of EC optimization and human evaluation. *Proceedings of the IEEE*, *89*(9), 1275–1296. https://doi.org/10.1109/5.949485

Todella, E. (2023). The architectural design practice in the folds of decision-making processes: La pratica della progettazione architettonica nelle pieghe dei processi decisionali. *Valori e Valutazioni*, *33*, 3–16. https://doi.org/10.48264/VVSIEV-20233302

Trapold, E., & Saldana Ochoa, K. (2023). Deriving Architectural Inspiration from Big Data. *UF Journal of Undergraduate Research*, *25*. https://doi.org/10.32473/ufjur.25.133060

Uzunkaya, A., & Paker Kahvecioğlu, N. (2022). Architectural design research through reflection: A sub-approach under "research by design." *Open House International*, *47*(4), 688–709. Scopus. https://doi.org/10.1108/OHI-07-2021-0155

Van Ameijde, J., & Leung, C. K. S. (2022). *UAV-based People Location Tracking and Analysis for the Data-Driven Assessment of Social Activities in Public Spaces*. 293–302. https://doi.org/10.52842/conf.caadria.2022.1.293

Van Geest, M., Tekinerdogan, B., & Catal, C. (2021). Design of a reference architecture for developing smart warehouses in industry 4.0. *Computers in Industry*, *124*, 103343. https://doi.org/10.1016/j.compind.2020.103343

Vehlken, S. (2014). Computational swarming: A cultural technique for generative architecture. *Footprint*, *15*, 9–24. Scopus.

Vitruvius. (1960). *The ten books on architecture* (M. H. Morgan, Trans.; New Dover ed., unabridged and unaltered republ. of the 1. ed. 1914). Dover.

Ward, J. S., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions* (No. arXiv:1309.5821). arXiv. https://doi.org/10.48550/arXiv.1309.5821

Weiss, S. M., & Indurkhya, N. (1998). *Predictive data mining: A practical guide* (Nachdr.). Morgan Kaufmann.

Wiggins, C., & Jones, M. L. (2023). *How data happened: A history from the age of reason to the age of algorithms* (First edition). W. W. Norton & Company.

Yaneva, A. (2009). *The Making of a Building: A pragmatist approach to architecture*. Peter Lang. https://research.manchester.ac.uk/en/publications/the-making-of-a-building-a-pragmatist-approach-to-architecture

Zhu, Y. (2016). Multi-Modality Human Action Recognition. *Graduate Theses, Dissertations, and Problem Reports*. https://doi.org/10.33915/etd.7054

Zins, C. (2006). Redefining information science: From "information science" to "knowledge science." *Journal of Documentation*, *62*, 447–461. https://doi.org/10.1108/00220410610673846

Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. *JASIST*, *58*, 479–493. https://doi.org/10.1002/asi.20508

Zins, C. (2011). Redefining information science: From "information science" to the "science of knowledge." *Informacao e Sociedade*, *21*, 155–167.