Building Information - Representation and Management - Fundamentals and principles

Alexander Koutamanis

TU Delft open

Building information - representation and management

Building information - representation and management

Fundamentals and principles

ALEXANDER KOUTAMANIS

Building information - representation and management by Delft University of Technology is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](), except where otherwise noted.

Cover image by the author is licensed under CC BY-NC-SA 4.0
Every attempt has been made to ensure the correct source of images and other potentially copyrighted material
was ascertained, and that all materials included in this book has been attributed and used according to its
license. If you believe that a portion of the material infringes someone else's copyright, please the author
directly on: A.Koutamanis@tudelft.nl

ISBN hardcopy: 978-94-6366-159-1
ISBN ebook: 978-94-6366-160-7
DOI: https://doi.org/10.5074/T.2019.003

Contents

Part I. Digitization

Part II. Building representation

Part III. Information: theory and management

Part IV. Exercises

Preface

Perhaps typically of me, this book was triggered by questions I was asked over the years about information, representation, digitization and management. The more I quoted standard answers from standard literature, the more restless I became because I perceived a lack of coherence in my answers. There seemed to be too many holes and grey areas, and, rather more worryingly, too few connections between the various parts of the underlying body of knowledge.

This led to a number of fundamental questions I had to ask myself before attempting to answer what others asked me. I tried to understand one by one the multiple layers and aspects involved in the phenomena that intrigued me, without losing sight of the whole. Thankfully, I was able to find enough enlightenment on these layers and aspects. There have been quite a few clever people who have attacked the same questions before me and managed to come up with convincing answers. My own contribution lies primarily in the interpretation of their theories and the connections I suggest between them and with the domain of buildings.

Note that in contrast to earlier publications of mine, I talk about buildings rather than architecture. The reason for doing so is that buildings and built environments have a larger scope than architecture, as suggested by the relation between the Dutch terms ‘bouwkunde’ and ‘architectuur’: the latter is a specialization within the former. It is unfortunate that both are translated into English as ‘architecture’ (the less said about terms like ‘building science’ the better).

I am grateful to the people who formulated the theories discussed in this book. I have learned a lot from them. In a more practical sense, I was also assisted by a number of people who merit a mention and my profound thanks: Monique de Bont for the meticulous copyright control; Saskia Roselaar for her thorough and insightful proofreading; Michiel de Jong for giving me the opportunity to publish this book as an open textbook and for managing every step of the production process. Polyxeni Mantzou, Paul Chan and Thanos Economou reviewed the book. I am indebted to them for their time and constructive criticism.

On May 1, 2019, a group of students who follow the Information Management course in the MBE master track at the Faculty of Architecture & the Built Environment (Faculteit Bouwkunde, in Dutch), Delft University of Technology, will be the first to use this open textbook. I hope they will enjoy working with it and thank them in advance for their tolerance of any mistakes that may have slipped through in this first edition.

A.K.

Delft, 19.02.2019

Introduction

This book is about the foundations and principles of building information, its representation and management. It does not tell you which software or policies to choose for representing buildings and managing the resulting information. In fact, the book argues that one should not start with practical steps before fully understanding the reasoning behind any such choice. The basis of this reasoning comprises the structure of information and of the representations that contain it, the purposes of managing information in the context of these representations and the processes in which the representations are used; in a nutshell, how information relates to a specific domain. Without adequate reasoning that takes into account all syntactic, semantic and pragmatic aspects, adopting one software or another, implementing one practical measure or another simply subjugates information processing to some prescriptive or proscriptive framework that may be unproductive, incompatible or inappropriate for the domain.

To explain these foundations and principles, the book brings together knowledge from various areas, including philosophy and computer science. Its perspective, nevertheless, remains bounded by the application domain: external knowledge is not imposed on domain practices but used to elucidate domain knowledge. Building information has its own peculiarities, drawn more from convention than necessity, and digitization has yet to address such matters, let alone resolve them. General knowledge about information and representation can be of assistance in developing approaches fit for the digital era. The approach advocated in this book is above all parsimonious: in a world inundated with digital information (Chapter 1), one should not resort to brute force and store or process everything. Instead, one should organize information intelligently, so that everything remains accessible but with less and more focused effort.

The first part of the book focuses on representation: many of the problems surrounding information and its management stem from a lack of understanding that most information, certainly regarding buildings, comes organized into representations. Knowing the structure of these representations provides insights into how information is produced and processed. Chapter 2 explains symbolic representations and analyses familiar spatial representations from the symbolic perspective. The analogue representations that still dominate building information are the subject of Chapter 3. Digitization is primarily considered with respect to BIM, as the first generation of truly symbolic digital building representations (Chapter 4).

Information theory and management are the subjects of the second part of the book. Particular emphasis is on the meaning of information (semantics) as a foundation for utility and relevance. For this reason, this part starts by introducing a semantic theory of information that complements symbolic representation (Chapter 5). Next, Chapter 6 explains information management and how it applies to building information and BIM, concluding with the principles that should guide building information management. Chapter 7 rounds off the second part by explaining how one can represent processes and the information contained in them.

Having explained the foundations and principles of representation and information management, the book concludes with some larger exercises, which can be used as individual or group assignments. Through these exercises, readers can test their understanding of the approach advocated in this book and hone their skills for its application in research or practice.

List of terms and abbreviations

A

AECO: architecture, engineering, construction and operation of buildings

Arc (graphs): directed edges

B

BIM: building information modelling

BIM checker: computer program in which one primarily views and analyses a model

BIM editor: computer program in which one primarily develops and modifies a model

Bridge (graphs): an edge that divides a graph into two unconnected parts

C

CAAD (computer-aided architectural design): the discipline covering all aspects of computerization in AECO

CAD (computer-aided design): a category of software primarily aimed at the computerization of design representations, including engineering drawings (2D) and models (3D)

Center of graph: the vertices with an eccentricity equal to the radius of the graph

Closeness of a vertex (graphs): its mean distance to all other vertices in the graph

Connected graph: a graph in which each vertex connects to every other vertex by some sequence of edges and vertices

Co-termination: the condition of two entities (e.g. walls) having a common endpoint

D

Degree of a vertex (graphs): the number of edges connected to it

Degree sequence (graphs): sequence obtained by listing the degrees of vertices in a graph

DM: design management

Diameter (graphs): the greatest eccentricity of any vertex in a graph

Directed graph (or digraph): a graph where edges have a direction (arcs)

Distance (graphs): the number of edges in the shortest path between two vertices

E

Eccentricity (graphs): the greatest distance between a vertex and any other vertex in a graph

Edge (graphs): usually a relation between two things (represented as nodes)

Exabyte = a million Terabytes = a billion Gigabytes = 10^{18} bytes

G

Graphs: mathematical structures that describe pairwise relations between things

I

IFC (Industry Foundation Classes): a standard underlying BIM

IM: information management

IoT: Internet of things

L

LoD: level of development (or detail) in BIM

M

MEP: mechanical, electrical and plumbing

Moore's "law": the number of transistors on a chip doubles every year while the costs are halved

MTC: mathematical theory of communication, proposed by Claude Shannon

N

Node (graph): synonym of vertex

P

Path (graphs): a sequence of edges and vertices in which no vertex occurs more than once

Periphery of a graph: the vertices with an eccentricity equal to the diameter of the graph

PDF: portable document format

R

Radius (graphs): the smallest eccentricity of any vertex in a graph

V

Vertex (graphs): usually the representation of a thing

Z

Zettabyte = a thousand Exabytes = a billion Terabytes = a trillion Gigabytes = 10^{21} bytes

PART I

DIGITIZATION

1. Digital information

This chapter presents the key challenges AECO is facing with the digitization of information and outlines the content of this book with respect to these challenges.

Information explosion

One of the key characteristics of our era is the explosive increase in information production and registration. It has been estimated that human societies had accumulated roughly 12 exabytes until the digital era. Then, annual information growth rates of 30% raised the total to 180 exabytes by 2006 and to over 1.800 exabytes by 2011. In the most recent period, the total more than doubled every two years, towards a projected 44 zettabytes by 2020 and 180 zettabytes by 2025.[1]

Such astounding calculations are updated regularly, with even more dramatic projections, so future totals may become even higher. The main reason for this is that the population of information users and producers keeps on increasing and is currently expanding to cover devices generating and sharing data on the Internet of Things. But even if we reach a plateau at some point, as with Moore's "law" concerning the growth of computing capacity,[2] we already have an enormous problem in our hands.

The situation is further complicated by changing attitudes concerning information. Not so long ago, most people were afraid of information overload.[3] Nowadays with the general excitement about big data we have moved to the opposite view. From being a worry, the plethora of information we produce and consume has become an opportunity. Attitudes may change further, moreover in unpredictable ways, as suggested by reactions to the Facebook – Cambridge Analytica data breach in 2018.

Regardless of such attitudes, two things will not change. The first is that we have to manage existing information efficiently, effectively, securely and safely. The second is that the means of information production, dissemination and management will remain primarily digital.

Information explosion in AECO

The explosive growth of digital information relates to AECO in various ways. On one end of the spectrum, we have new information sources that produce big data, such as smartphones and sensors. These tell us a lot about users and conditions in the built environment, and so promise a huge potential for the analysis and improvement of building performance, while requiring substantial investment in technologies and organization. On the other end of the spectrum, there are established information and communication technologies that have already become commonplace and ubiquitous, also in AECO. Email, for instance, appears to dominate communication and information exchange[4] by offering a digital equivalent to analogue practices like letter writing. Replication of analogue practices that dominate digital information processing is typical of AECO: digital technologies and information standards are still geared towards the production of conventional documents like floor plans and sections.

In between these two extremes, we encounter domain-specific technologies that aim to structure AECO processes and knowledge. Currently paramount among these is BIM, an integrated approach that is usually justified with respect to performance.[5] Performance improvement through BIM involves intensive and extensive collaboration, which adds to both the importance and the burden of information. The wide adoption of BIM means rapid expansion to cover more aspects and larger projects, which accentuates interoperability, capacity and coordination problems. In a recent survey, 70% of AECO professionals claim that project information deluge actually impedes effective collaboration, while 42% feel unable to integrate new digital tools in their organizations.[6] This surely impedes the deployment of solutions to their information needs: AECO appears to share many of the problems of the digital information explosion, yet to profit relatively little from the information-processing opportunities of the digital era.

Digitization in AECO: origins and outcomes

AECO has always been an intensive producer and consumer of information. In fact, most of its disciplines primarily produce information on buildings rather than buildings, e.g. drawings and related documents that specify what should be constructed and how. Especially drawings have been a major commodity in AECO;

they are ubiquitous in all forms of specification and communication, and quite effective in supporting all kinds of AECO tasks.

The history of digitization in AECO starts quite early, already in the 1960s, but with disparate ambitions. Some researchers were interested in automating design (even to the extent of replacing human designers with computers), while others were keen to computerize drawing. The two coexisted in the area of CAAD, with design automation been generally treated as the real goal. With the popularization of computers in the 1990s, however, it was computerized drawing that became popular in AECO practice.

As with other software, the primary use of computerized drawing systems has been the production of analogue documents: conventional drawings like floor plans and sections on paper. For many years, the advantages of computerized drawing were presented in terms of efficiency improvement over drawing by hand on paper: faster production of drawings, easier modification and compact storage. Even after the popularization of the Internet, the emphasis on conventional documents remained; the only difference was that, rather than producing and exchanging paper-based documents, one would produce and exchange digital files like PDFs.

A main consequence of this has been that AECO remained firmly entrenched in conventional, document-based processes. While other analogue documents like telephone directories were being replaced by online information systems and apps, and people adapted to having their day planners and address lists on mobile phones, AECO stubbornly stuck to analogue practices and documents, prolonging their life into the digital era.

BIM: radical intentions

Drawing from product modelling, BIM emerged as a radical improvement of computerized drawing that should provide a closer relation to design. The difference with earlier design automation attempts was that it did not offer prescriptive means for generating a design but descriptive support to design processes: collaboration between AECO disciplines, integration of aspects and smooth transition between phases. By doing so, it shifted attention from drawings to the information they contained.

The wide acceptance of BIM is unprecedented in AECO computerization. Earlier attempts at computerization were often met with reluctance, not in the least for the cost of hardware, software and training to use them. The reception of BIM, by contrast, was much more positive, even though it was more demanding than its predecessors in terms of cost. Arguably more than its attention to information or collaboration, it was its apparent simplicity (a Lego-like assembly of a building) that made it appealing, especially to non-technical stakeholders. The arcane conventions and practices of analogue drawing no longer seemed necessary or relevant.

Nevertheless, BIM remained rooted in such conventions. It may have moved from the graphic to the symbolic but it did so through interfaces laden with graphic conventions. For example, entering a wall in BIM may be done in a floor plan projection as follows: the user selects the wall type and then draws a line to indicate its axis. As soon as the axis is drawn, the wall symbol appears fully detailed according to the wall type that has been chosen: lines, hatches and other graphic elements indicating the wall materials and layers. The axis is normally not among the visible graphic elements. Such attachment to convention makes it rather hard for users to understand that they are actually entering a symbol in the model rather than somehow generating a drawing.

More on such matters follows later in the book. For the moment, it suffices to note that BIM may indicate a step forward in the digitization of AECO information but it remains a hybrid environment that may confuse or obscure fundamental information issues. As such, it deserves particular attention and, being the best option for AECO for the moment, it is used as the main information environment discussed in this book. Future technologies are expected to follow the symbolic character of BIM, so any solutions developed on the basis of BIM will probably remain applicable.

Representation

Ideas about information and how it works can be vague or even confusing if one fails to realize that most of it is not unstructured or haphazard but organized in meaningful representations. These representations allow us to understand and utilize information effectively and economically. Consequently, they are critical for both information and digitization. As intensive but generally intuitive users of representations, we have to become aware of their structure and characteristics in order to understand how we process and disseminate information. We also have to appreciate that existing representations are not necessarily appropriate for the computer era. Computers have different capacities to humans, therefore familiar representations we have been using successfully for centuries may have to be adapted or even abandoned.

This is evident in changes that have already occurred but are not always apparent, even to avid computer users. Anticipating the following chapters on representation, let us consider just one example of the effects of computerization: humans mostly use decimal numbers, arguably because we have ten fingers to help us with calculations, while computers use binary numbers because they are built out of components with two possible states (on an off). Humans are capable of using binary numbers but they require significantly more effort than decimal ones. As a result, while computers use binary numbers, user interfaces translate them into decimal ones. Despite the added burden of having to employ and connect two different representations, this solution works well for the symbiosis of computers and humans.

In dealing with information, one must therefore be aware of all representations involved, their connections and utility. This is a prerequisite to effective and reliable computerization, e.g. concerning the role and operation of interfaces. The same applies to the treatment of digital information: knowing the characteristics of a representation allows one to ascertain which data are well-formed and meaningful in the particular context.

Information management

Managing information is not just a task for managers and computer specialists. It involves everyone who disseminates, receives or stores information. Very few people are concerned with information management just for the sake of it; most approach information and its management in the framework of their own activities, for which information is an essential commodity. This makes management of information not an alien, externally imposed obligation but a key aspect of everyone’s information processing, a fundamental element in communication and collaboration, and a joint responsibility for all those involved. Given the amounts of information currently produced and exchanged, its careful management is a necessity for anyone who relies on information for their functioning or livelihood.

For these reasons, in this book we view management issues from two complementary perspectives: that of design management, as representative of all management, coordination and collaboration activities in AECO, and that of generic information management, not restricted to AECO, as a source of generally applicable principles and guidelines. As we shall see, the one depends on the other for providing a suitable solution to information management problems. As with all aspects of this book, emphasis is not on technical solutions but on the conceptual and operational structure of information management: the definition of clear approaches and transparent criteria for guiding people to a better performance and selecting or evaluating means that support them towards this goal.

The reasons for doing so are already rather pressing. Despite the broad acknowledgement of the information deluge in AECO, the development of effective IM approaches appears to be lagging behind. Information may hold a central position in AECO computerization, as the "I" in BIM testifies, yet IM in AECO is generally poorly specified as an abstract, background obligation in management — as something that additional computer systems should solve or as a reason to create additional management roles, such as project information managers, BIM and CAD managers and coordinators, so as to cover the increased technical complexity (not just quantity) of digital information. Such new computer systems and technical specializations nevertheless add to the complexity of IM by their mere presence, especially if they operate without clear goals.

A primary cause for confusion and uncertainty is the lack of a clear definition of information. Despite wide acknowledgement of its importance in all AECO products and processes, to the extent that perceptions of information in DM vary from a key means of communication and decision support to the main goal of design management, there is considerable fuzziness concerning what constitutes information in AECO. Many adopt a conventional view and equate information to drawings and other documents, even in the framework of BIM. As a result, IM is reduced to document management and to the use of document management systems, which often exist parallel to BIM, increasing redundancy and lowering overall efficiency.

Considering a document as information goes beyond using the carrier as a metaphor for the content, in the same way that we say "the Town Hall" to indicate the local authority accommodated in the building. It also reflects a strong adherence of AECO to conventional practices that have managed to survive into the digital era and may be uncritically replicated in digital information processing. For IM, this means that coordination of information production, exchange and utilization is in danger of being reduced to merely ensuring the presence of the right files, while most content-related matters, including quality assessment, are deferred to the human information users. It is therefore not surprising that both industry and academia complain that AECO has yet to define clear goals for information management and governance, even within BIM. Lots of data are captured but they are not always organized in ways that support comprehensive utilization.

IM literature is not particularly helpful in this respect. Arguably consistently with its broad scope, IM is rather inclusive concerning what is to be managed and covers documents, applications, services, schemes and metadata. To make such disparate material coherent and usable, IM literature proposes processing it in ways that establish correlations between data or with specific contexts, classify and categorize or condense data. This may be apply to conventional practices in AECO but is incompatible with new directions towards integration of information, as represented by BIM.

Finally, it should be stressed that IM is not a matter of brute force (by computers or humans) but of information organization. One can store all documents, files and models and hope for the best but stored information is not necessarily accessible and usable. As we know from search machines on the Internet, they can be very clever in retrieving what there is but this does not mean that they return the answers we need. If one asks for the specific causes of a fault in a building, it is not enough to receive all documents on the building from all archives to browse and interpret. Being able to identify the precise documents that refer to the particular part or aspect of the building depends on how the archives and the documents have been organized and maintained. To do that, one can rely on labour-intensive interpretation, indexing and cross-referencing of each part of each document – or one can try to understand the fundamental structure of these documents and build intelligent representations and management strategies based on them.

Key Takeaways

- *Computerization has added substantial possibilities to our information-processing capacities and also promoted the accumulation of huge amounts of information, which keep on increasing*
- *Computerization in AECO is still in a transitional stage, bounded by conventions from the analogue era and confused by its dual origins: automation of design and digitization of drawing*
- *Information is often organized in representations; understanding how representations are structured and operate is a prerequisite to both computerization of information and its management*
- *Information management is becoming critical for the utilization of digital information; instead of relying on brute-force solutions, one should consider the fundamental principles on which it should be based*

Exercises

1. Calculate how much data you produce per week, categorized in:
    1. Personal emails
    2. Social media (including instant messaging)
    3. Digital photographs, video and audio for personal use
    4. Study-related emails
    5. Study-related photographs, video and audio
    6. Study-related alphanumeric documents (texts, spreadsheets etc.)
    7. Study-related drawings and diagrams (CAD, BIM, renderings etc.)
    8. Other (please specify)
2. Specify how much of the above data is stored or shared on the Internet and how much remains only on personal storage devices (hard drives, SSD, memory cards etc.)
3. Calculate how much data a design project may produce and explain your calcu-

lations analytically, keeping in mind that there may be several design alternatives and versions. Use the following categories:

    1. CAD or BIM files
    2. PDFs and images produced from CAD & BIM or other software
    3. Alphanumeric files (texts, spreadsheets, databases etc.)
    4. Other (please specify)

4. Calculate how much of the above data is produced by different stakeholders, explaining your calculations analytically:

    1. Architects
    2. Structural engineers
    3. MEP engineers
    4. Clients
    5. Managers

Notes

---

1. Calculations and projections of information accumulated by human societies can be found in: Lyman, P. & Varian, H.P. 2003, "How much information." http://groups.ischool.berkeley.edu/archive/how-much-info/; Gantz, J. & Reinsel, D., 2011, "Extracting value from chaos." 2011, https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf;Turner, V., Reinsel D., Gantz J. F., & Minton S., 2014. "The Digital Universe of Opportunities" https://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm
2. Simonite, T., 2016. "Moore's law Is dead. Now what?" Technology Review https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/
3. The notion of information overload was popularized in: Toffler, A., 1970. *Future shock*. New York: Random House.
4. The dominance of email in AECO communication is reported in several sources, including a 2015 survey: https://www.newforma.com/news-resources/press-releases/70-aec-firms-say-information-explosion-impacted-collaboration/
5. Performance and in particular the avoidance of failures and related costs are among the primary reasons for adopting BIM, as argued in:Eastman, C., Teicholz, P.M., Sacks, R., & Lee, G., 2018. *BIM handbook* (3rd ed.). Hoboken NJ: Wiley.
6. Research conducted in the UK in 2015: https://www.newforma.com/news-resources/press-releases/70-aec-firms-say-information-explosion-impacted-collaboration/

PART II

BUILDING REPRESENTATION

2. Representation

This chapter introduces representations, in particular symbolic ones: how they are structured and how they describe things, including spatial ones. It explains that spatial symbolic representations are frequently graphs and presents some of the advantages of using such mathematical foundations. The chapter concludes with the paradigmatic and syntagmatic dimensions of representations, and their relevance for interpretation and management.

Symbolic representations

Many of the misunderstandings concerning information stem from our lack of understanding of representations and how these convey information. Representations are so central to our thinking that even if the sender of some information has failed to structure it in a representation, the receiver does so automatically. A representation can be succinctly defined as a system for describing a particular class of entities. The result of applying a representation to an entity is therefore a description. Representations of the symbolic kind, which proliferate human societies, consist of two main components:

- A usually finite set of symbols
- Some rules for linking these symbols to the entities they describe

The decimal numeral system is such a symbolic representation. Its symbols are the familiar Hindu-Arabic numerals:

$$S_D = \{0,1,2,3,4,5,6,7,8,9\}$$

The rules by which these symbols are linked to the quantities they describe can be summarized as follows:

$$n_n \cdot 10^n + n_{n-1} \cdot 10^{n-1} + \ldots + n_1 \cdot 10^1 + n_0 \cdot 10^0$$

These rules underlie positional notation, i.e. the description of a quantity as:

$$n_n n_{n-1} \ldots. n_1 n_0$$

For example, the description of seventeen becomes:

$$1 \cdot 10^1 + 7 \cdot 10^0 \Rightarrow 17$$

The binary numeral system is essentially similar. Its symbol set consists of only two numerals and its rules employ two as a base instead of ten:

$$S_B = \{0,1\}$$

$$n_n \cdot 2^n + n_{n-1} \cdot 2^{n-1} .+ \ldots + n_1 \cdot 2^1 + n_0 \cdot 2^0$$

This means that seventeen becomes:

$$1 \cdot 2^4 + 0 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0 \Rightarrow 10001$$

There are often alternative representations for the same class of entities. Quantities, for example, can be represented by (from left to right) Roman, decimal and binary numerals, as well as one of many tally mark systems:

XVII = 17 = 10001 = |||| |||| |||| ||

A representation makes explicit only certain aspects of the described entities. The above numerical representations concern quantity: they tell us, for example, that there are seventeen persons in a room. The length, weight, age or other features of these persons are not described. For these, one needs different representations.

Each representation has its advantages. Decimal numerals, for example, are considered appropriate for humans because we have ten fingers that can be used as an aid to calculations. Being built out of components with two states (on and off), computers are better suited to binary numerals. However, when it comes to counting ongoing results like people boarding a ship, tally marks are better suited to the task. Some representations may be not particularly good at anything: it has been suggested that despite their brilliance at geometry, ancient Greeks and Romans failed to develop other branches of mathematics to a similar level because they lacked helpful numeral representations.

Symbols and things

The correspondence between symbols in a representation and the entities they denote may be less than perfect. This applies even to the Latin alphabet, one of the most successful symbolic representations and a cornerstone of computerization. Using the compact set of symbols in an alphabet instead of syllabaries or logographies (i.e. graphemes that correspond to syllables or words) is an economical way of describing sounds (phonemes) in a language. This turns a computerized text into a string of ASCII characters that combine to form all possible words and sentences. Imagine how different text processing in the computer would be if its symbols were not alphabetic characters but pixels or lines like the strokes we make to form the characters in handwriting.

At the same time, the correspondence between Latin alphabet graphemes and the phonemes in the languages that employ them is not straightforward. In English, for example, the letter *A* may denote different phonemes:

- ɑː (as in ‘car’)
- æ (as in ‘cat’)
- ɒ (as in ‘call’)
- ə (as in ‘alive’)
- ɔː (as in ‘talk’)

The digraph *TH* can be either:

- θ (as in ‘think’) or
- ð (as in ‘this’)

Conversely, the phoneme eɪ can be written either as:

- *AY* (as in ‘say’)
- *EI* (as in ‘eight’)

The lesson we learn from these examples is that abstraction and context are important in representation. Abstraction allows for less strict yet still clear relations between symbols and things, as with the letter *A* which represents only vowels. A one-to-many correspondence like that is trickier than a simple one-to-one relation but is usually clarified thanks to the context, in our case proximal alphabetic symbols: ‘car’ and ‘cat’ are very similar strings but most English learners soon learn that they are pronounced differently and associate the right phoneme with the word rather than the letter. Similarly, in the floor plan of a building one soon learns to distinguish between two closely spaced lines denoting a wall and two very similar lines representing a step (Figure 1).

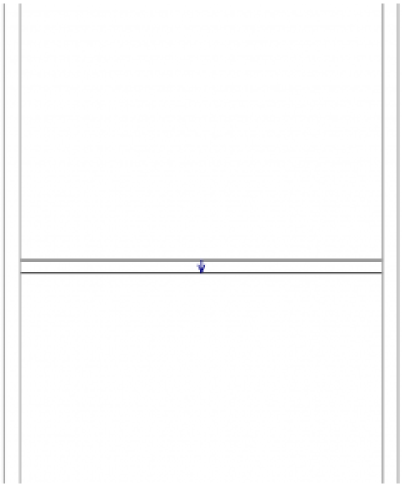


Figure 1. Walls and step in a floor plan: both types of entities are represented by two closely spaced parallel lines

Spatial symbolic representations

Symbolic representations are also used for spatial entities. A familiar example are metro and similar public transport maps. A common characteristic of many such maps is that they started life as lines drawn on a city map to indicate the route of each metro line and the position of the stations (Figure 2). As the size and complexity of the transport networks increased, the metro lines and stations were liberated from the city maps and became separate, diagrammatic maps: spatial symbolic representations, comprising symbols for stations and connections between stations (Figure 3). The symbols are similar for each line but may be differentiated e.g. by means of shape or colour, so that one can distinguish between lines. The symbol set for a metro network comprising two lines (the red O line and the blue Plus line) would therefore consist of the station symbol for the red line, the station symbol for the blue line, the connection symbol for the red line and the connection symbol for the blue line:

$$S_M = \{o, +, |_o, |_+\}$$

The rules that connect these symbols to real-world entities can be summarized as follows:

- Each station on a metro line (regardless of the complexity of the building that accommodates it) is represented by a station symbol of that line
- Each part of the rail network that connects two stations of the same line is represented by a line symbol of that line

These common-sense, practical principles underlie many intuitive attempts at spatial representation and, as discussed later on, even a branch of mathematics that provides quite useful and powerful means for formalizing and analysing symbolic spatial representations.

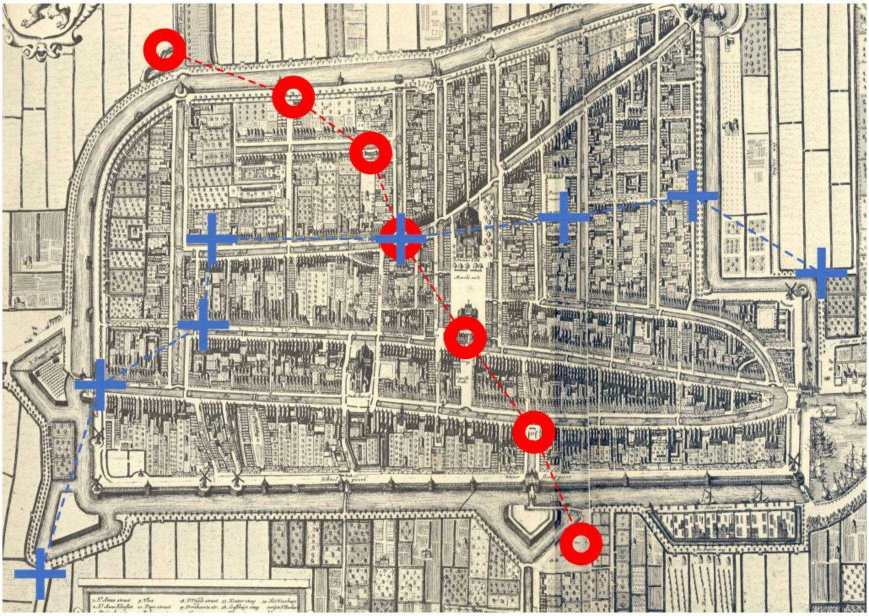


Figure 2. Metro lines and stations on a city map

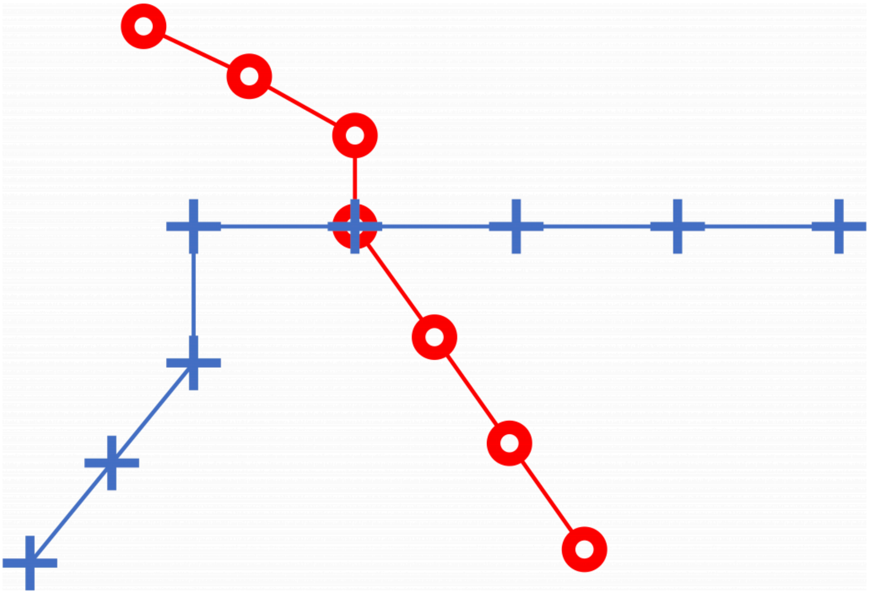


Figure 3. Metro map

Our familiarity with metro maps is to a large degree due to their legibility and usability, which make them excellent illustrations of the strengths of a good representation. As descriptions of a city transport system, they allow for easy and clear planning of travels, facilitate recognition of interchanges and connections, and generally provide a clear overview and support easy understanding. To manage all that, metro maps tend to be abstract and diagrammatic (as in Figure 2), in particular by simplifying the geometry of the metro lines (usually turning them into straight lines) and normalizing distances between stations (often on the basis of a grid). As a consequence, metro diagrams are inappropriate for measuring geometric distances between stations. Still, as travelling times on a metro often depend mostly on the number of stations to be traversed, metro maps are quite useful for estimating the time a trip may take. However, for finding the precise location of a station, city maps are far more useful.

A comparison of metro maps to numerals leads to the suggestion that the increase in dimensionality necessitates explicit representation of relations between symbols. In the one-dimensional numerals, relations are implicit yet unambiguous: positional notation establishes a strict order that makes evident which numeral stands for hundreds in a decimal number and how it relates to the numerals denoting thousands and tens. Similarly, in another kind of one-dimensional representation, spaces and punctuation marks are used in alphabetic texts to indicate the clustering of letters into words, sentences and paragraphs, and thus facilitate understanding of not only phonemes but also meanings in the text.

In two-dimensional representations like the metro diagrams, proximity between two station symbols does not suffice for inferring the precise relation between them. One needs an explicit indication like a line that connects the two symbols. A metro map missing such a connection (Figure 4) is puzzling and ambiguous: does the missing connection mean that a metro line is still under development or simply that the drawings is incomplete by mistake? Interestingly, such an omission in a metro diagram is quite striking and does not normally go unnoticed, triggering questions and interpretations, which will be discussed in the chapter on information theory (in relation to anti-data).

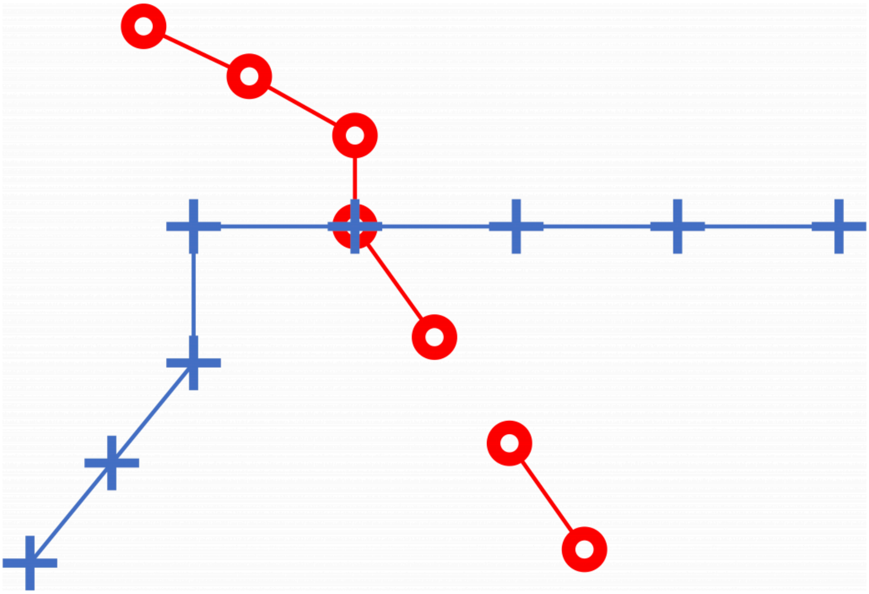


Figure 4. Metro map missing a connection between stations

Similarly puzzling is a metro map where stations of different lines are close to each other, even touching (Figure 5): does this indicate that the stations are housed in the same building, so that one can change from one line to the other, or that the stations are close by but separate, in which case one has to exit the metro and enter it again (which may involve having to buy a new ticket)? In a metro map where stations are clearly connected or coincide (Figure 3), there is no such ambiguity concerning interchange possibilities.

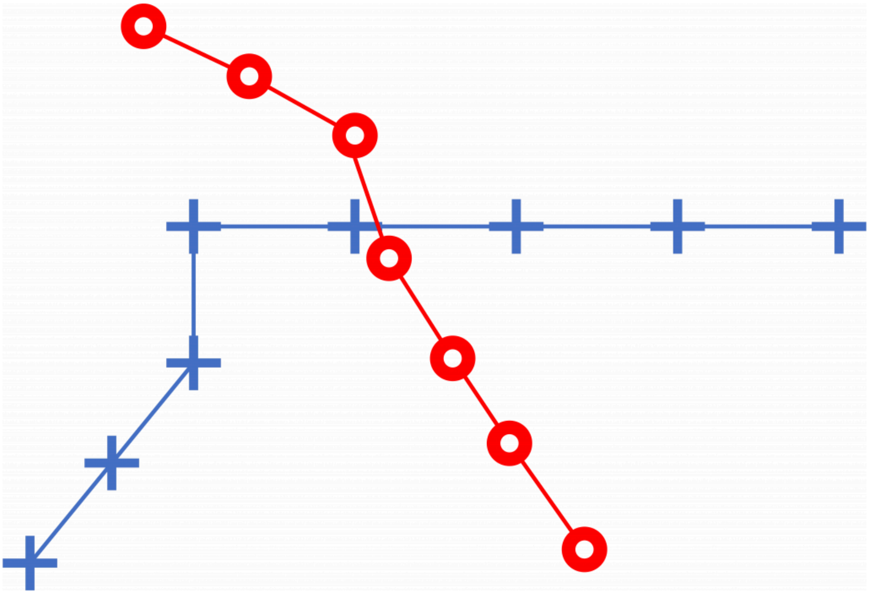


Figure 5. Metro map unclear about interchange possibilities

Graphs

Diagrams like these metro maps are *graphs*: mathematical structures that describe pairwise relations between things. Graph theory in mathematics began in 1736 with Euler's study of paths that crossed the bridges of Königsberg only once and has since gone from strength to strength. A key element of their success is that graphs are fairly simple but strictly structured diagrams consisting of *vertices* (or nodes) and *edges* (or lines) that link pairs of vertices. Vertices usually denote things and edges relations. In Figure 3, each metro station is a vertex and each connection between two stations an edge.

Graphs have a wide range of applications, from computer networks and molecular structures to the organization of a company or a family tree. The tools supplied by graph theory help analyse and quantify many of aspects of such networks. For example, the *degree* of a vertex (the number of edges connected to it) is a good indication of complexity: in a metro map it indicates the number of lines that connect there. The degree can therefore be used to identify interchanges, as well as a basic measure of how busy each interchange might be. Another measure is the *closeness* of a vertex: its mean distance to all other vertices in the graph (distance being the number of edges in the shortest path between two vertices). Closeness is a good indication of a vertice's centrality in a graph.

The *degree sequence* of a graph is a sequence that is obtained by listing the degrees of its vertices. In a map of a metro line this sequence is a good expression not only of opportunities for crossing over to other lines but also an indication of how busy the line may become as passengers make use of such opportunities. One can measure complexity in the whole graph in other ways, too, e.g. through *eccentricity:* the greatest distance between a vertex and any other vertex in the graph. The eccentricity of a metro station relates to its remoteness or poor connectivity. The *diameter* of the graph is the greatest eccentricity of any vertex in it and its *radius* the smallest eccentricity of any vertex. Vertices with an eccentricity equal to the radius are the center of the graph, while those with an eccentricity equal to the diameter are the periphery, In a metro system, therefore, it is interesting to know how many stations form te center and should consequently be easily and quickly accessible, and how many are in the periphery.

Finally, in order to be able to travel on the metro, the graph has to be *connected:* each vertex should connect to every other vertex by some sequence of edges and vertices (the graph in Figure 5 is therefore not connected). In fact, this sequence should be a *path*: no vertex should occur twice. Any edge that divides a graph into two parts (as in Figure 4) is called a *bridge*. In our metro example, all edges are bridges, making the metro particularly sensitive: any problem between two stations can render it unusable, as passengers cannot move along alternative routes.

What the above examples illustrate is that a well-structured representation can rely on mathematical tools that help formalize its structure and analyses. This is important for two reasons: firstly, formalization makes explicit what one may recognize intuitively in a representation; secondly, it allows for automation, especially of analyses. Allowing computers to perform painstaking and exhaustive analyses complements, liberates and supports the creative capacities of humans.

Graphs and buildings

Graph-like representations are also used for buildings: architects, for example, use bubble and relationship diagrams to express schematically the spatial structure of a design (Figure 3). In such diagrams nodes usually denote spaces where some specific activities take place (e.g. “Expositions” or “Library”), while edges or overlaps indicate proximity or direct access.

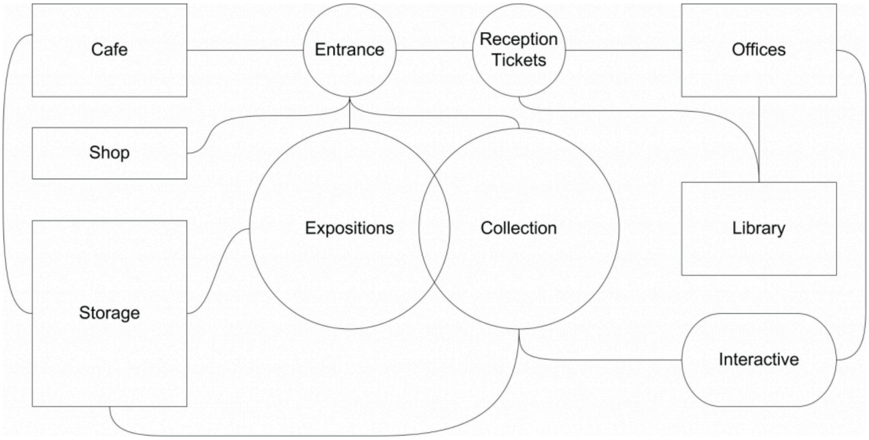


Figure 6. Relationship diagram

On the basis of graph theory, more formal versions of such diagrams have been developed, such as *access graphs*.Here nodes represent spaces and edges openings like doors, which afford direct connection between spaces. Access graphs are particularly useful for analysing circulation in a building.[1]

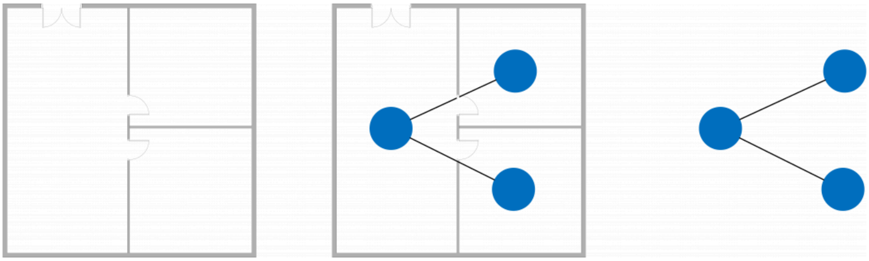


Figure 7. Floor plan and its access graph

The access graph demonstrates the significance of explicit structure: pictorially it may have few advantages over relationship diagrams, as both make explicit the entities in a representation and their relations. However, imposing the stricter principles of a mathematical structure reduces vagueness and provides access to useful mathematical tools. In a relationship diagram one may use both edges and overlaps to indicate relations, and shapes, colours and sizes to indicate properties of the nodes. In a graph, one must use only nodes and edges, and label them with the necessary attributes. This improves consistency and clarity in representation, similarly to the standardization of spelling in a language. It also facilitates application of mathematical measures which give clear indications of design performance. For example, the eccentricity of the node representing the space from where one may exit a building is a useful measure of how long it may take for people to leave the building, which is critical for e.g. fire egress. Similarly, the significance of a space for pedestrian circulation is indicated by its degree in the access graph, while spaces that form bridges are opportune locations for circulation control. For all these reasons, graphs are a representational basis to which we will returning in several parts of this book.

Paradigmatic and syntagmatic dimensions

In a symbolic representation we can analyse descriptions along two dimensions: the paradigmatic and the syntagmatic.[2] The paradigmatic dimension concerns the symbols in the representation, e.g. letters in a text. The syntagmatic dimension refers to the sequence by which these symbols are entered in the description. The meaning of the description relies primarily on the paradigmatic dimension: the symbols and their arrangement in the description. Syntagmatic aspects may influence the form of these symbols and their arrangement but above all reveal much about the cognitive and social processes behind the representation and its application, as well as mechanical aspects. For instance, in a culture where left-to-right writing is dominant, one would expect people to write numerals from left to right, too. However, the Dutch language uses a ten-before-unit structure for number words between 21 and 99 (as opposed to the unit-and-ten structure in English), e.g. "vijfentwintig" (five-and-twenty). Consequently, when writing by hand, e.g. noting down a telephone number dictated by someone else, one often sees Dutch people first enter the ten numeral, leaving space for the unit, and then backtrack to that space to enter the unit numeral. With a computer keyboard such backtracking is not possible, so the writer normally pauses while listening to the ten numeral, waits for the unit numeral and then enters them in the reverse order. Matching the oral representation to the written one may involve such syntagmatic peculiarities, which are moreover constrained by the implementation means of the representation (writing by hand or typing).

In drawing by hand, one may use a variety of guidelines, including perspective, grid and frame lines, which prescribe directions, relations and boundaries. These lines are normally entered first in the drawing, either during the initial setup or when the need for guidance emerges. The graphic elements of the building representation are entered afterwards, often in direct reference to the guidelines: if a graphic element has to terminate on a guideline, one may draw it from the guideline or, if one starts from the opposite direction, slow down while approaching the guideline, so as to ensure clear termination. Similar constraining influences may also derive from already existing graphic elements in the drawing: consciously or unconsciously one might keep new graphic elements parallel, similarly sized or proportioned as previously entered ones, terminate them against existing lines etc. Such mechanical and proportional dependence on existing graphic elements has led to the development of a wide range of object-snap options and alignment facilities in computerized drawing.

Any analysis of the paradigmatic dimension in a description aims at identifying symbols, e.g. relating each stroke in a handwritten text to a letter. To do that, one has to account for every stroke with respect to not only all symbols available in the representation but also various alternatives and variations, such as different styles of handwriting. Analyses of the syntagmatic dimension have to take into account not only the paradigmatic dimension (especially symbols and implementation mechanisms) but also cognitive, social, mechanical aspects that may have played a role in the temporal process of making a description, such as the tendency to draw *from* an existing graphic element to endure clear termination. Similarly, in most BIM editors, one enters openings like doors or windows only after the walls that host them have been entered in the model, while rooms are defined only after the bounding walls have been completed.

As all that relates to the organization of a design project and the relations between members of a design team, the syntagmatic dimension is of particular relevance to the management of information processes. Thankfully, there are sufficient tools for registering changes in a digital representation, since adding a time stamp to the creation, modification and eventual deletion of a symbol in a computer program is easy and computationally inexpensive. Making sense of what these changes mean requires thorough analysis of the sequences registered and clear distinctions between possible reasons for doing things in a particular order.

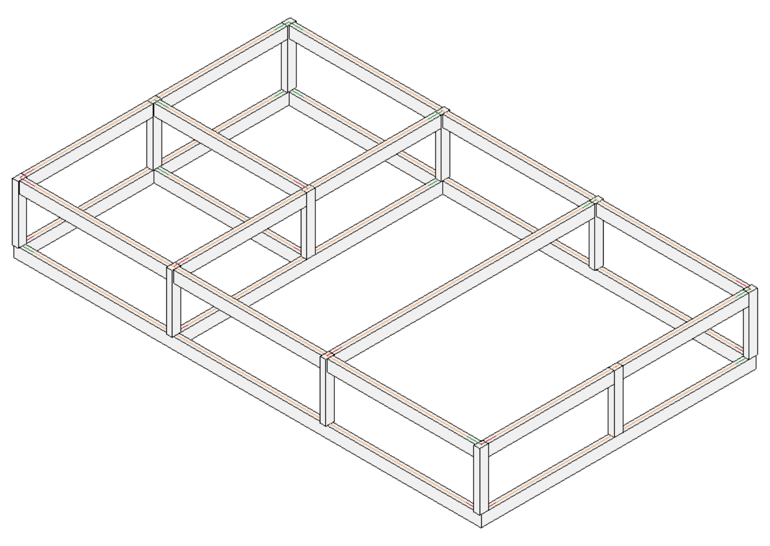
The significance of the syntagmatic dimension increases with the dimensionality of the representation: in a one-dimensional representation like a text, the sequence by which letters are entered is quite predictable, including peculiarities like the way Dutch words for numbers between 21 and 99 are structured. In representations with two or more dimensions, one may enter symbols in a variety of ways, starting from what is important or opportune and moving iteratively through the description until it is complete (although completeness may be difficult to ascertain syntagmatically, making it unclear when the process should terminate). This clearly indicates the significance of the syntagmatic dimension for the management of 3D and 4D representations of buildings.

Key Takeaways

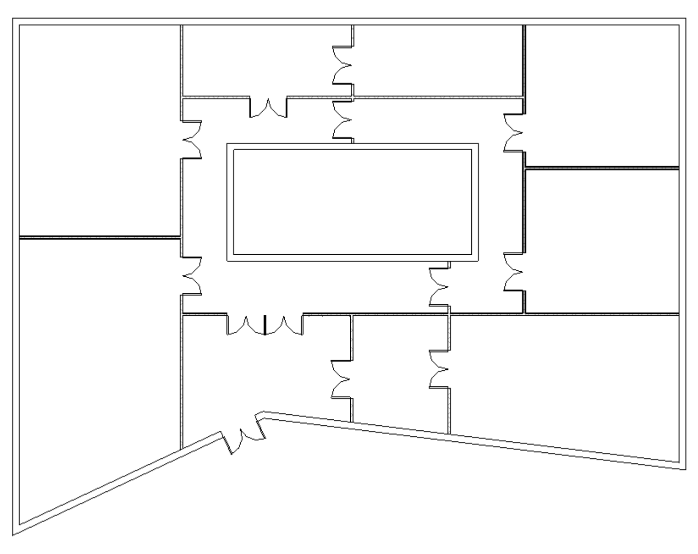
- *Symbolic representations employ usually finite sets of symbols and rules to relate these symbols to specific classes of entities and produce descriptions of these entities*
- *Familiar spatial symbolic representations like metro diagrams are graphs: mathe-*

matical structures that describe pairwise relations between things, using nodes for the things and edges for the relations
- *Graphs are a useful representational basis for buildings because they make symbols and relations between symbols explicit and manageable*
- *Symbolic descriptions have a paradigmatic and a syntagmatic dimension, relating respectively to the symbols they contain and the sequence by which the symbols have been entered in the description*
- *Interpretation of a description relies primarily on the paradigmatic dimension, while management strongly relates to the syntagmatic dimension*

Exercises



1. Draw graphs for the above post-and-beam structure:
    1. One using vertices for the posts and beams and edges for their connections
    2. One using vertices for the junctions and edges for the posts and beams
2. Calculate the following for the above graphs:
    1. The degree and eccentricity of each vertex
    2. The diameter and radius of each graph
3. Draw an access graph for the following floor plan:



4. In the access graph:
    1. Calculate the degree and eccentricity of each vertex
    2. Calculate the diameter and radius of the graph
    3. Indicate the vertices belonging to the center and the periphery
    4. Identify any bridges in the access graph

Notes

1. Graph-based applications in the representation of buildings are discussed extensively in: Steadman, P., 1983. *Architectural morphology: an introduction to the geometry of building plans.* London: Pion.
2. The discussion on the paradigmatic and syntagmatic dimensions in visual representations draws from: Van Sommers, P., 1984. *Drawing and cognition: descriptive and experimental studies of graphic production processes.* Cambridge: Cambridge University Press.

3. Analogue representations

> *To understand many of the problems surrounding building information, we first need to examine the analogue representations that still dominate AECO. The chapter presents some of the key characteristics that have made these representations so successful, although they do not necessarily agree with digital environments. Effective computerization relies on replacing the human abilities that enable analogue representations with capacities for information processing by machines.*

Pictorial representations and geometry

Familiar building representations tend to be drawings on paper, such as orthographic projections like floor plans and sections, and projective ones, including isometrics and axonometrics: two-dimensional depictions of three-dimensional scenes, through which one tries to describe the spatial arrangement, construction or appearance of a building. What these drawings have in common is:

- They are pictorial representations (not symbolic)
- They rely heavily on geometry

Even though drawings were used in building design already in antiquity, it was in the Renaissance that applied geometry revolutionized the way Europeans represented and conceptualized space, in many cases raising the importance of the graphic image over the written text. Geometry was not merely a handy foundation for descriptive purposes, i.e. formalizing pictorial representations of buildings, but also a means of ordering space, i.e. organizing people's experiences and thoughts to reveal some inherent order (including that of the cosmos). Consequently, building drawings evolved from schematic to precise and detailed representations that matched the perception of actual buildings, as well as most levels of decision making and communication about building design and construction.

Such empowerment gave geometry a central position in building design, with many architects and engineers becoming absorbed in geometric explorations closely linked to some presumed essence or ambition of their profession. With geometry forming both an overlay and underlay to reality, a complex relation developed between building design and geometry, involving not only the shape of the building but also the shape of its drawings. In turn, this caused building drawings to become semantically and syntactically dense pictorial representations, where any pictorial element, however small, can be significant for interpretation. By the same token, in comparison to more diagrammatic representations, the interpretation of building drawings involves a larger number of pictorial elements, properties and aspects, such as colour, thickness, intensity and contrast. As representations, building drawings were therefore considered a mixed and transitional case.[1]

The computerization of such complex, highly conventional analogue representations was initially superficial, aiming at faithful reproduction of their appearance. To many, the primary function of digital building representations, including not only CAD but also BIM, is the production of conventional analogue drawings either on paper (prints) or as identical computer files (e.g. a PDF of a floor plan). This makes computerization merely an efficiency improvement, especially concerning ease of drawing modification, compactness of storage and speed of dissemination. This is a testimony to the power and success of analogue building drawings but at the same time a major limitation to a fuller utilization of the information-processing capacities of computers. Analogue drawings work well in conjunction with human abilities for visual recognition, allowing us to develop efficient and effective means of specification and communication: most people recognize the same number of spaces in a floor plan on paper; scanning the floor plan transforms it into a computer file but computers generally only recognize it as an array of pixels. Recognizing the rooms and counting them by computer requires explicit representation of spaces.

Visual perception and recognition

Building drawings are surprisingly parsimonious: they manage to achieve quite a lot with a limited repertory of graphic primitives. With just a few kinds of lines, they produce floor plans, sections, perspectives etc., as well as depict a wide variety of shapes and materials in all these projections. To a large degree this is due to the ingenious ways they trigger the human visual system and allow us to see things. For example, we tend to associate similar elements if they are proximal. Therefore, closely parallel lines become depictions of walls but if the distance between the lines increases (beyond what might be plausible for a thick wall), they become just parallel lines. Seeing two lines as a wall does not necessarily mean they have to be strictly parallel or straight (Figure 1).

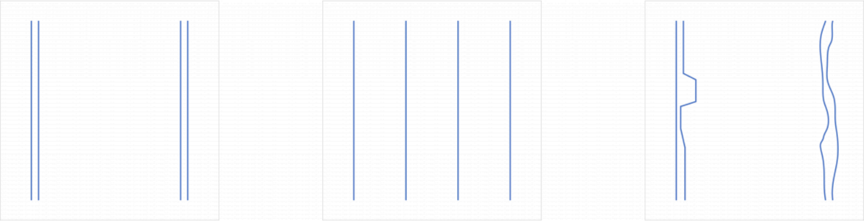


Figure 1. In the context of a building floor plan, closely spaced parallel lines are often paired into depictions of walls (left); if the distance between parallel lines increases, perceiving them as walls becomes hard or impossible (middle); perturbations or irregularity of shape do not necessarily disqualify closely spaced, roughly parallel lines as wall depictions (right)

It is similarly easy to identify columns in a floor plan. Even more significantly, the arrangement (repetition, collinearity, proximity etc.) and similarity of columns allow us to recognize colonnades: groups of objects with a specific character (Figure 2). The colonnade may be recognizable even if the columns are not identical and their arrangement not completely regular (Figure 3). However, if the arrangement is truly irregular, proximity or similarity do not suffice for the recognition of a colonnade (Figure 4).

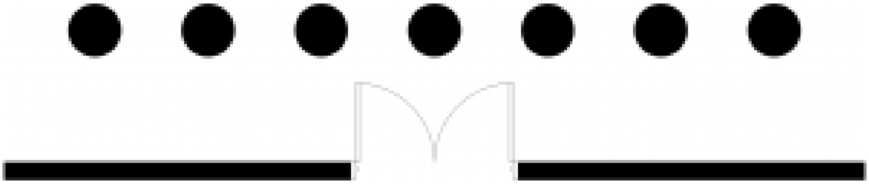


Figure 2. Colonnade in floor plan: recognition of the columns as a group is based on their arrangement and similarity

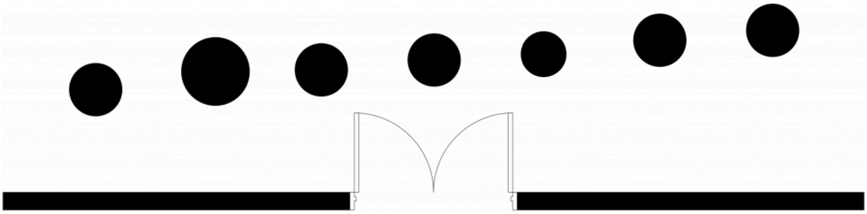


Figure 3. A colonnade may be recognized even if there are irregularities in the size and arrangement of the columns

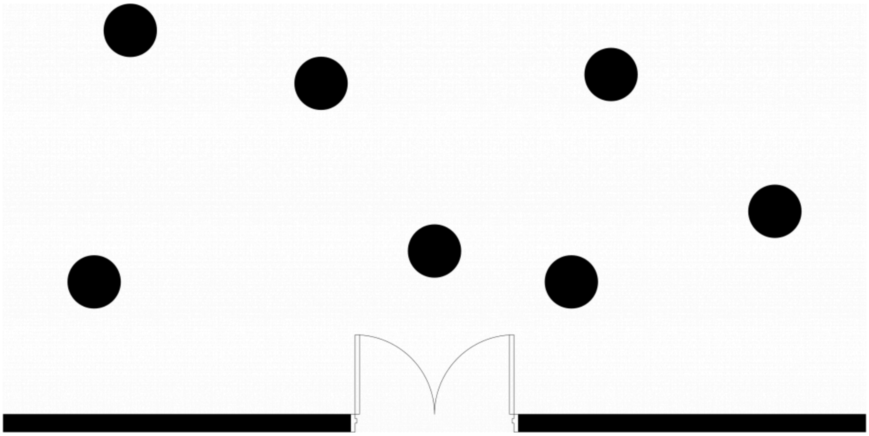


Figure 4. Randomly placed columns do not make a colonnade

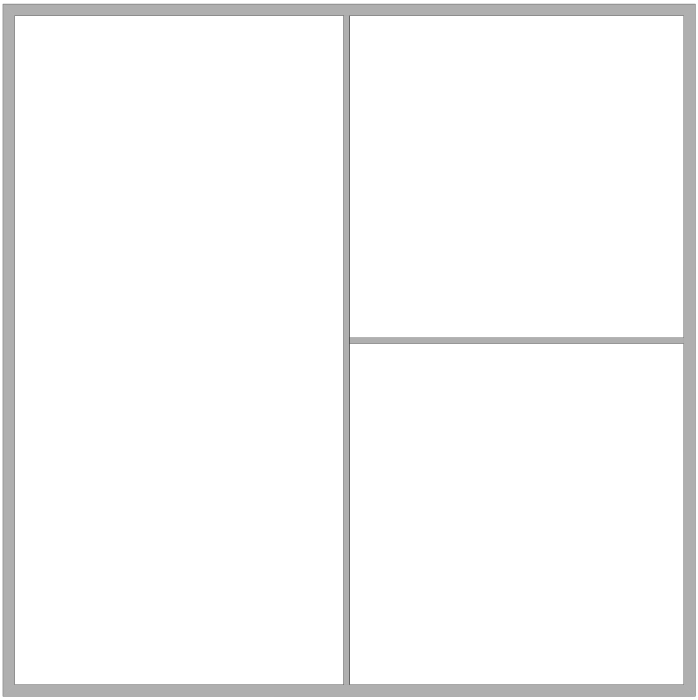


Figure 5. Floor plan of a building with three rooms: the drawing consists of just the walls but the rooms are instantly recognizable

Probably the most unnoticed and yet striking part of reading a drawing concerns the recognition of spaces: in a floor plan, one enters graphic elements that develop into depictions of building elements and components, like walls, doors and windows. Spaces are what is left over on paper, essentially background coming through the drawing. Yet most people with a basic understanding of building drawings are capable of recognizing the spaces in a floor plan (inferring them from the bounding building elements) with precision, accuracy and reliability (Figure 5).

Abstraction and incompleteness

Pictorial representations are characterized by a high potential for abstraction, which is evident in the different scales of building drawings: a wall at a scale like 1:20 is depicted by a large number of lines indicating various layers and materials; at 1:100 the wall may be reduced to just two parallel lines; at 1:500 it may even become a single, relatively thick line. Similarly, a door in a floor plan at 1:20 is quite detailed (Figure 6), at 1:100 it is abstracted into a depiction that primarily indicates the door type (Figure 7) and at 1:500 it becomes just a hole in a wall (Figure 8). At all three scales both the wall and the door are clearly recognizable, albeit at different scales of specificity and detail. Such abstraction is largely visual: it mimics the perception of a drawing (or, for that matter, any object) from various distances. It also corresponds to the design priorities in different stages: in early, conceptual design, one tends to focus on general issues, zooming out of the drawing to study larger parts, while deferring details to later stages. Therefore, the precise type, function and construction of a door may be relatively insignificant, making abstraction at the scale of 1:500 suitable. However, that abstraction level is inappropriate for the final technical design, when one has to specify not just the function and construction of a door but also its interfacing with the wall. To do so, one has to zoom in and use a scale like 1:20 to view and settle all details.

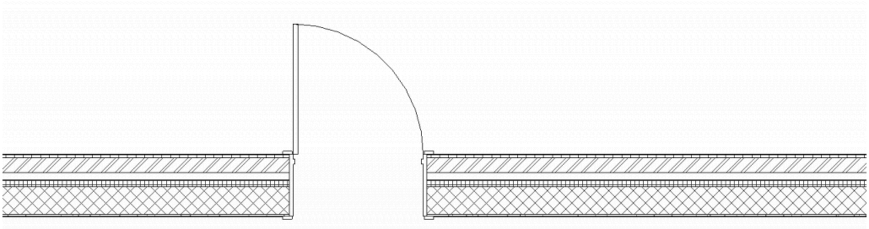


Figure 6. Wall and door at 1:20

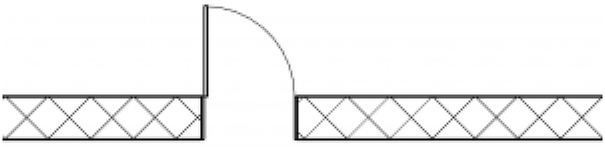


Figure 7. Wall and door at 1:100

Figure 8. Wall and door at 1:500

In addition to visual abstraction, one may also reduce common or pertinent configurations, however complex, into a single, named entity, e.g. an Ionic or Corinthian column or a colonnade (Figure 2) or "third floor" and "north wing". Such mnemonic or conceptual abstraction is constrained by visual recognition, as outlined above, but also relies on cultural convention: it is clearly not insignificant that we have a term for a colonnade. As such, mnemonic abstraction plays a more important role in symbolic representations than purely visual abstraction.

Pictorial representations are also relatively immune to incompleteness: a hastily drawn line on paper, with bits missing, is still perceived as a line (Figure 9). A house partially occluded by an obstacle is similarly perceived as a single, complete and coherent entity (Figure 10).

Figure 9. An imperfectly drawn line may still be perceived as a line

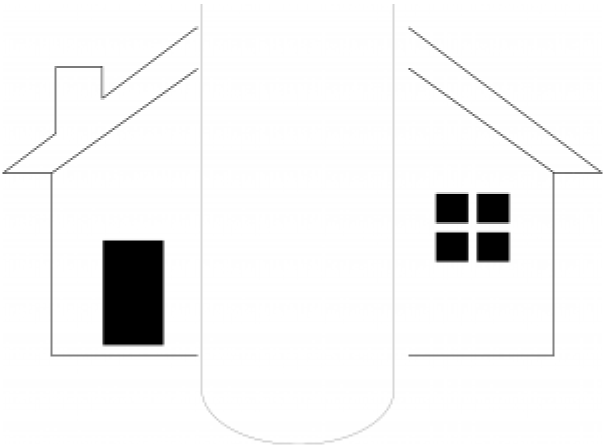


Figure 10. A house partially occluded by another object is still perceived as a single house

Dealing with incomplete descriptions is generally possible because not all parts are critical for understanding their meaning, even if they are not redundant. In English, for example, keeping only the consonants in a text may suffice for recognizing most words:

TH QCK BRWN FX JMPS VR TH LZY DG

(THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG)

This practice, currently known as disenvoweling, is widely applied in digital short messages. In the past, it was used to similar effect by telegraph operators, note takers and others who wanted to economize on message length and the time and effort required for writing or transmitting a message. Identifying the missing vowels is often a matter of context: 'DG' in a farmyard setting probably means 'DOG' but in an archaeological one it may stand for 'DIG'. If a word contains many vowels, it may be hard even then: 'JMPS' is highly probably 'JUMPS' in most contexts but 'DT' as a shorthand of 'IDIOT' may be far from effective in any context.

Likewise in images, some parts are more critical than others for recognition. A basic example is dashed lines: even with half of the line missing, the human visual system invariably recognizes the complete lines and the shapes they form (Figure 11).

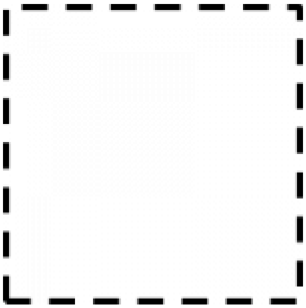


Figure 11. A square drawn with dashed lines

Interestingly, a shape drawn with dashed lines is recognized more easily if the line junctions are present. This relates to a general tendency of the human visual system to rely on points of maximum curvature in the outline of shapes.[2] Corners, in particular, are quite important in this respect: the presence of corners makes it possible to perceive illusory figures (Figure 12). The form of a corner gives perceivers quite specific expectations concerning the position and form of other corners connected to it, regardless of rectilinear or curvilinear geometry (Figure 13). The presence of compatible corners in the image leads to perception of an illusory figure occluding other forms. Perception of the illusory figure weakens if occlusion occurs at non-critical parts of the figure, such as the middle of its sides (Figure 14).

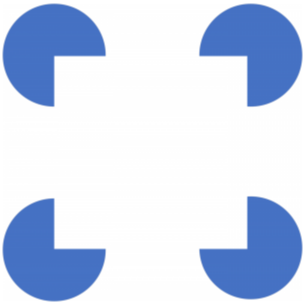


Figure 12. An illusory square

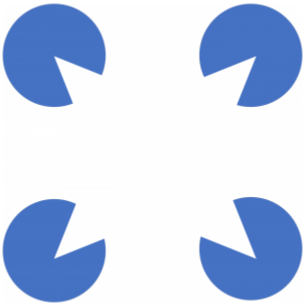


Figure 13. A curved illusory form

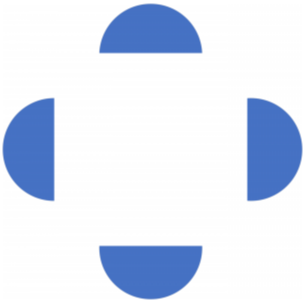


Figure 14. Missing the corners makes perception of illusory figures harder or more uncertain: in this case, one cannot be certain if the illusory square has rounded-off or bevelled corners

The importance of corners underlay one of the early successes in artificial intelligence: using a typology of edge junctions (Figure 15) and expectations about the connectivity of these types and the orientation of surfaces that met there, researchers were able to use constraint propagation to recognize the composition of scenes with trihedral geometric forms: faces, volumes and their relative positions (Figure 16).[3]

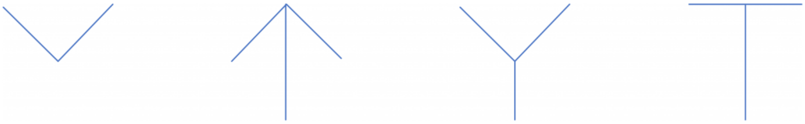


Figure 15. The four basic edge junction types in trihedral scenes

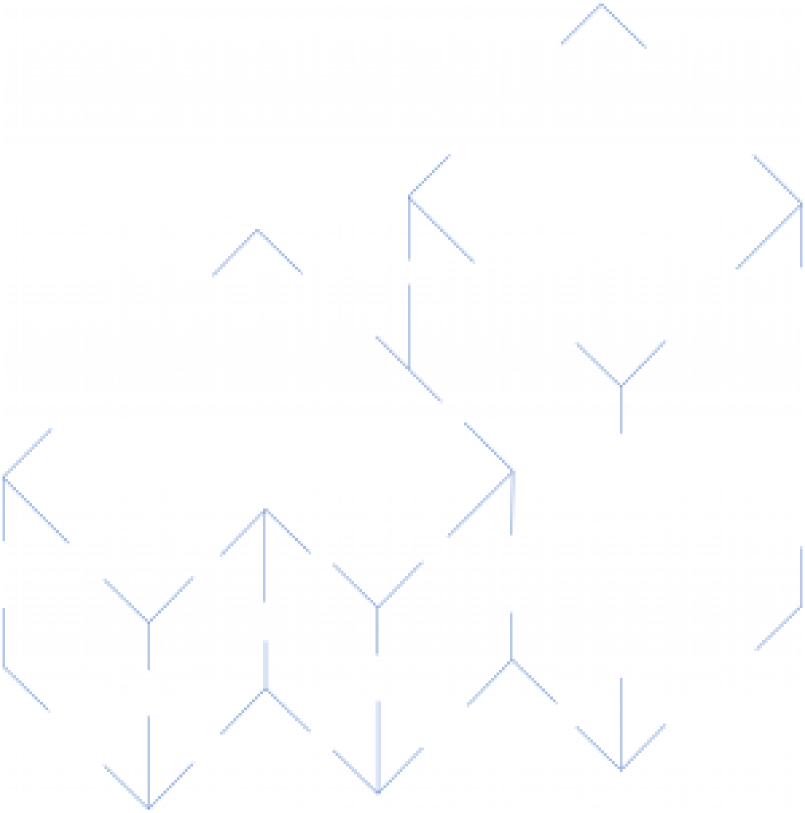


Figure 16. Recognition of objects in a trihedral scene can be based on the types of edge junctions in Figure 15

The above examples illustrate how analogue representations can be parsimonious and simultaneously effective but only if complemented with quite advanced and expensive recognition capacities. Empowering computers with such capacities is an emerging future but for the moment at least symbolic representations that contain explicit information are clearly preferable.

Implementation mechanisms

Another problem with analogue building representations is the overemphasis on geometry and the resulting dominance of implementation mechanisms over symbols. As symbols have to be implemented in various environments, one has to use means appropriate to each environment. A letter of the alphabet can be handwritten on paper with ink or graphite particles, depending on the writing implement (although one might claim that the strokes that comprise the letter are the real implementation mechanisms with respect to both the paradigmatic and the syntagmatic dimensions). In the computer, the same letter is implemented as an ASCII character in a text processing, spreadsheet and similar programs. In a drawing program, it may comprise pixels or vectors corresponding to the strokes (depending on the type of the program). In all cases, the symbol (the letter) is the same; what changes is the mechanisms used for its implementation.

With geometric primitives forming the graphic implementation mechanisms in pictorial building representations (underlay) and the ordering influence of geometry on building design (overlay), it has been easy to sidetrack attention to the geometric implementation mechanism of building representations, not only in the analogue but also in the digital versions. This geometric fixation meant lack of progress in CAD and also many misunderstandings in BIM.

To understand the true significance of geometric implementation mechanisms for the symbols in a building representation, consider the differences between alternative depictions of the same door in a floor plan (Figure 17). Despite differences between the graphic elements and their arrangement, they all carry the same information and are therefore equivalent and interchangeable. Many people reading the floor plan are unlikely to even notice such differences in notation, even in the same drawing, if the doors are not placed close to each other.

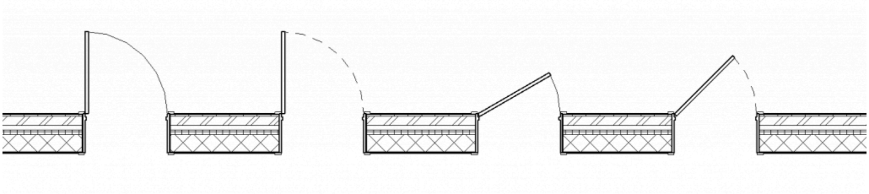


Figure 17. Alternative depictions of the same door

Using different door depictions for the same door type in the same drawing makes little sense. Differences in notation normally indicate different types of doors (Figure 18): they trigger comparisons that allow us to identify that there are different door types in the design and facilitate recognition of the precise differences between these types, so as to be able to judge the utility of each door in the design.

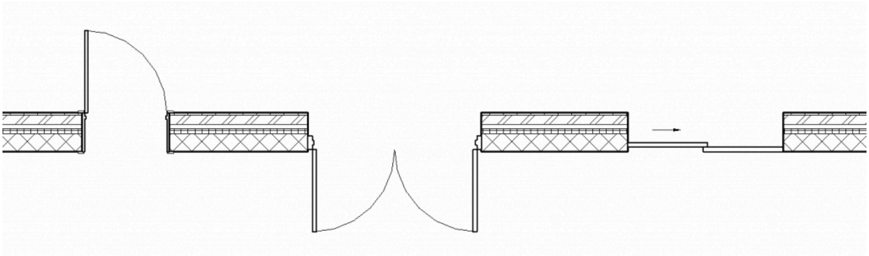


Figure 18. Alternative types of doors

In conclusion, one of the key advantages of symbolic representations is the pre-eminence of symbols and the attenuation of confusion between symbols and implementation mechanisms relative to pictorial representations. In computerized texts, letters are not formed by handwritten strokes that produce the required appearance; the appearance of letters is added to the letter symbols through properties like their font and size. Analogue building representations are similar to handwritten texts in that they may put too much emphasis on graphic elements because it is only through the interpretation of these that one can know e.g. the materials and layers that comprise a wall. In a symbolic representation, the materials and composition of the wall are explicit properties of an explicit symbol, which can also be described alphanumerically. This removes ambiguity and makes visual displays one of the possible views of building information.

Key Takeaways

- *Analogue building representations are mostly pictorial and rely heavily on geometry*
- *Visual perception and recognition are essential for the success of pictorial representations*
- *The reliance of analogue building representations on geometry leads to overemphasis on implementation mechanisms like graphic elements, even in digital environments*

Exercises

1. Identify the building elements and components in Figure 6 and list the properties described graphically and geometrically in the drawing
2. List and explain the differences between the above and what appears in Figure 7 and Figure 8

Notes

---

1. There are many treatises on building drawings, their history, significance and relation to geometry. The summary presented here draws in particular from: Cosgrove, D., 2003. Ptolemy and Vitruvius: spatial representation in the sixteenth-century texts and commentaries. A. Picon & A. Ponte (eds) *Architecture and the sciences: exchanging metaphors*. Princeton NJ: Princeton University Press; Evans, R., 1995. *The Projective Cast: Architecture and Its Three Geometries.* Cambridge MA: MIT Press; Goodman, N., 1976. *Languages of art; an approach to a theory of symbols* (2nd ed.). Indianapolis IN: Hackett·
2. The significance of points of maximum curvature, corners and other critical parts of an image is described among others in: Attneave, F., 1959. *Applications of information theory to psychology; a summary of basic concepts, methods, and results.* New York: Holt; Kanizsa, G., 1979. *Organization in vision: essays on Gestalt perception.* New York: Praeger.
3. The algorithmically and conceptually elegant recognition of scenes with trihedral objects

was finalized in: Waltz, D., 1975. Understanding line drawings of scenes with shadows. P.H. Winston (ed) *The psychology of computer vision*. New York: McGraw-Hill.

4. Building representations in BIM

This chapter offers an overview of symbolic building representations in BIM, including their key differences to analogue representations and how these were implemented in CAD. It explains how a model is built out of symbols that may have an uneasy correspondence with real-world objects and how abstraction can be achieved using these symbols.

Symbols and relations in BIM

BIM[1] is the first generation of truly symbolic digital building representations. CAD also used discrete symbols but these referred to implementation mechanisms: the geometric primitives that comprised a symbol in analogue representations. In BIM the symbols explicitly describe discrete building elements or spaces – not their drawings. BIM symbols usually appear as “libraries” of elements: predefined symbols of various types. The types can be specific, such as windows of a particular model by a certain manufacturer or abstract, e.g. single-hung sash windows or even just windows. The hierarchical relations between types enable specificity and abstraction in the representation, e.g. deferring the choice of a precise window type or of a window manufacturer to a later design stage, without missing information that is essential for the current stage: all relevant properties of the window, like its size, position and general type, are present in the generic window symbol at a suitable abstraction level.

Entering an instance of any kind in a model normally follows the following procedure:

- The user selects the symbol type from a library menu or palette
- The user positions and dimensions the instance in a geometric view like a floor plan, usually interactively by:
    - Clicking on an insertion point for the location of the instance, e.g. on the part of a wall where a window should be
    - Clicking on other points to indicate the window width and height relative

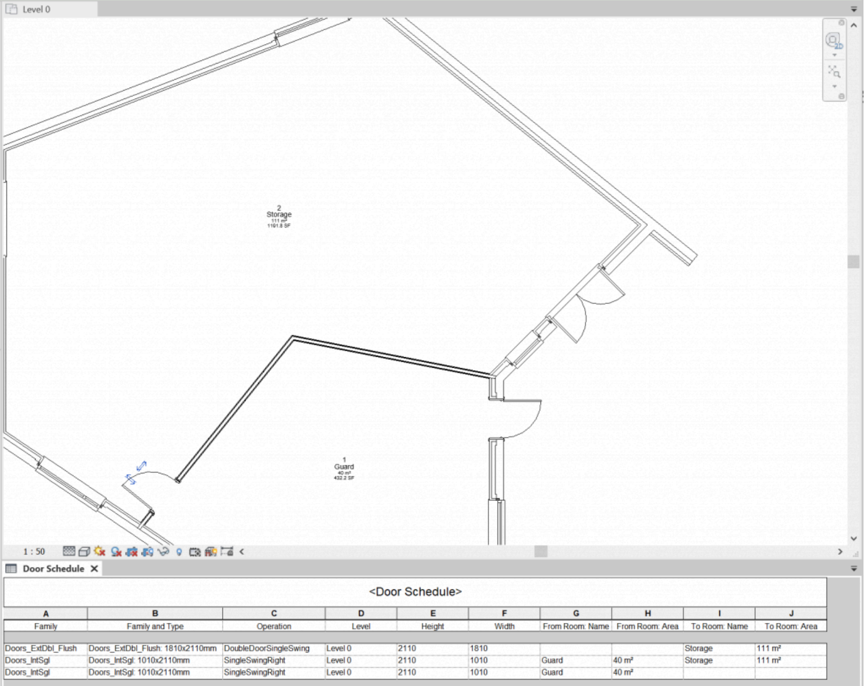
to the insertion point (this only if the window does not have a fixed size)

Modifications of the instance are performed in three complementary ways:

- Changes of essential properties such as the materials of a component amount to change of type. This is done by selecting a different symbol type from the library menu or palette and linking it to the instance.
- Changes in the geometry of an instance involve either repositioning the reference points or numerically changing the relevant values in any of the ways allowed by the program interface: in dialogue boxes that pop up by right-clicking on the instance, in properties palettes, through dimension lines or schedules.
- Changes in additional properties that do not conflict with the type, e.g. the occupancy of a space or the stage where a wall should be demolished, are entered through similar facilities in the interface, like a properties palette. Some of these properties are built in the symbols, while others can be defined by the user.

BIM symbols make all properties, geometric or alphanumeric, explicit: the materials of a building element are not inferred from its graphic appearance but are clearly stated among its properties, indicated either specifically or abstractly, e.g. “oak” or “wood”. Most properties in an instance are inherited from the type – not just materials but also any fixed dimensions: each wall type typically has a fixed cross section. Changing type properties like materials means crossing over to a different type, not changes in the instance properties. This ensures consistency in the representation by keeping all similar windows truly similar in all critical respects. This is essential for many tasks, such as cost estimation or procurement.

Many of the relations between symbols are also present in BIM, even if they are not always directly accessible. Openings like doors and windows, for example, are hosted by a wall. Therefore, normally they can only be entered after the hosting wall has been placed in the representation and in strict connection to it: trying to move a window out of a wall is not allowed. Connected walls may also have a specific relation, e.g. co-termination: if one is moved, the others follow suit, staying connected in the same manner. Similarly, spaces know their bounding elements (which also precede them in the representation) and if any of these is modified, they automatically adapt themselves. Through such relations, some of the possibilities offered by graphs become available in BIM, albeit often in indirect ways. A door schedule (Figure 1) reveals that, in addition to its hosting wall, a door knows which two spaces it connects (or separates when closed).



<Door Schedule>

A	B	C	D	E	F	G	H	I	J
Family	Family and Type	Operation	Level	Height	Width	From Room: Name	From Room: Area	To Room: Name	To Room: Area
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110mm	DoubleDoorSingleSwing	Level 0	2110	1810			Storage	111 m²
Doors_IntSgl	Doors_IntSgl: 1010x2110mm	SingleSwingRight	Level 0	2110	1010	Guard	40 m²	Storage	111 m²
Doors_IntSgl	Doors_IntSgl: 1010x2110mm	SingleSwingRight	Level 0	2110	1010	Guard	40 m²		

Figure 1. A door schedule in BIM reveals that each door is aware of the spaces it connects

The explicit symbolic representation of both the ‘solids’ out of which a building is constructed (building elements like walls, floors, doors and windows) and the ‘voids’ of the building (the spaces bounded by the building elements) is important. In analogue representations, the spaces are normally implicit, i.e. inferred by the reader. Having them explicit in BIM means that we can manipulate them directly and, quite significantly from the perspective of this book, attach to them information that cannot be linked to building elements: similarly to specifying that a window is made of oak wood, one can specify that a space is intended for a particular use, e.g. “office”, and even for specific activities like “small group meeting” or “CEO’s meeting room”. Such characterizations relate to various performance specifications, such as acoustics or daylighting, which can also be attached to the space and be used to guide and evaluate the design. Making spaces explicit in the representation therefore allows for full integration of building information in BIM and, through that, higher specificity and certainty. Spaces, after all, are the main reason and purpose of buildings, and most aspects are judged by how well spaces accommodate user activities.

BIM symbols and things

BIM has many advantages but, in common with other symbolic representations, also several ambiguities. Arguably the most important of these concerns the correspondence between symbols and real-world things. Building representations in BIM are truly symbolic, comprising discrete symbols. Unfortunately, the structure of building elements often introduces fuzziness in the definition of these symbols, similarly to the one-to-many correspondence between graphemes and phonemes we have seen in alphabets. In general, there are two categories of ‘solids’ in buildings. The first is building elements that are adequately represented by discrete symbols: doors and windows, for example, are normally complete assemblies that are accommodated in a hole in a wall. Walls, on the other hand, are typical representatives of the second category: conceptual entities that are difficult to handle in three respects. Firstly, walls tend to consist of multiple layers of brickwork, insulation, plaster, paint and other materials. Some of these layers continue into other elements: the inner brick layer of an external wall may become the main layer of internal walls, forming a large, complex and continuous structure that is locally incorporated in various walls (Figure 2).

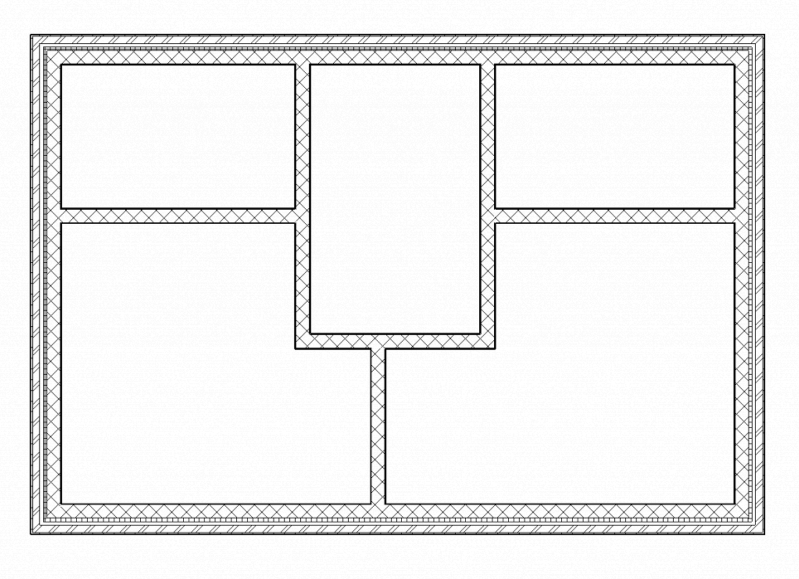


Figure 2. Continuous brick layer locally incorporated in two different kinds of wall

Secondly, BIM retains some of the geometric bias of earlier building representations, especially in the definition of elements like walls that have a fixed cross section but variable length or shape. When users have to enter the axis of a wall to describe this length or shape, they inevitably draw a geometric shape. BIM usually defines symbols on the basis of the most fundamental primitives in this shape. Even if one uses e.g. a rectangle to describe the axis, the result is four interconnected yet distinct walls, each corresponding to a side of the rectangle. Similarly, a wall with a complex shape, but conceptually and practically unmistakably a single structure, is analysed into several walls, each corresponding to a line segment of its shape (Figure 3).

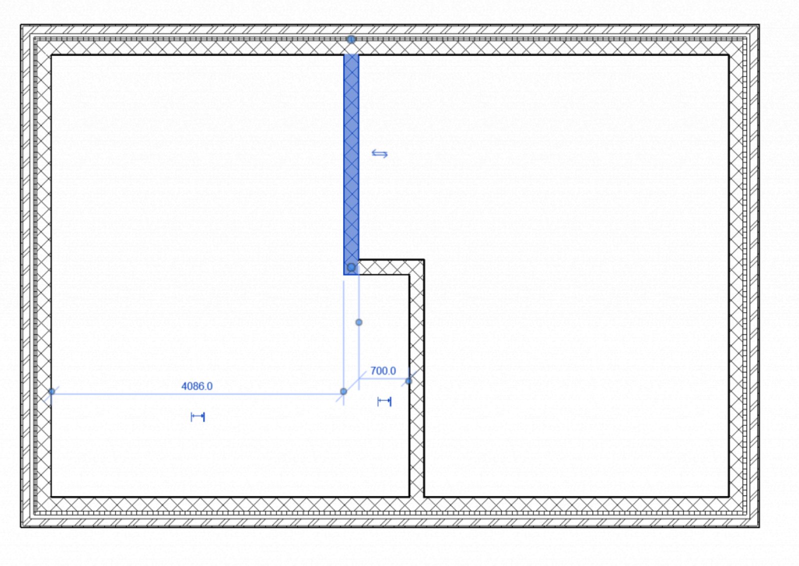


Figure 3. The internal wall is clearly one structure but in BIM each segment is represented as a distinct wall

Thirdly, our own perception of elements like walls may get in the way. Standing on one side of a wall, we see only the portion of the wall that bounds the room we are in. Standing on the other side, we perceive not only a different face but possibly also a different part of the wall (Figure 4). As a result, when thinking from the perspective of either space, we refer to parts of the same entity as if they were different walls.

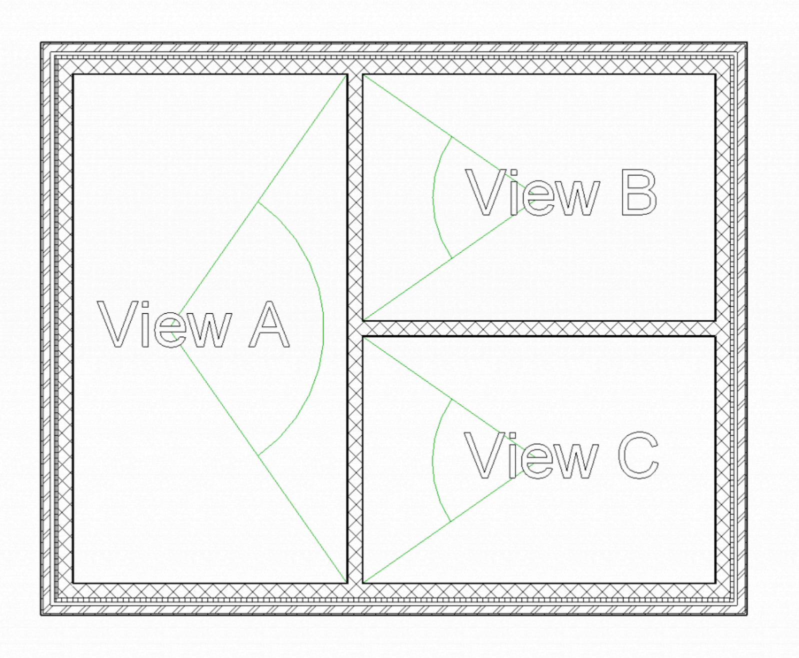


Figure 4. Three different views of the same wall

The inevitable conclusion is that some symbols in BIM may still require further processing when considered with respect to particular goals. One may have to analyse a symbol into parts that then have to be combined with parts of other symbols, e.g. for scheduling the construction of the brickwork in Figure 2. Other symbols have to be grouped together, like the internal wall in Figure 3. Such manipulations should not reduce the integrity of the symbols; it makes little sense to represent each layer of a wall separately. At the same time, one has to be both consistent and pragmatic in the geometric definition of building elements. In most cases, acceptance of the BIM preference for the simplest possible geometry is the least painful option: the vertical internal wall in Figure 4 should be represented as a single entity and not split into two parts in order to simplify the adjacency of walls to spaces. Looking at it in any way beyond the three perspectives indicated in the figure and the spaces that frame them, it cannot be anything else than a single building element.

Paradigmatic and syntagmatic dimensions in BIM

Even with the issues discussed above, the symbolic character of BIM has obvious advantages for the paradigmatic dimension: each symbol is explicit and integral in a building representation. The same holds for the syntagmatic dimension, in three different respects. The first concerns the practical side of developing a building representation in BIM: in common with most computerized programs, BIM editors can record the sequence of user actions and so make the history of a representation accessible and transparent. This allows users to undo actions and backtrack to earlier states.

More significantly, the sequence of user actions is often organized in prescriptive procedures because their order is not trivial. As we have seen, one has first to select the type of a new symbol in a model and then indicate the geometry of the instance. Such procedures ensure consistency in the symbols and their structure, as well as register many relations between symbols, for example the anchoring of a window or a wash basin to a hosting wall.

The third advantage of BIM for the syntagmatic dimension relates to 4D modelling: the addition of a time property to symbols, for example the moment the symbolized element should be constructed. This supports the scheduling of construction, demolition and other real-world activities from within the building representation, and reduces inconsistencies or other errors that emerge from poor communication between building representations and scheduling activities or software.

Abstraction and grouping in BIM

BIM symbols cover a wide range of abstraction levels, from generic symbols like “internal wall” without any further specifications to highly detailed symbols, representing e.g. a very specific wall type, including precise descriptions of materials from particular manufacturers. Usually a building representation in BIM starts with abstract symbols, which become progressively more specific. It is also possible to backtrack to a higher abstraction level rather than sidestep to a different type on the same level, e.g. when some conflict resolution leads to a dead end and one needs to reconsider their options. This typologic abstraction is one of the strong points of BIM but also something one has to treat with care because a model may contain symbols at various abstraction levels. Managing the connections between them, e.g. deciding on the interfacing between a highly specific window and an abstract wall, requires attention to detail. On the positive side, one can use such connections to guide decision making, e.g. restrict the choice of exact wall type to those that fit the expectations from the window.

Symbolic representations also have considerable capacities for bottom-up mnemonic abstraction on the basis of explicit relations between symbols, ranging from similarity (e.g. all vowels in a text) to proximity (all letters in a word). As it typical of digital symbolic representations, BIM allows for multiple groupings of symbols to produce mnemonic structures of all kinds, e.g. selecting all instances of the same door type in a design, identifying all spaces with a particular use on the second floor or determining which parts of a design belong to the north wing. For the latter, some additional input from the user may be required, such as drawing a shape that represents the outline of the north wing or labelling every symbol with an additional wing property. No user input is required for relations built into the behavioural constraints of a symbol, e.g. the hosting of openings in walls.

Through the combination of standard symbol features (like their properties) and arbitrary, user-defined criteria (like the outline of a wing), one can process the representation at any relevant abstraction level and from multiple perspectives, always in direct reference to specific symbols. For example, it is possible to consider a specific beam in the context of its local function and connections to other elements but simultaneously with respect to the whole load-bearing structure of its floor and wing or of the whole building. Any decision taken locally, specifically for this beam, relates transparently to either the instance or the type and may therefore lead not only to changes in the particular beam but also reconsideration of the beam types comprising the structure, e.g. a change of type for all similar beams. Reversely, any decision concerning the general type of the structure can be directly and automatically propagated to all of its members and their arrangement.

The automatic propagation of decisions relates to parametric modelling: the connection of symbol properties so that any modification to one symbol causes all others to adapt accordingly. In addition to what is built into the relations between types and instances or the behaviours like hosting, one can explicitly link instance properties, e.g. make several walls remain parallel to each other or vertical to another wall. One can also specify that the length of several walls is the same or a multiple of an explicitly defined parameter. Changing the value of the parameter leads to automatic modification of the length of all related walls. Parametric design holds significant promise. People have envisaged building representations in which it suffices to change a few values to produce a completely new design (or variation). However, establishing and maintaining the constraint propagation networks necessary for doing so in a reliable manner remains a major challenge. For the moment, parametric modelling is a clever way of grouping symbols with explicit reference to the relation underlying the grouping, e.g. parallelism of walls. Still, even in such simple cases, the effects of parametric relations in combination with built-in behaviours can lead to unpredictable and unwanted results.

In views which replicate conventional drawings, BIM software often also incorporates visual abstraction that mimics that of scales in analogue representations. By selecting e.g. “1:20” and “fine” one can make the visual display of a floor plan more detailed than with “1:200” and “coarse”. Such settings are useful only for visual inspections; they alter only the appearance of symbols, not their type or structure.

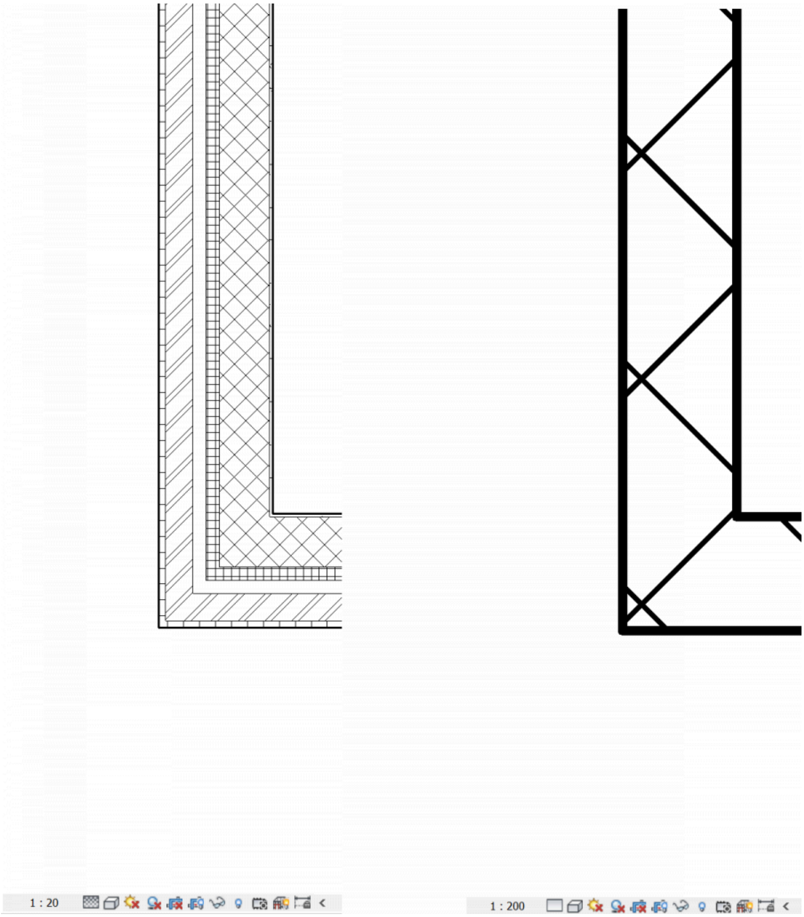


Figure 5. Display of the same wall in a BIM floor plan, under settings 1:20 and fine (left), and 1:200 and coarse (right)

The LoD in BIM is also related to abstraction. LoD specifications attempt to standardize the specificity of information in a model or preferably in a symbol, as a model may contain elements at various LoD. Many LoD standards have been proposed but strict adherence to them is a throwback to analogue standards regarding drawing scale. Such adherence fails to appreciate that information in a model has a reason and a purpose: some people have taken decisions concerning some part or aspect of a design. The specificity of these decisions and of the resulting representations is not accidental or conventional. Rather, it reflects what is needed for that part or aspect at the particular stage of a project. The LoD of the model that accommodates this information can only be variable, as not all parts or aspects receive the same attention at the same stages.

Specificity should therefore be driven by the need for information rather than by convention. If information in a representation is at a higher specificity level, one should not discard it but simply abstract in a meaningful way by focusing on relevant properties, relations or symbols. A useful analogy is with how human vision works: in your peripheral vision, you perceive vague forms and movement, e.g. something approaching you rapidly. If you turn your eyes and pay attention to these forms, you can see their details and recognize e.g. a friend rushing to meet you. As soon as you turn to these forms, other parts of what you perceive become vague and schematic. In other words, the world is as detailed as it is; your visual system is what makes some of its parts more abstract or specific, depending on your needs. By the same token, the specificity of a building representation should be as high as the available information allows. Our need for information determines the abstraction level at which we consider the representation, as well as actions by which we can increase the specificity of some of its parts.

Implementation mechanisms in BIM

Despite its symbolic structure, BIM uses the same implementation mechanisms as CAD: the same geometric primitives that reproduce the graphic appearance of analogue representations. The key difference is that these primitives are just part of pictorial views, in which they express certain symbol properties. The type of a door, for example, is explicitly named, so that we do not have to infer its swing from the arc used to represent it in a floor plan; the width of a wall is a numerical property of its symbol, so that we do not have to measure the distance between the two lines indicating the outer faces of the wall. On the contrary, this distance is determined by the width property of the symbol.

As we have seen, however, implementation mechanisms still influence the structure of a building representation in other respects: a wall is still partly determined by drawing its axis and so by the geometric shape one draws. On the whole, therefore, one should consider BIM as largely immune to undue influences from implementation mechanisms but at the same time remain aware of persistent geometric biases in both building representation in BIM and in the mindset of BIM users.

Key Takeaways

- *BIM is a truly symbolic building representation that employs discrete symbols to describe building elements and spaces. Symbols in BIM integrate all properties of the symbolized entities, which determine their pictorial appearance.*
- *This makes BIM symbols largely independent of graphic implementation mechanisms and immune to most geometric biases.*
- *The correspondence between BIM symbols and some building elements is problematic in certain respects due to the structure of these elements, persisting geometric biases and human perception and cognition.*
- *The symbolic structure of BIM representations has advantages for the paradigmatic dimension (it makes symbols explicit) and the syntagmatic dimension (through prescriptive procedures for user input, as well as parametric modelling).*
- *Abstraction in BIM is both typological (as symbols are at various abstraction levels) and mnemonic (based on similarity of properties and relations like proximity and hosting between symbols). Mnemonic abstraction amounts to grouping of symbols and relates to parametric modelling.*

Exercises

1. In a BIM editor of your choice (e.g. Revit), make an inventory of all wall types (*Families* in Revit) in the supplied library. Classify these types in terms of abstraction, clearly specifying your criteria.
2. In a BIM editor of your choice, make a simple design of a space with four walls and two floors around it. Identify properties of the building elements and space

symbols that connect them (e.g. dimensions) and overlapping properties (e.g. space properties that refer to finishings of the building elements). Make schedules that illustrate your findings.

3. Expand your design with another space and a door that connects them. Make a schedule that illustrates some relations between the spaces.
4. In the same design, describe step by step how a change in the size of one room is propagated to other symbols in the model.

Notes

---

1. A comprehensive general introduction to BIM, which may be necessary, depending on the reader's experience with it, is: Eastman, C., Teicholz, P.M., Sacks, R., & Lee, G., 2018. *BIM handbook* (3rd ed.). Hoboken NJ: Wiley.

PART III

INFORMATION: THEORY AND MANAGEMENT

5. Data and information

Previous chapters have explained how information is organized in representations. A question that remains to be answered is what exactly constitutes information, i.e. what one should consider as information and data in these representations. This chapter introduces relevant theories and explains how they apply to building information and representations.

Theories and definitions

There is nothing more practical than a good theory: it supplies the definitions people need to agree what to do, how and why; it explains the world, providing new perspectives from where to see and understand it; it establishes targets for researchers keen to improve or refute the theory and so advance science and knowledge. In our case, there is a clear need for good, transparent and operational definitions. Terms like ‘information’ and ‘data’ are used too loosely, interchangeably and variably to remove ambiguities in information processing and management. Computerization adds to this vagueness, especially with subjects like buildings: as we have seen in previous chapters, there may be a big gap between the analogue representations still used in most AECO processes and the capacities of computers.

A theory that resolves these problems cannot draw from the AECO domains only. It needs a firm foundation in general theories of information, especially those that take the capacities and peculiarities of digital means and environments into account. Thankfully, there are enough candidates for this.

Syntactic, semantic and pragmatic theories

When one thinks of information theory in a computing context, Shannon’s MTC springs to mind.[1] The MTC is indeed foundational and preeminent among formal theories of information. It addresses what has been visualized as the innermost circle in information theory (Figure 1):[2] the syntactic core of information, dealing with the structure and basic, essential aspects of information, including matters of probability, transmission flows and capacities of communication facilities – the subjects of the technical side of information theory.

The outermost circle in the same visualization is occupied by pragmatics: real-life usage of meaningful information. Information management theories (which will be discussed in a later chapter) populate this circle, providing a general operational framework for supporting and controlling information quality and flow. To apply this framework, one requires pragmatic constraints and priorities from application areas: a notary and a facility manager have different interests with regard to the same building information.

Between the syntactic and the pragmatic lies the intermediate circle of semantics, which deals with how meaning is added to the syntactical components of information before they are utilized in real life. As syntactic approaches are of limited help with the content of information and its interpretation, establishing a basis for IM requires that we turn to semantic theories of information.

Arguably the most appealing of these is by Luciano Floridi, who is credited with establishing the subject of philosophy of information. His value goes beyond his position as a modern authority on the subject. The central role of semantics in his work is an essential contribution to the development of much-needed theoretical principles in a world inundated with rapidly changing digital technologies. In our case, they promise a clear and coherent basis for understanding AECO information and establishing parsimonious structures that link different kinds of information and data. These structures simplify IM in a meaningful and relevant manner: they allow us to shift attention from *how* one should manage information (the technical and operational sides) to *which* information and *why.*

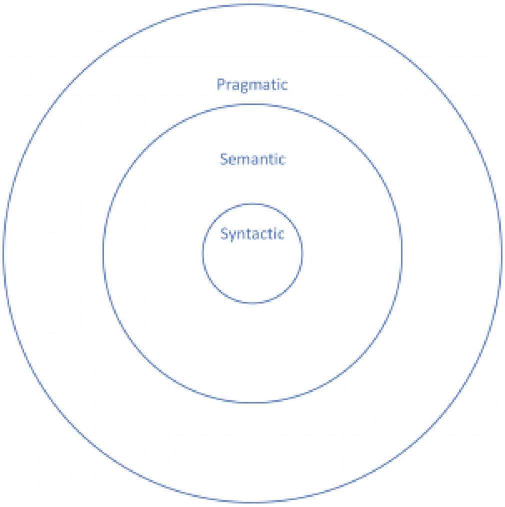


Figure 1. A classification of information theories

Before moving on to explaining this theory and applying it to building information, it should be noted that management, computing and related disciplines abound with rather too easy, relational definitions of data, information, knowledge, strategy etc., e.g. that data interpreted become information, information understood turns into knowledge and so forth. Such definitions tend to underestimate the complexity of various cognitive processes and are therefore not to be trusted. In this book, we focus on data, information and their relation. The rest concerns utilization of information and benefits that may be derived for individuals, enterprises, disciplines or societies – matters that require extensive analyses well beyond the scope of the present book. Information certainly contributes to achieving these benefits and in many cases it may even be a prerequisite but seldom suffices by itself. Rather than making unfounded claims about knowledge and performance, we focus on more modest goals concerning IM: understanding building information, its quality and flows, and organizing them in ways that may help AECO take informed decisions, in the hope that informed also means better.

A semantic theory for building information

Data and information instances

A fundamental definition in Floridi's theory[3] concerns the relation between data and information: an instance of information consists of one or more data which are well-formed and meaningful. Data are defined as lacks of uniformity in what we perceive at a given moment or between different states of a percept or between two symbols in a percept. For example, if a coffee stain appears on a floor plan drawing on paper (Figure 3), this is certainly a lack of uniformity with the earlier, pristine state of the drawing but it is neither well-formed nor meaningful within the context of architectural representations. It tells us nothing about the representation or the represented design, only that someone has been rather careless with the drawing (the physical carrier of the representation).

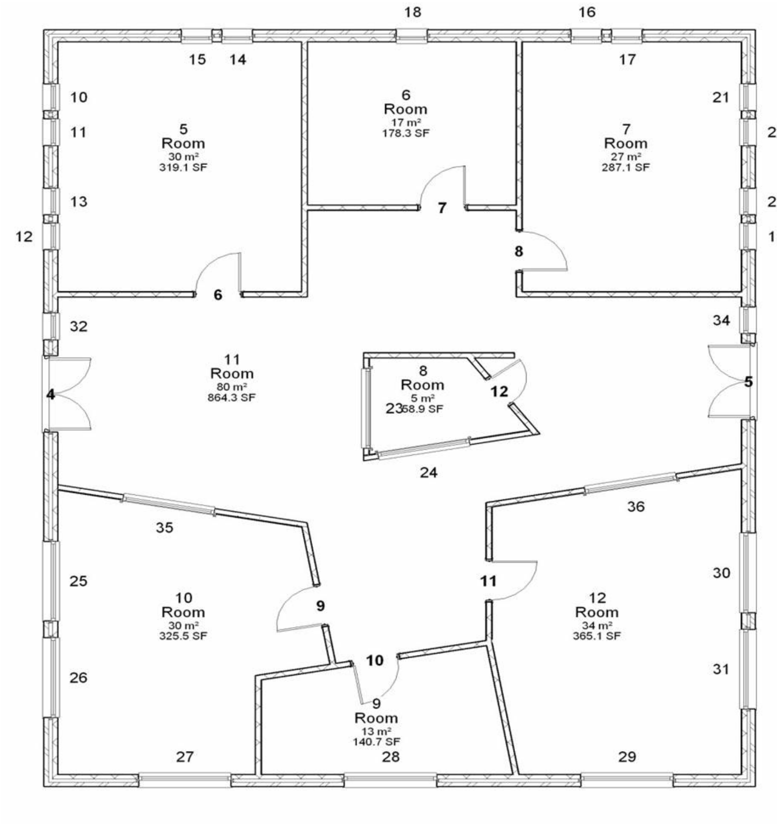


Figure 2. Floor plan

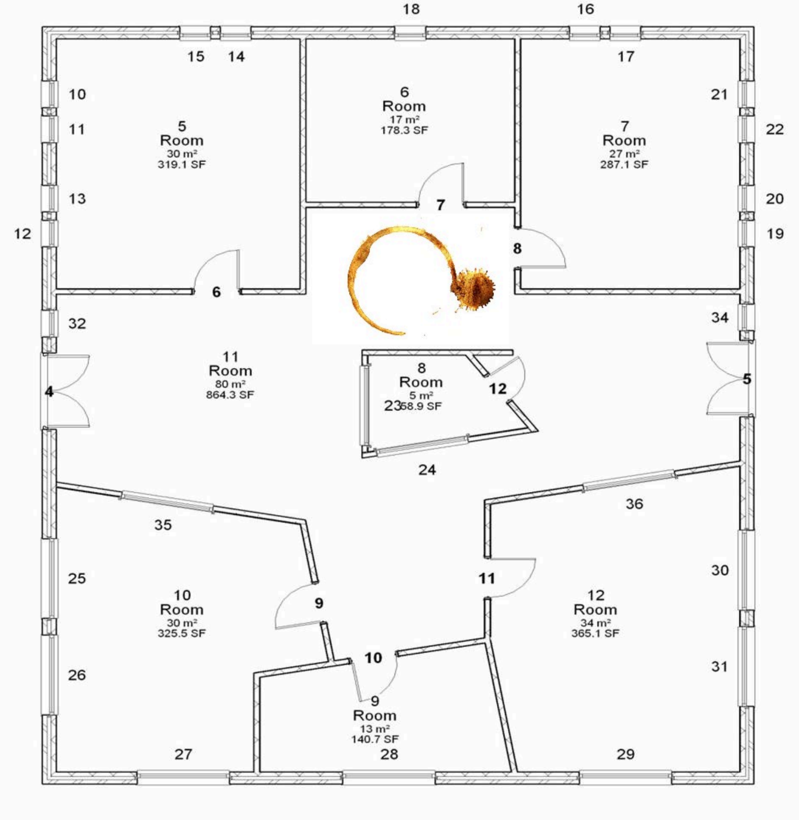


Figure 3. A new state of the floor plan: the coffee stain is neither well-formed nor meaningful in the framework of a line drawing

On the other hand, if the lack of uniformity between the two states is a new straight line segment across a room in a floor plan (Figure 4), this is both well-formed (as a line in a line drawing) and meaningful (indicating a change in the design, possibly that the room has now a split-level floor).

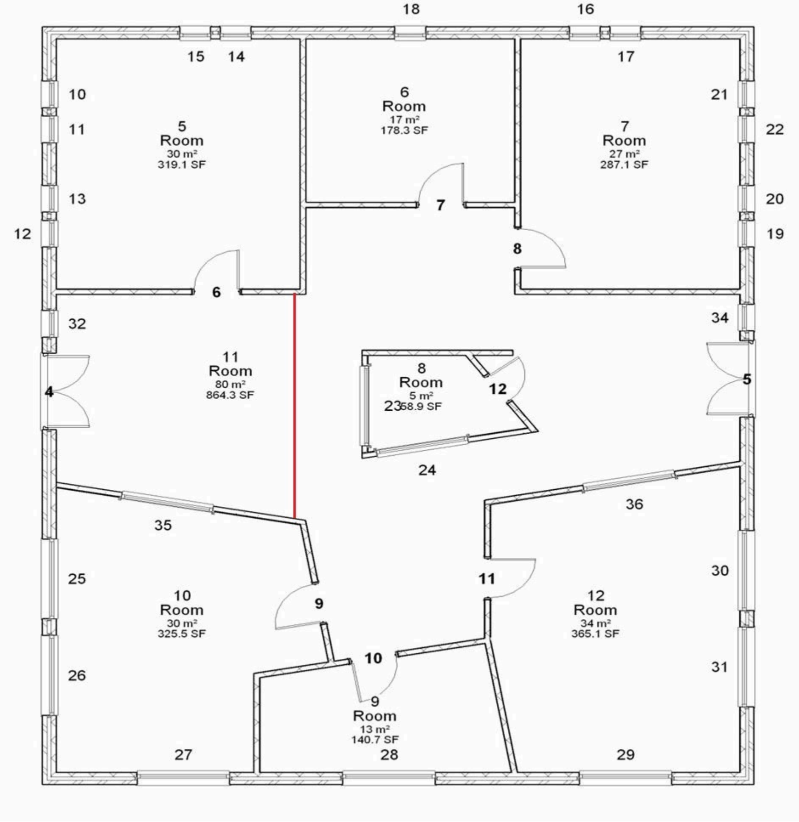


Figure 4. A different new state of the floor plan: the line segment is both well-formed and meaningful

Data and information types

The typology of data is a key component in Floridi's approach. Data can be:

- *Primary*, like the name and birth date of a person in a database, or the light emitted by an indicator lamp to show that a radio receiver is on.
- *Anti–data,*[4] i.e. the absence of primary data, like the failure of an indicator lamp to emit light or silence following having turned the radio on. Anti-data are informative: they tell us that e.g. the radio or the indicator lamp are defective.

- *Derivative:* data produced by other, typically primary data, which can therefore serve as indirect indications of the primary ones, such as a series of transactions with a particular credit card as an indication of the trail of its owner.
- *Operational*: data about the operations of the whole system, like a lamp that indicates whether other indicator lamps are malfunctioning.
- *Metadata*: indications about the nature of the information system, like the geographic coordinates that tell where a digital picture has been taken.

These types also apply to information instances, depending on the type of data they contain: an information instance containing metadata is meta-information.

In the context of analogue building representations like floor plans (Figure 5), lines denoting building elements are primary data. They describe the shape of these elements, their position and the materials they comprise.

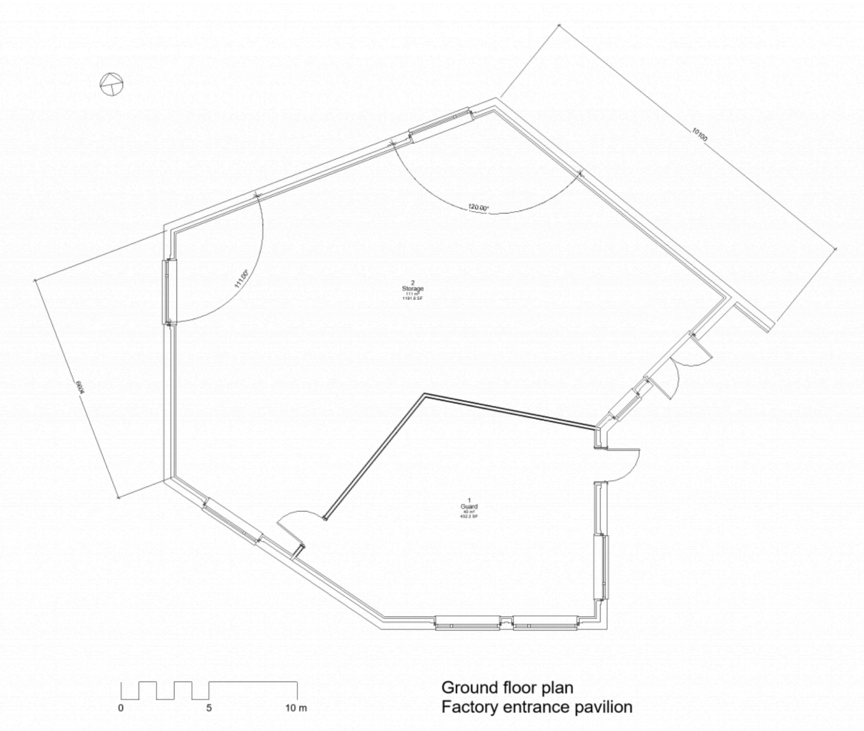


Figure 5. In an analogue floor plan, lines denoting building elements are primary data

In addition to such geometric primary data, an analogue floor plan may contain alphanumeric primary data, such as labels indicating the function of a room or dimension lines (Figure 6). A basic principle in hand drawing is that such explicitly specified dimensions take precedence over measurements in the drawing because amending these dimensions is easier than having to redraw the building elements.

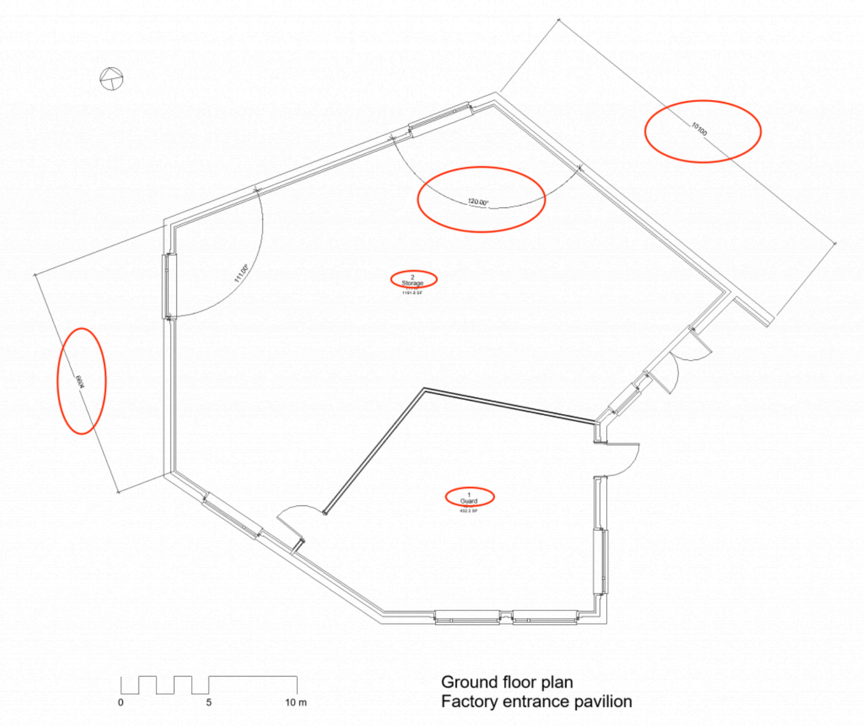


Figure 6. Alphanumeric primary data in an analogue floor plan

Anti-data are rather tricky to identify in building representations because of the abstraction and ellipsis that characterize them. Quite often it is hard to know if something is missing in a representation. One should therefore consider absence as anti-data chiefly when absence runs contrary to expectation and is therefore directly informative: a door missing from the perimeter of a room indicates either a design mistake or that the room is inaccessible (e.g. a shaft). Similarly, a missing room label indicates either that the room has no specific function or that the drawer has forgotten to include it in the floor plan (Figure 7).

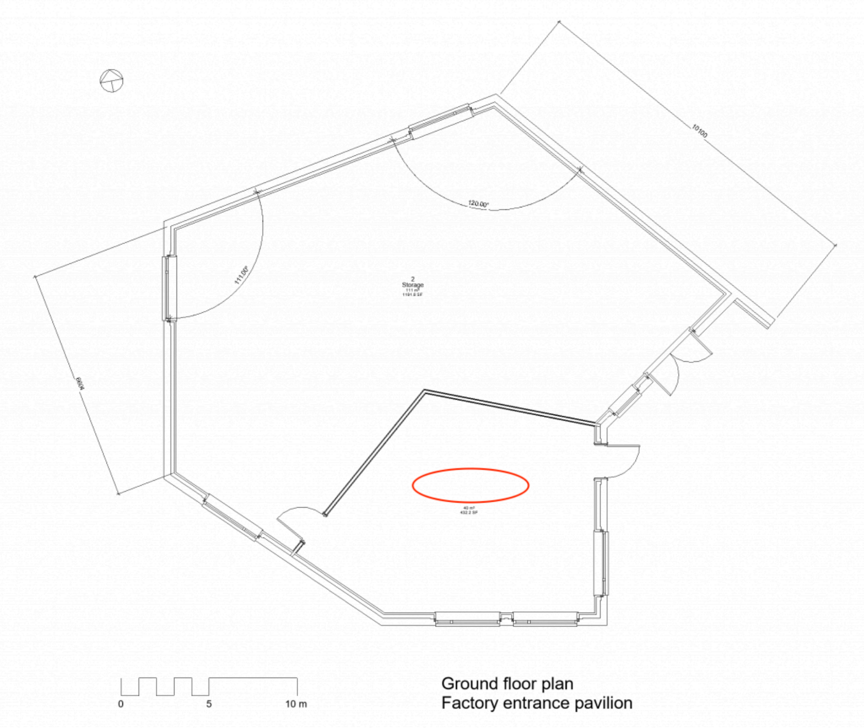


Figure 7. Anti-data in an analogue floor plan

Derivative data in building representations generally refer to the abundance of measurements, tables and other data produced from primary data in the representation, such as floor area labels in a floor plan (Figure 8). One can recognize derivative data from the fact that they can be omitted from the representation without reducing its completeness or specificity: derivative data like the area of a room can be easily reproduced when necessary from primary data (the room dimensions). An important point is that one should always keep in mind the conventions of analogue representations, like the precedence of dimension lines over measurement in the drawing, which turns the former into primary data.

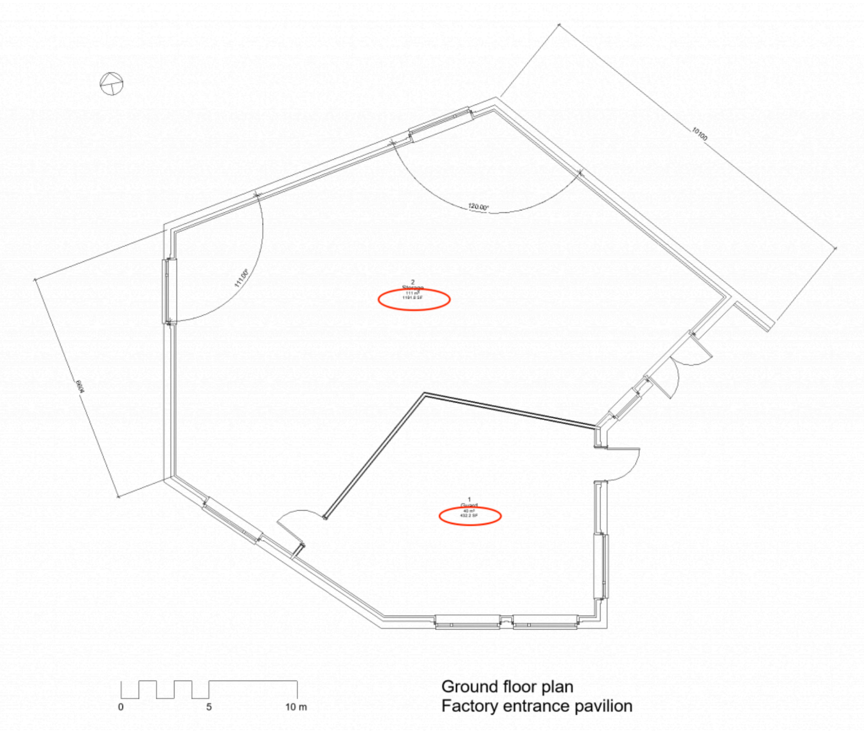


Figure 8. Derivative data in an analogue floor plan

Operational data reveal the structure of the building representation and explain how data should be interpreted. Examples include graphic scale bars and north arrows, which indicate respectively the true size of units measured in the representation and the true orientation of shapes in the design (Figure 9).

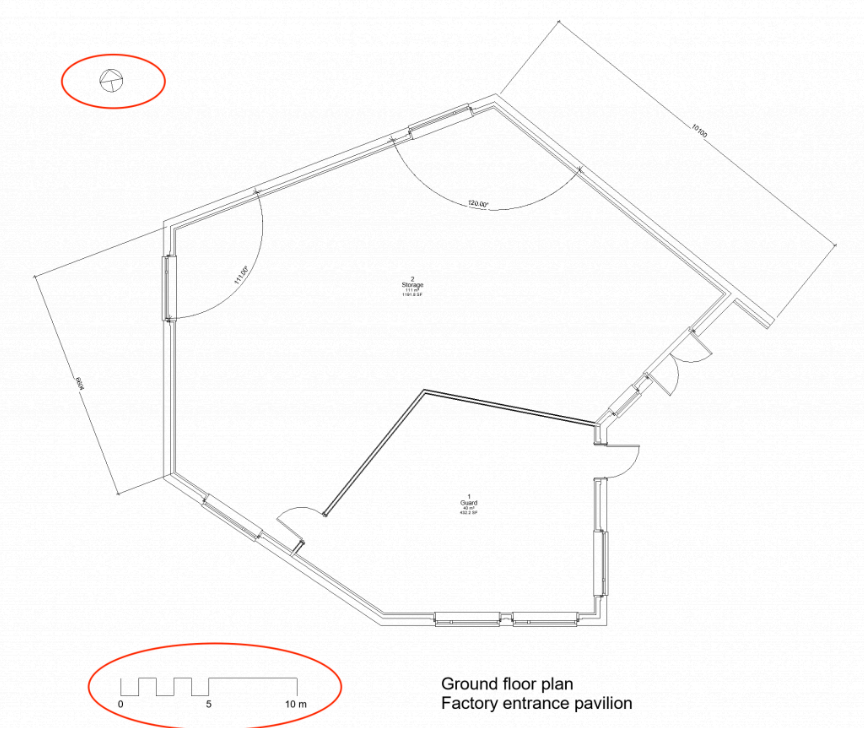


Figure 9. Operational data in an analogue floor plan

Finally, metadata describe the nature of the representation, such as the projection type and the design project or building, e.g. labels like 'floor plan' (Figure 10).

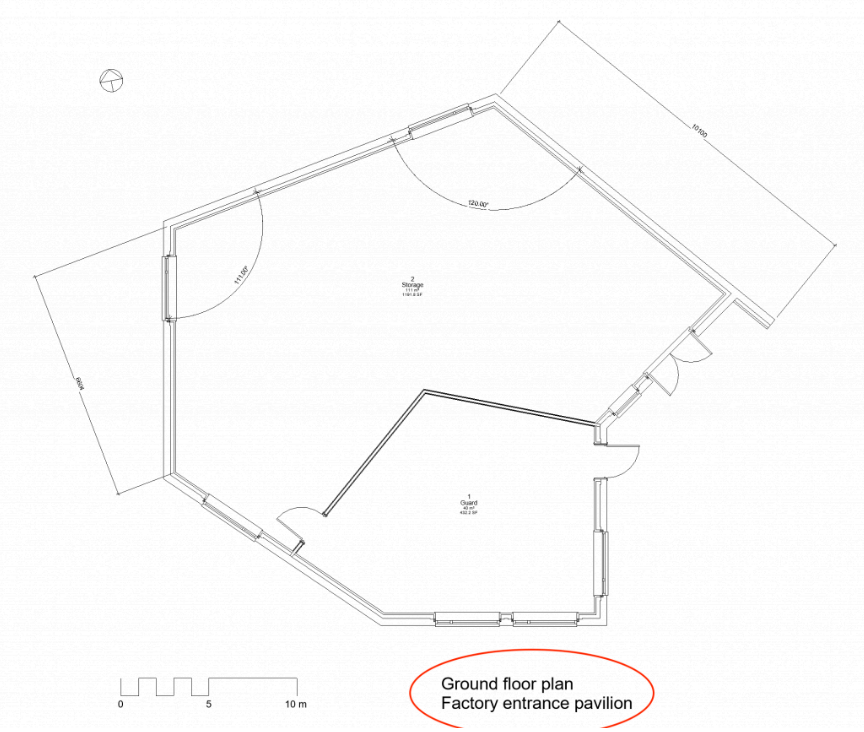
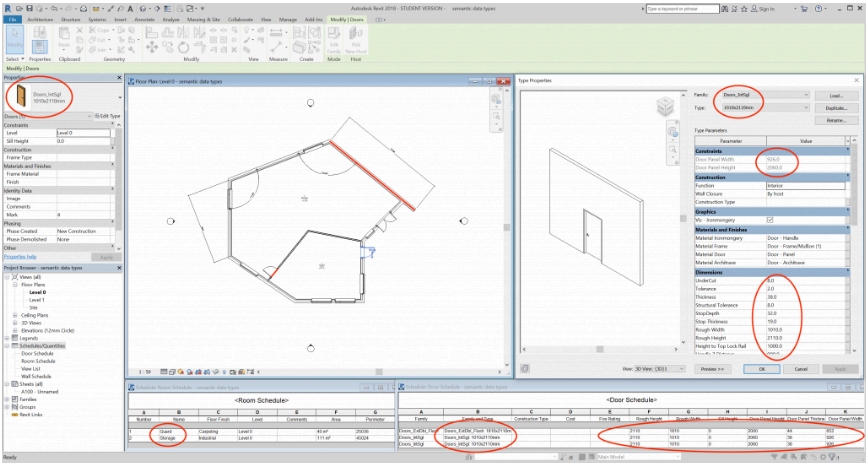


Figure 10. Metadata in an analogue floor plan

BIM, information and data

Data types in BIM

As we have seen in previous chapters, computerization does not just reproduce analogue building representations. Digital representations may mimic their analogue counterparts in appearance but can be quite different in structure – something that becomes apparent when we examine the data types they contain. Looking at a BIM editor on a computer screen, one cannot help observing a striking shift in primary and derivative data (Figure 11 & Figure 12): most graphic elements in views like floor plans are derived from properties of symbols. In contrast to analogue drawings, in BIM, dimension lines and values are derivative, pure annotations like floor area calculations in a space. This may be understandable given the ease with which one can modify a digital representation but even the lines denoting the various materials of a building element are derivative, determined by the type of the symbol: if the type of a wall changes, then all these graphic elements change accordingly. In analogue representations the opposite applies: we infer the wall type from the graphic elements that describe it in terms of layers of materials and other components.



Doors_IntSgl
1010x2110mm

Doors (1)

Constraints	
Level	Level 0
Sill Height	0.0
Construction	
Frame Type	
Materials and Finishes	
Frame Material	
Finish	
Identity Data	
Image	
Comments	
Mark	4
Phasing	
Phase Created	New Construction
Phase Demolished	None
Other	

Views (all)
- Floor Plans
  - Level 0
  - Level 1
  - Site
- Ceiling Plans
- 3D Views
- Elevations (12mm Circle)
- Legends
- Schedules/Quantities
  - Door Schedule
  - Room Schedule
  - View List
  - Wall Schedule
- Sheets (all)
  - A100 - Unnamed
- Families
- Groups
- Revit Links

Family: Doors_IntSgl
Type: 1010x2110mm
View: 3D View: {3D}1

Type Parameters

Parameter	Value
Constraints	
Door Panel Width	926.0
Door Panel Height	2060.0
Construction	
Function	Interior
Wall Closure	By host
Construction Type	
Graphics	
Vis - Ironmongery	☑
Materials and Finishes	
Material Ironmongery	Door - Handle
Material Frame	Door - Frame/Mullion (1)
Material Door	Door - Panel
Material Architrave	Door - Architrave
Dimensions	
UnderCut	8.0
Tolerance	2.0
Thickness	38.0
Structural Tolerance	8.0
StopDepth	32.0
Stop Thickness	19.0
Rough Width	1010.0
Rough Height	2110.0
Height to Top Lock Rail	1000.0

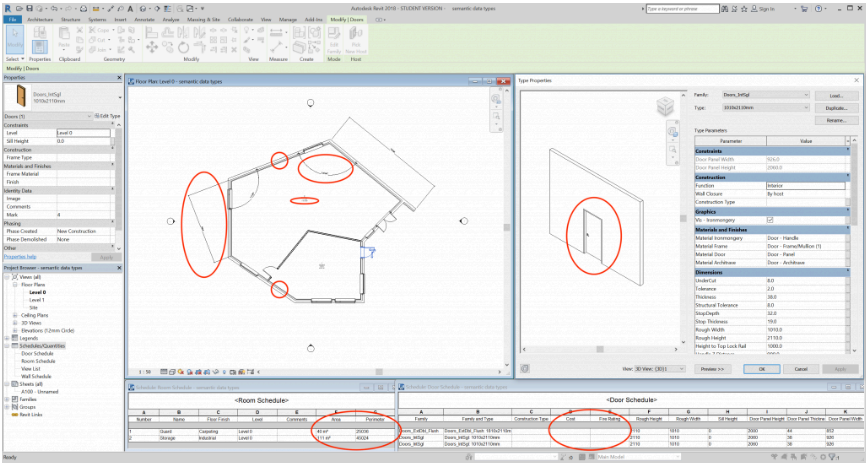
<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

<Door Schedule>

A	B	C	D	E	F	G	H	I	J	K
Family	Family and Type	Construction Type	Cost	Fire Rating	Rough Height	Rough Width	Sill Height	Door Panel Height	Door Panel Thickne	Door Panel Width
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110m				2110	1810	0	2000	44	852
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926

Figure 11. Primary data in BIM



Properties

Doors_IntSgl
1010x2110mm

Doors (1)

Constraints	
Level	Level 0
Sill Height	0.0
Construction	
Frame Type	
Materials and Finishes	
Frame Material	
Finish	
Identity Data	
Image	
Comments	
Mark	4
Phasing	
Phase Created	New Construction
Phase Demolished	None
Other	

Family: Doors_IntSgl
Type: 1010x2110mm

Type Parameters

Parameter	Value
Constraints	
Door Panel Width	926.0
Door Panel Height	2060.0
Construction	
Function	Interior
Wall Closure	By host
Construction Type	
Graphics	
Vis - Ironmongery	☑
Materials and Finishes	
Material Ironmongery	Door - Handle
Material Frame	Door - Frame/Mullion (1)
Material Door	Door - Panel
Material Architrave	Door - Architrave
Dimensions	
UnderCut	8.0
Tolerance	2.0
Thickness	38.0
Structural Tolerance	8.0
StopDepth	32.0
Stop Thickness	19.0
Rough Width	1010.0
Rough Height	2110.0
Height to Top Lock Rail	1000.0

View: 3D View: {3D}1

Schedule: Room Schedule - semantic data types

<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

Schedule: Door Schedule - semantic data types

<Door Schedule>

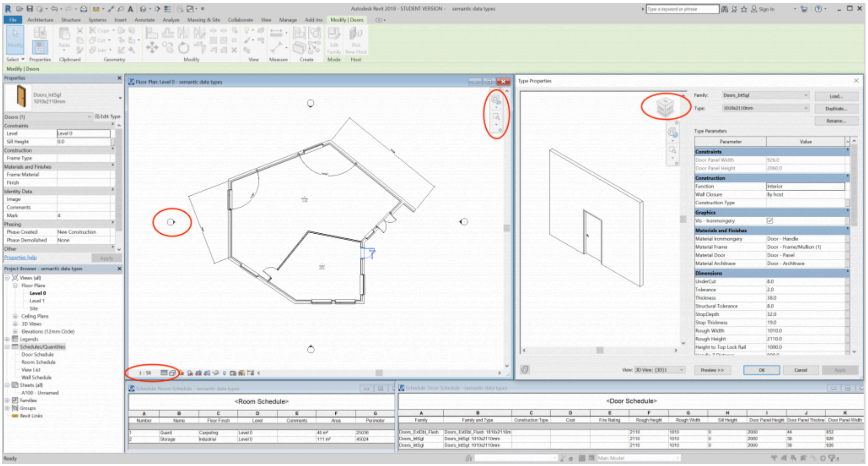
A	B	C	D	E	F	G	H	I	J	K
Family	Family and Type	Construction Type	Cost	Fire Rating	Rough Height	Rough Width	Sill Height	Door Panel Height	Door Panel Thickne	Door Panel Width
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110m				2110	1810	0	2000	44	852
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926

Figure 12. Derivative data in BIM

The main exception is the geometry of symbols. As described in the previous chapter, when one enters e.g. a wall in BIM, the usual procedure is to first choose the type of the wall and then draw its axis in a geometric view like a floor plan. Similarly, modifications to the location or shape of the wall are made by changing the same axis, while other properties, like layer composition and material properties of each layer, can only be changed in the definition of the wall type. One can also change the axis by typing new coordinates in some window but in most BIM editors the usual procedure is interactive modification of the drawn axis with a pointer device like a mouse. Consequently, primary data may appear dispersed over a number of views and windows, including ones that chiefly contain derivative data.

One should not be confused by the possibilities offered by computer programs, especially for the modification of entities in a model. The interfaces of these programs are rich with facilities to change shapes and values. It seems as if programmers have taken the trouble to allow users to utilize practically everything for this purpose. For example, one may be able to change the length of a wall by typing a new value for its dimension line, i.e. via derivative data. Such redundancy of entry points is highly prized for user interaction but may be confusing in terms of IM, as it tends to obscure the type of data and the location where each type can be found. To reduce confusion and hence the risk of mistakes and misunderstandings, one should consider the character of each view or window and how necessary it is for defining an entity in a model. A schedule, for example, is chiefly meant for displaying derivative data, such as area or volume calculations, but may also contain primary data for reasons of overview, transparency or legibility. Most schedules are not necessary for entering entities in a model, in contrast to a window containing the properties of a symbol, from where one chooses the type of the entity to be entered. In managing the primary data of a symbol one should therefore focus on the property window and its contents.

Computer interfaces also introduce more operational data, through which users can interact with the software. Part of this interaction concerns how other data are processed, including in terms of appearance, as with the scale and resolution settings in drawing views mentioned in the previous chapter (Figure 13).



Doors_IntSgl
1010x2110mm
Doors (1) Edit Type

Parameter	Value
Constraints	
Level	Level 0
Sill Height	0.0
Construction	
Frame Type	
Materials and Finishes	
Frame Material	
Finish	
Identity Data	
Image	
Comments	
Mark	4
Phasing	
Phase Created	New Construction
Phase Demolished	None
Other	

Properties help

Family: Doors_IntSgl
Type: 1010x2110mm

Type Parameters

Parameter	Value
Constraints	
Door Panel Width	926.0
Door Panel Height	2060.0
Construction	
Function	Interior
Wall Closure	By host
Construction Type	
Graphics	
Vis - Ironmongery	☑
Materials and Finishes	
Material Ironmongery	Door - Handle
Material Frame	Door - Frame/Mullion (1)
Material Door	Door - Panel
Material Architrave	Door - Architrave
Dimensions	
UnderCut	8.0
Tolerance	2.0
Thickness	38.0
Structural Tolerance	8.0
StopDepth	32.0
Stop Thickness	19.0
Rough Width	1010.0
Rough Height	2110.0
Height to Top Lock Rail	1000.0

View: 3D View: {3D}1

Schedule: Room Schedule - semantic data types

<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

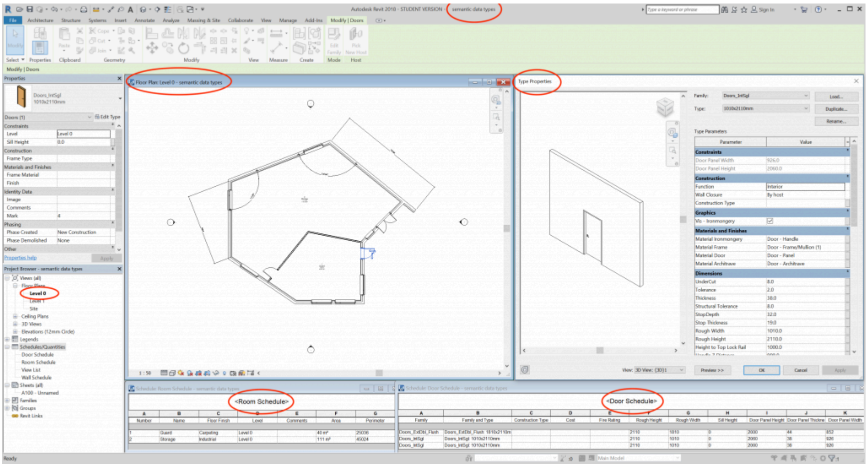
Schedule: Door Schedule - semantic data types

<Door Schedule>

A	B	C	D	E	F	G	H	I	J	K
Family	Family and Type	Construction Type	Cost	Fire Rating	Rough Height	Rough Width	Sill Height	Door Panel Height	Door Panel Thickne	Door Panel Width
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110m				2110	1810	0	2000	44	852
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926

Figure 13. Operational data in BIM

The presence of multiple windows on the screen also increases the number of visible metadata, such as window headers that describe the view in each window (Figure 14).



Doors_IntSgl
1010x2110mm

Doors (1)

Constraints	
Level	Level 0
Sill Height	0.0
Construction	
Frame Type	
Materials and Finishes	
Frame Material	
Finish	
Identity Data	
Image	
Comments	
Mark	4
Phasing	
Phase Created	New Construction
Phase Demolished	None
Other	

Family: Doors_IntSgl

Type: 1010x2110mm

Type Parameters

Parameter	Value
Constraints	
Door Panel Width	926.0
Door Panel Height	2060.0
Construction	
Function	Interior
Wall Closure	By host
Construction Type	
Graphics	
Vis - Ironmongery	☑
Materials and Finishes	
Material Ironmongery	Door - Handle
Material Frame	Door - Frame/Mullion (1)
Material Door	Door - Panel
Material Architrave	Door - Architrave
Dimensions	
UnderCut	8.0
Tolerance	2.0
Thickness	38.0
Structual Tolerance	8.0
StopDepth	32.0
Stop Thickness	19.0
Rough Width	1010.0
Rough Height	2110.0
Height to Top Lock Rail	1000.0

View: 3D View: {3D}1

Schedule: Room Schedule - semantic data types

<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

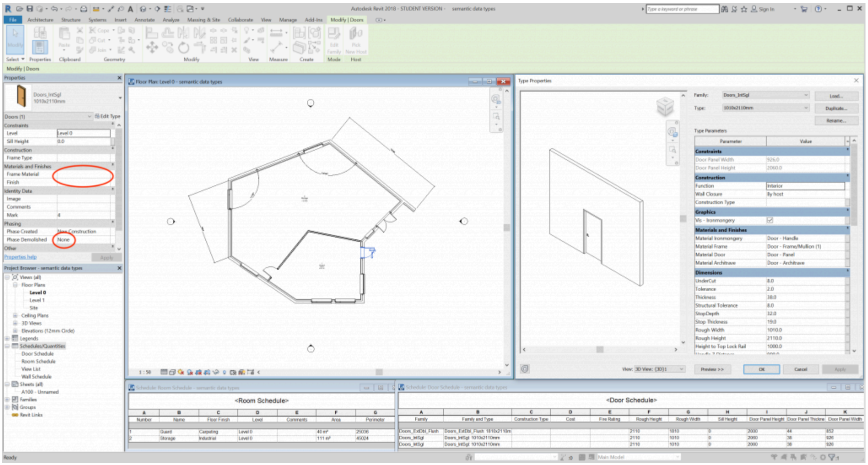
Schedule: Door Schedule - semantic data types

<Door Schedule>

A	B	C	D	E	F	G	H	I	J	K
Family	Family and Type	Construction Type	Cost	Fire Rating	Rough Height	Rough Width	Sill Height	Door Panel Height	Door Panel Thickne	Door Panel Width
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110m				2110	1810	0	2000	44	852
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926

Figure 14. Metadata in BIM

Anti-data remain difficult to distinguish from data missing due to abstraction or deferment. The lack of values for e.g. cost or fire rating for some building elements may merely indicate that their calculation has yet to take place, despite the availability of the necessary primary data. After all, both are calculated on the basis of materials present in the elements: if these materials are known, cost and fire ratings are easy to derive. One should remember the inherent duality of anti-data: they do not only indicate missing primary data but the presence of anti-data is significant and meaningful by itself. For example, not knowing the materials and finishes of a window frame, although the window symbol is quite detailed, signifies that the interfacing of the window to a wall is a non-trivial problem that remains to be solved. Interfacing typically produces anti-data, especially when sub-models meet in BIM, e.g. when the MEP and architectural sub-models are integrated, and the fastenings of pipes and cables to walls are present in neither. Anti-data generally necessitate action: no value (or "none") for the demolition phase of an entity suggests that the entity has to be preserved during all demolition phases – not ignored but actively preserved with purposeful measures, which should be made explicit (Figure 15).



Properties

Doors_IntSgl
1010x2110mm

Doors (1) · Edit Type

Parameter	Value
Constraints	
Level	Level 0
Sill Height	0.0
Construction	
Frame Type	
Materials and Finishes	
Frame Material	
Finish	
Identity Data	
Image	
Comments	
Mark	4
Phasing	
Phase Created	New Construction
Phase Demolished	None
Other	

Project Browser - semantic data types
- Views (all)
  - Floor Plans
    - **Level 0**
    - Level 1
    - Site
  - Ceiling Plans
  - 3D Views
  - Elevations (12mm Circle)
- Legends
- Schedules/Quantities
  - Door Schedule
  - Room Schedule
  - View List
  - Wall Schedule
- Sheets (all)
  - A100 - Unnamed
- Families
- Groups
- Revit Links

Family: Doors_IntSgl
Type: 1010x2110mm

Type Parameters

Parameter	Value
Constraints	
Door Panel Width	926.0
Door Panel Height	2060.0
Construction	
Function	Interior
Wall Closure	By host
Construction Type	
Graphics	
Vis - Ironmongery	☑
Materials and Finishes	
Material Ironmongery	Door - Handle
Material Frame	Door - Frame/Mullion (1)
Material Door	Door - Panel
Material Architrave	Door - Architrave
Dimensions	
UnderCut	8.0
Tolerance	2.0
Thickness	38.0
Structural Tolerance	8.0
StopDepth	32.0
Stop Thickness	19.0
Rough Width	1010.0
Rough Height	2110.0
Height to Top Lock Rail	1000.0

Schedule: Room Schedule - semantic data types

<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

Schedule: Door Schedule - semantic data types

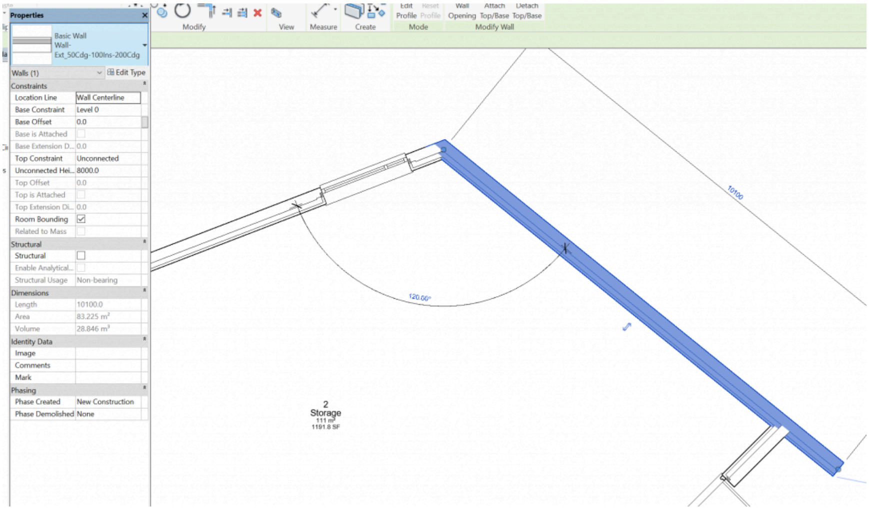
<Door Schedule>

A	B	C	D	E	F	G	H	I	J	K
Family	Family and Type	Construction Type	Cost	Fire Rating	Rough Height	Rough Width	Sill Height	Door Panel Height	Door Panel Thickne	Door Panel Width
Doors_ExtDbl_Flush	Doors_ExtDbl_Flush: 1810x2110m				2110	1010	0	2000	44	852
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926
Doors_IntSgl	Doors_IntSgl: 1010x2110mm				2110	1010	0	2060	38	926

Figure 15. Anti-data in BIM

Information instances in BIM

Identifying information instances in BIM starts with recognizing the data. As described in the previous section, data are to be found in the symbols: their properties and relations. In the various views and windows of BIM software, one can easily find the properties of each symbol, either of the instance (Figure 16, Figure 18) or of the type (Figure 17).



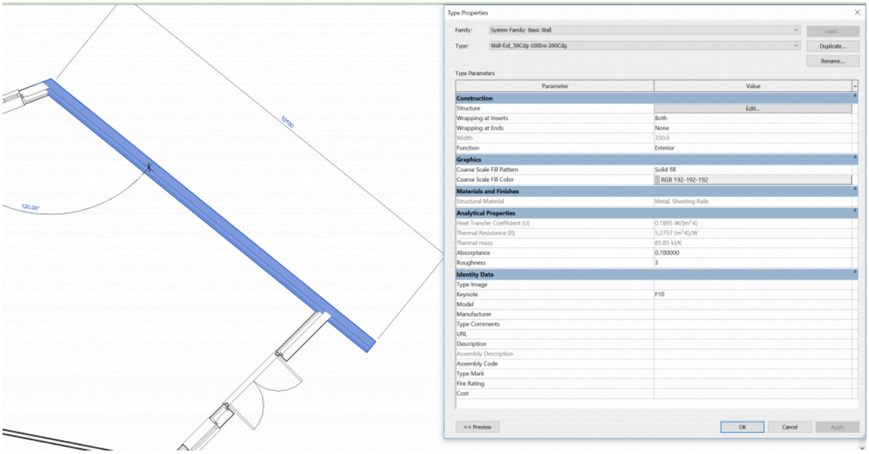
Properties

Basic Wall
Wall-Ext_50Cdg-100Ins-200Cdg

Walls (1) | Edit Type

Constraints	
Location Line	Wall Centerline
Base Constraint	Level 0
Base Offset	0.0
Base is Attached	☐
Base Extension D...	0.0
Top Constraint	Unconnected
Unconnected Hei...	8000.0
Top Offset	0.0
Top is Attached	☐
Top Extension Di...	0.0
Room Bounding	☑
Related to Mass	☐
Structural	
Structural	☐
Enable Analytical...	☐
Structural Usage	Non-bearing
Dimensions	
Length	10100.0
Area	83.225 m²
Volume	28.846 m³
Identity Data	
Image	
Comments	
Mark	
Phasing	
Phase Created	New Construction
Phase Demolished	None

Figure 16. Instance properties palette in a BIM editor (Revit)



Type Properties

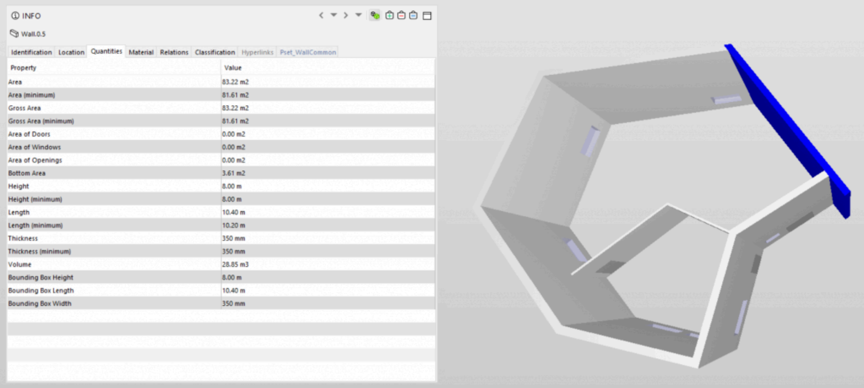
Family: System Family: Basic Wall

Type: Wall-Ext_50Cdg-100Ins-200Cdg

Type Parameters

Parameter	Value
Construction	
Structure	Edit...
Wrapping at Inserts	Both
Wrapping at Ends	None
Width	350.0
Function	Exterior
Graphics	
Coarse Scale Fill Pattern	Solid fill
Coarse Scale Fill Color	RGB 192-192-192
Materials and Finishes	
Structural Material	Metal, Sheeting Rails
Analytical Properties	
Heat Transfer Coefficient (U)	0.1895 W/(m²·K)
Thermal Resistance (R)	5.2757 (m²·K)/W
Thermal mass	85.85 kJ/K
Absorptance	0.700000
Roughness	3
Identity Data	
Type Image	
Keynote	F10
Model	
Manufacturer	
Type Comments	
URL	
Description	
Assembly Description	
Assembly Code	
Type Mark	
Fire Rating	
Cost	

Figure 17. Type properties window in a BIM editor (Revit)



Property	Value
Area	83.22 m2
Area (minimum)	81.61 m2
Gross Area	83.22 m2
Gross Area (minimum)	81.61 m2
Area of Doors	0.00 m2
Area of Windows	0.00 m2
Area of Openings	0.00 m2
Bottom Area	3.61 m2
Height	8.00 m
Height (minimum)	8.00 m
Length	10.40 m
Length (minimum)	10.20 m
Thickness	350 mm
Thickness (minimum)	350 mm
Volume	28.85 m3
Bounding Box Height	8.00 m
Bounding Box Length	10.40 m
Bounding Box Width	350 mm

Figure 18. Properties window in a BIM checker (Solibri)

What one sees in such a view or window is a mix of different data types, with derivative data like a volume calculation or thermal resistance next to primary data, such as the length and thickness of a wall. Moreover, no view or window contains a comprehensive collection of properties. As a result, when a property changes in one view, the change is reflected in several other parts of the interface that accommodate the same property or data derived from it.

Any lack of uniformity in these properties, including the addition of new symbols and their properties to a model, qualifies as data. One can restrict the identification of data to each view separately but it makes more sense for IM to include all clones of the same property, in any view. On the other hand, any derivative data that are automatically produced or modified as a result of the primary data count as different data instances. So, any change in the shape of a space counts as a single data instance, regardless of the view in which the user applies the change or of in how many views the change appears. The ensuing change in the space area value counts as a second instance of data; the change in the space volume as a third.

Relations between symbols are even more dispersed and often tacit. They can be found hidden in symbol behaviours (e.g. in that windows, doors or wash basins tend to stick to walls or in that walls tend to retain their co-termination), in explicit parametric rules and constraints, as well as in properties (e.g. construction time labels) that determine incidental grouping. Discerning lacks of uniformity in relations is therefore often hard, especially since most derive variably from changes in the symbols. For example, modifying the length of a wall may inadvertently cause its co-termination with another wall to be removed or, if the co-termination is retained, to change the angle between the walls.

Many relations can be made explicit and controllable through appropriate views like schedules. As we have seen, window and door schedules make explicit relations between openings and spaces. This extends to relations between properties of windows or doors and of the adjacent spaces, e.g. connects the fire rating of a door to whether a space on either side is part of a main fire egress route or the acoustic isolation offered by the door to the noise or privacy level of activities accommodated in either adjacent space.

Information instances can be categorized by the type of their data: primary, derivative, operational etc. This is important for IM, as it allows one to, firstly, prioritize in terms of significance and, secondly, to link information to actors and stakeholders. Primary information obviously carries a higher priority than derivative. Moreover, primary information (e.g. the shape of spaces) is produced or maintained by specific actors (e.g. designers), preferably with no interference by others who work with derivative information (e.g. fire engineers). So, information instances concerning space shape are fed forward from the designers to the fire engineers, whose observations or recommendations are fed back to the designers, who then initiate possible further actions and produce new data. Understanding these flows, the information types they convey and transparently linking instances to each other and to actors or stakeholders is essential for IM.

Another categorization of information instances concerns scope. This leads to two fundamental categories:

1. Instances comprising one or more properties or relations of a single symbol: the data are produced when one enters the symbol in the representation or when the symbol is modified, either interactively by a user or automatically, e.g. on the basis of a built-in behaviour, parametrization etc. Instances of this category are basic and homogeneous: they refer to a single entity of a particular kind, e.g. a door. The entity can be:
    1. Generic in type, like an abstract internal door
    2. Contextually specific, such as a door for a particular wall in the design, i.e.

    partially defined by relations in the representation
    3. Specific in type, e.g. a specific model of a particular manufacturer, fixed in all its properties
2. Instances comprising one or more properties or relations of multiple symbols, added or modified together, e.g. following a change of type for a number of internal walls, or a resizing or repositioning of the building elements bounding a particular space. Consequently, instances of this category can be:
    1. Homogeneous, comprising symbols of the same type, e.g. all office spaces in a building
    2. Heterogeneous, comprising symbols of various types, usually related to each other in direct, contextual ways, e.g. the spaces and doors of a particular wing that make up a fire egress route

These two categories can account for all data and abstraction levels in a representation, from sub-symbols (like the modification of the geometry of a door handle in the definition of a door type) to changes in the height of a floor level that affects the location of all building elements and spaces on that floor, the size and composition of some (e.g. stairs) and potentially also relations to entities on adjacent floors.

Key Takeaways

- *An instance of information consists of one or more data which are well-formed and meaningful.*
- *Data are lacks of uniformity in what we perceive at a given moment or between different states of a percept or between two symbols in a percept.*
- *Data can be primary, anti-data, derivative, operational or metadata.*
- *There are significant differences between analogue and digital building representations concerning data types, with symbols like dimension lines being primary in the one and derivative in the other.*
- *In BIM lacks of uniformity can be identified in the properties and relations of symbols.*
- *Information instances can be categorized by the semantic type of their data and by their scope in the representation.*

Exercises

1. Identify the semantic data types in the infobox of a Wikipedia biographic lemma (the summary panel on the top right), e.g. https://en.wikipedia.org/wiki/Aldo_van_Eyck (Figure 19),[5] and in the basic page information of the same lemma (e.g. https://en.wikipedia.org/w/index.php?title=Aldo_van_Eyck&action=info)

Aldo van Eyck

Aldo van Eyck in 1970

Born	March 16, 1918 Driebergen, Netherlands
Died	January 14, 1999 (aged 80) Loenen aan de Vecht, Netherlands
Nationality	Dutch
Alma mater	ETH Zurich
Occupation	Architect

Figure 19. Infobox in Wikipedia

2. Explain the information instances produced in BIM when one inserts a door in an existing wall. Use the following notation (expanding it if necessary): *(scope; symbol; name of property or relation; value of property or relation; time; semantic data type)*
If the instances concern multiple symbols, use the notation to describe each symbol separately.

3. Explain the information instances produced in BIM when one moves an existing door to a slightly different position in an existing wall. Use the above notation for each concerned symbol separately.
4. In BIM it is claimed that one can add information dimensions to the three geometric dimensions, turning 3D into *n*D: 4D comes with the addition of time (e.g. when the symbolized entity is constructed), 5D with the addition of cost, 6D with sustainability, 7D with facility management a, 8D with accident prevention (or safety) etc. For something to qualify as a dimension, it should be primary and not derivative, otherwise area and volume would be dimensions, too. Which of the above “dimensions” contain primary data and should therefore be acceptable as true dimensions? Do you know of other properties of a symbol that would qualify as dimensions? Use doors, windows, walls or floors as examples.
5. IFC (Industry Foundation Classes) is a standard underlying BIM, in particular for the way each entity is represented. Identify the semantic data types in the IFC wall base quantities, i.e. quantities that are common to the definition of all occurrences of walls (http://www.buildingsmart-tech.org/ifc/IFC4/final/html/schema/ifcsharedbldgelements/qset/qto_wallbasequantities.htm), with particular attention for derivative quantities present in the specification. If each of the quantities becomes a symbol property in BIM, calculate how much of a typical model consists of derivative data, both in percentage and megabytes (assuming that what holds for walls also holds for all entities in BIM).

Notes

---

1. There are several fundamental sources on the MTC, starting with the original publication: Shannon, C., 1948. A mathematical theory of communication. *Bell System Technical Journal, 27*(July, October), 379-423, 623-656; Shannon, C.E., & Weaver, W., 1998. *The mathematical theory of communication*. Urbana IL: University of Illinois Press; Cover, T.M., & Thomas, J.A., 2006. *Elements of information theory* (2nd ed.). Hoboken NJ: Wiley-Interscience; Pierce, J.R., 1980. *An introduction to information theory : symbols, signals & noise* (2nd, rev. ed.). New York: Dover.
2. The classification of theories of information is after: Sommaruga, G., 2009. Introduction. G. Sommaruga (ed), Formal Theories of Information: From Shannon to semantic information theory and general concepts of information. Berlin, Heidelberg: Springer.
3. Floridi's theory has been published in: Floridi, L., 2008. Trends in the philosophy of information. P. Adriaans & J. v. Benthem (eds), *Philosophy of information*. Amsterdam: North-Holland; Floridi, L., 2009. Philosophical conceptions of information. G. Sommaruga (ed), *Formal Theories of Information: From Shannon to semantic information theory and general concepts of information*. Berlin, Heidelberg: Springer; Floridi, L., 2016. Semantic con-

ceptions of information. *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2016/entries/information-semantic/

4. In later publications Floridi has preferred the term secondary data instead of anti-data but the new name seems rather confusing, suggesting data of a lesser importance rather than the converse of primary data.
5. Source: https://en.wikipedia.org/wiki/Aldo_van_Eyck) ; photograph credit:: Aldo van Eyck in 1970 by Bert Verhoef is licensed under CC BY-SA 3.0 NL

6. Information management

This chapter introduces the general goals of IM and connects them to information sources on buildings in order to determine the fundamental principles of IM with BIM.

The need for information management

With the information explosion we have been experiencing, it is hardly surprising that IM seems to have become a self-evident technical necessity. Handling the astounding amounts of information produced and disseminated every day requires more robust and efficient approaches than ever. Nevertheless, IM is considered mosty as a means to an end, usually performance in a project or enterprise: with effective IM, one can improve the chances of higher performance. Consequently, IM usually forms a key component of overall management.

This is widely acknowledged in building design management. Even before the digital era, the evident dependence of AECO on information coming from various sources and regarding various but interconnected aspects of a building had led to agreement that information and the way it is handled can be critical for communication and decision making. DM often focuses on information completeness, relevance, clarity, accuracy, quality, value, timeliness etc., so as to enable greater productivity, improve risk management, reduce errors and generally raise efficiency and reliability. The dependence on information is such that some even go so far as to suggest that DM is really fundamentally about IM: managing information flows so that stakeholders receive the right information at the right time.[1]

In practical terms, however, there was little clarity concerning what should be managed and how. DM sources often simply affirm that information is important and should be treated with care. What makes information usable, valuable, relevant etc. is assumed to be known tacitly. Information is fundamentally correctly defined as data in usable form. Predictably, however, it is also equated to the thousands of drawings and other documents produced during the lifecycle of a building. If the right document is present, then it is assumed that stakeholders also possess the right information and are directly capable of judging the veracity, completeness, coherence etc. of the information they receive or need. However, equating information with documents not only places a heavy burden on users, it also prolongs attachment to analogue practices in the digital era.

It is arguably typical of AECO and DM that, in the face of operational and especially technical complexity, they invest heavily in human resources. This goes beyond the interpretation of documents in order to extract information; it also includes the invention of new roles that assume a mix of new and old tasks and responsibilities. So, in addition to project and process managers, one encounters not only information managers but also BIM managers, CAD managers, BIM coordinators and CAD coordinators, working together in complex, overlapping hierarchies. These new roles are usually justified by the need for support concerning new technologies, which may be yet unfamiliar to the usual participants in an AECO project. At the same time, however, they increase complexity and reduce transparency by adding more intermediaries in the already multi-layered structure of AECO. They moreover increase the distance between AECO stakeholders and new technologies, frequently limiting learning opportunities for the stakeholders.

New roles, either temporary or permanent, may be inevitable with technological innovation. In the early days of motorcars, for example, chauffeurs were more widely employed to drive them than today, while webmasters have become necessary by the invention and popularity of the World Wide Web and remain so for the foreseeable future, despite growing web literacy among general users. However, such new roles should be part of a sound and thorough plan of approach rather than an easy alternative to a good approach. The plan should determine what is needed and why, taking into account the increasing familiarity and even proficiency of many users with various technologies, to a degree that they require little day-to-day support. In our case, one may expect that AECO professionals will eventually become quite capable not only of using BIM directly but also of coordinating their BIM activities, with little need for technical intermediaries. After all, that was the case with analogue drawings in the past. To achieve this, AECO needs practical familiarization with the new technologies but above all clear comprehension of what these technologies do with information. Based on that, one can develop a sound IM approach that takes into account both domain needs and the capacities of digital technologies, determine changes in the tasks, responsibilities and procedures of existing AECO roles, and develop profiles for any additional roles.

Information sources

Inclusiveness

IM^2 has a broad scope and, as a result, is quite inclusive. It pays no attention to issues of representation and accepts as information sources all kinds of documents, applications, services and schemes. This is due to three reasons. Firstly, IM covers many application areas and must therefore be adaptable to practices encountered in any of them. Secondly, in many areas there is a mix of analogue and digital information, as well as various channels, for example financial client transactions with a shop using cash and debit or credit cards, either physically or via a web shop. IM provides means for bringing such disparate material together into more coherent forms, ensuring that no outdated or inappropriate information is used and preventing that information is missing, inaccessible or deleted by error. These means include correlation with context (e.g. time series displays relative to other data), classification and condensation (aggregation, totalling, filtering and summarization). Thirdly, IM has a tenuous relation to computerization, often relying on it but also appearing weary of putting too much emphasis on technology to the detriment of information and organization.

The inclusiveness of IM with respect to information sources means that it may end up not only tolerating the redundancy of analogue and digital versions of the same information but also supporting outdated practices and conventions, even prolonging their life through superficial digitization. It may also reduce IM to mere document management, i.e. making sure that the necessary documents are retained and kept available. This seems like an easy way out of most domain problems. As the content and the expanse of the Internet suggest, there may be enough computer power and capacity to store and retrieve any document produced in a project or enterprise – in our case, throughout the whole lifecycle of a building (although one should question whether this also applies to all buildings in the world). On the other hand, however, both the information explosion in the digital era and big data approaches suggest the opposite: we already need more intelligent solutions than brute force. At this moment, we may think we still have control over the huge amounts of information in production and circulation but the IoT could change that soon, as smart things start communicating with each other with great intensity. For AECO this can be quite critical, since buildings are among the prime candidates for accommodating a wide range of sensors and actuators, e.g. for controlling energy consumption.

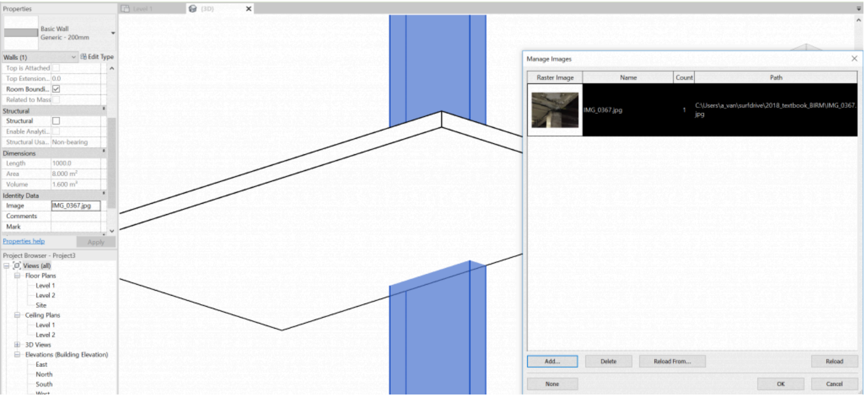
Structured, semi-structured and unstructured information

It is important for IM that BIM marks a transition not only to symbolic representation but also to holistic, *structured* information solutions for AECO. With regard to structure, there are three main categories:

- *Unstructured data* are the subject of big data approaches: sensor measurements, social media messages and other data without a uniform, standardized format. Finding relevant information in unstructured data is quite demanding because queries have to take into account a wide range of locations where meaningful data may reside and a wide variety of storage forms (including natural language and images).
- *Semi-structured data* are a favourite of IM: information sources with a loosely defined structure and flexible use. Analogue drawings are a typical example: one knows what is expected in e.g. a section but there are several alternative notations and few if any prohibitions concerning what may be depicted and how. IM thrives on semi-structured sources, adding metadata, extracting and condensing, so as to summarize relevant information into a structured overview.
- *Structured data* are found in sources where one knows precisely what is expected and where. Databases are prime examples of structured information sources. In a relational database, one knows that each table describes a particular class of entities, that each record in a table describes a single entity and that each field describes a particular property of these entities in the same, predefined way. Finding the right data in a structured source is therefore straightforward and less challenging for IM.

In contrast to analogue drawings, BIM is clearly structured, along the lines of a database. Each symbol belongs to a particular type and has specific properties. This structure is one of the driving forces behind BIM, in particular with respect to its capacity to integrate and process building information coherently. Given the effort put into developing structured models in BIM, it makes little sense to abandon the advantages they promise. More specifically, a parsimonious approach to IM with BIM should:

- Avoid having other primary information sources next to BIM: any building information should be integrated in BIM and all related data linked to it. Currently, there is general agreement that the price of a component, e.g. a washbasin, should be a property of the corresponding symbol. However, the same should apply to packaging information for this component, including the dimensions of the box in which the washbasin is brought to the building site, as this is useful for logistic purposes. Trying to retrieve this information from the manufacturer's catalogue is significantly less efficient than integrating the relevant data among the symbol properties. The same applies to a photograph of some part of the building during construction or use: this too should be connected to BIM as a link between the digital file of the photograph and relevant symbols in the model (Figure 1) or even mapped as a decal on the symbols (Figure 2).
- Desist from promoting BIM output to the level of a primary source: any view of a model, from a floor plan to a cost calculation, can be exported as a separate document (PDF, spreadsheet etc.). This may have its uses but one should not treat such exports as sources separate from the model. Any query about the building, including the history of such output, should start from the model. Using IM to ensure consistency between exports and the model is meaningless. This applies even to legally significant documents like contracts because these too can be expressed as views of the model (i.e. textual frames around data exported from the model).



Basic Wall
Generic - 200mm

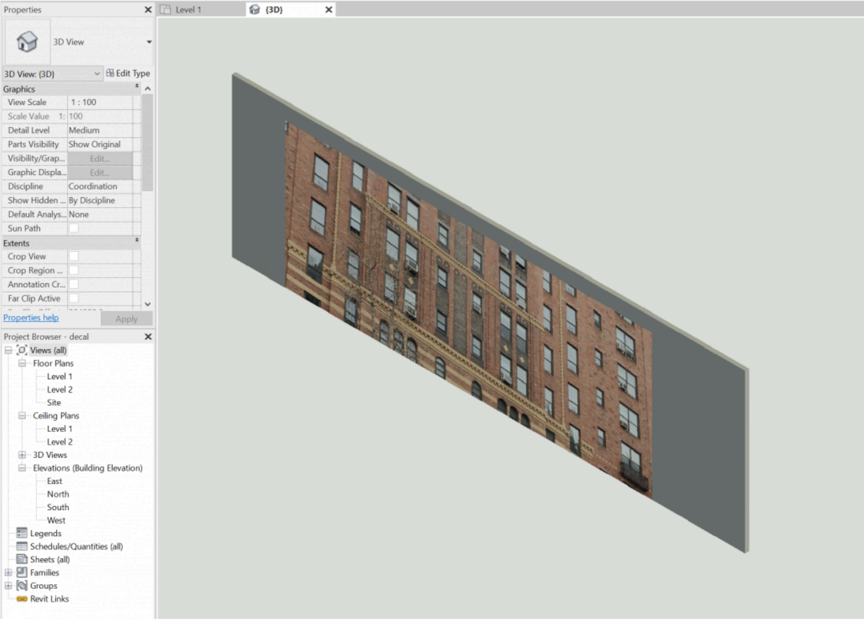
Walls (1)

Top is Attached	☐
Top Extension...	0.0
Room Boundi...	☑
Related to Mass	☐
Structural	
Structural	☐
Enable Analyti...	☐
Structural Usa...	Non-bearing
Dimensions	
Length	1000.0
Area	8.000 m²
Volume	1.600 m³
Identity Data	
Image	IMG_0367.jpg
Comments	
Mark	

Manage Images

Raster Image	Name	Count	Path
	IMG_0367.jpg	1	C:\Users\a_van\surfdrive\2018_textbook_BIRM\IMG_0367.jpg

Figure 1. Photograph of current state linked as image to relevant components in Revit



3D View: {3D}

Graphics	
View Scale	1 : 100
Scale Value 1:	100
Detail Level	Medium
Parts Visibility	Show Original
Visibility/Grap...	Edit...
Graphic Displa...	Edit...
Discipline	Coordination
Show Hidden ...	By Discipline
Default Analys...	None
Sun Path	☐
Extents	
Crop View	☐
Crop Region ...	☐
Annotation Cr...	☐
Far Clip Active	☐

Figure 2. Photograph of current state mapped as decal in Revit

From the above, a wider information environment emerges around the model, populated largely by files linked to the model, preferably to specific symbols. IM can assist with the organization of this environment, even allowing queries to be answered on the basis of such satellite documents, but the successful deployment of BIM depends on transparent links between these queries and documents and the model itself: any query should ultimately lead to primary data and their history in the model.

It is perhaps ironic that while the world is focusing on big, unstructured data, AECO should insist on structured data. One explanation is latency: AECO has been late with the development of structured information solutions because it continued to use analogue, semi-structured practices in digital facsimiles. As a consequence, AECO has yet to find the limits of structured data, although this may happen soon, when the IoT becomes better integrated in building design and management.

The emphasis on the structured nature of BIM also flies in the face of IM and its inclusiveness. In this respect, one should keep in mind that IM is a means, not an end, and that its adaptability has historical causes. It is not compulsory to retain redundant information sources next to BIM, simply because IM can handle redundancy and complexity. If the structured content of BIM suffices, then IM for AECO simply becomes easier and parsimonious.

Information management goals

Information flow

The first of the two main goals of IM is to regulate information flows. This is usually achieved by specifying precise processing steps and stages, which ensure that information is produced and disseminated on time and to the right people, until it is finally archived (or disposed of). In terms of the semantic information theory proposed in this book, this involves identifying and tracking information instances throughout a process, covering both the production and modification of data. In IM there is an emphasis on the sources and stores of information: the containers and carriers from where information is drawn, rests or is archived. BIM combines all these into a single information environment, shifting attention to the symbols, their properties and relations, where the data of information instances are found.

Managing information flow involves:

- *What:* the information required for or returned by each specific task in a process
- *Who:* the actors or stakeholders who produce or receive the information in a task
- *How:* the processing of information instances
- *When:* the timing of information instances

What is about information instances and symbols, as discussed in the previous section. However, despite the integration potential of BIM, which makes most information internal, some data may reside outside of models, e.g. weather data required for a thermal simulation. Connectivity to external sources is also part of IM.

For both internal and external information, it is critical to distinguish between *authorship* and *custodianship*: the actors who produce some information are not necessarily the same stakeholders who safeguard this information in a project, let alone during the lifecycle of a building. A typical example is briefing information: this is usually compiled in the initiative stage by a specialist on the basis of client and user input, as well as professional knowledge. In the development stage, custodianship may pass on to a project manager who utilizes it to evaluate the design, possibly adapting the brief on the basis of insights from the design. Then in the use stage, it becomes a background to facility and property management, before it develops into a direct or indirect source for a new brief, e.g. for the refurbishment of the building. Making clear in all stages who is the custodian of this information is of paramount importance in an integrated environment like BIM, where overlaps and grey areas are easy to develop.

How information flows are regulated relates to the syntagmatic dimension of a model: the sequence of actions through which symbols, their properties and relations are processed. The information instances produced by these actions generally correspond to the sequence of tasks in the process but are also subject to extrinsic constraints, often from the software: the presence of bounding walls is necessary for defining a space in most BIM editors, although in many design processes one starts with the spatial design rather than with construction. IM needs to take such conflicts into account and differentiate between the two sequences.

A useful device for translating tasks into information actions is the tripartite scheme *Input-Processing-Output (I-P-O),* which underlies any form of information processing: for any task, some actors deliver information as input; this input is then processed by some other (or even the same) actors. These return as output some other information, which usually becomes input for the next task. IM has to ensure that the right input is delivered to the right actors and that the right output is collected. By considering each decision task with respect to I-P-O, one can identify missing information in the input and arrange for its delivery.

The syntagmatic dimension obviously also relates to *when:* the moments when information instances become available. These moments usually form a coherent time schedule. The time schedule captures the process of actions and transactions, linking each to specific information instances. Here again one should differentiate between the sequence of tasks, which tends to be adequately covered by a project schedule, and the sequence of information actions, which may require additional refinement.

Information flow in BIM

We are used to viewing the early part of a design process as something almost magical: someone somehow puts a few lines on a scrap of paper and suddenly we have a basis for imagining what the building will look like. The same applies to BIM: one starts entering symbols in a model and the design is there for all to see and process. Building information flows seem to emerge out of nothing but this is far from true. The designers who make the first sketches or decide on the first elements in a model operate on the basis of general knowledge of their disciplines, specific knowledge of the kind of building they are designing and specific project information, including location characteristics and briefs. In other words, building representations are the product of cognitive processes that combine both tacit and overt information.

It is also widely assumed that the amount of information in a design process grows from very little in early design to substantial amounts by the end, when a building is fully specified. This actually refers to the specificity of conventional building representations, e.g. the drawing scales used in different design stages.

In fact, even before the first sketch is made, there usually is considerable information available on the building. Some of it predates the project, e.g. planning regulations and building codes that determine much of the form of a building and key features of its elements, such as the pitch of the roof and the dimensions of stairs. Other information belongs to the project, e.g. the brief that states accommodation requirements for the activities to be housed in the building, the budget that constrains cost or site-related principles like the continuation of vistas or circulation networks in the neighbourhood through the building site. Early building representations may conform to such specifications but most information remains in other documents or in the mind of the designers. For example, in many cases, one starts drawing or modelling a design with a site plan onto which building elements and spaces are placed but the site plan rarely includes planning regulations.

In managing building information, one should ensure that this information becomes explicit and is connected to subsequent tasks. In BIM, this amounts to augmenting the basic model setup (site plan, floor height and grids) with constraints from planning regulations (e.g. in the form of the permissible building envelope), use information from the brief and constraints on the kind of building elements that are admissible in the model (e.g. with respect to the fire rating of the building). Integration of such information amounts to *feedforward:* measurement and control of the information system before disturbances occur. Feedforward is generally more efficient and effective than feedback, e.g. checking if all building elements meet the fire safety requirements after they have been entered in the model.

It has also been suggested that early design decisions have a bigger impact on the outcome of a design process than later decisions. Having to decide on the basis of little overt information makes such decisions difficult and precarious. This conventional wisdom concerning early decisions may be misleading. Admittedly, early design decisions tend to concern basic features and aspects, from overall form to load-bearing structure, which determine much of the building and so have a disproportionate influence on cost and performance. However, such decisions are not exclusive to early design: the type of load-bearing structure can change late in the process, e.g. in relation to cost considerations or the need for larger spans. Such a late change can be more expensive because it also necessitates careful control of all interfacing between load-bearing and other elements in the design. From an IM perspective, what matters is to make all relevant information explicit in BIM, so as to know which data serve as input for a task (processing) and register the output of the task. Explicitness of information allows one to map decision making in a process and to understand the significance of any decision, regardless of process stage.

Information quality

The second main goal of IM is to safeguard or improve information quality.[3] Quality matters to IM in two respects. Firstly, concerning information utility: knowing that the information produced and disseminated in a process meets the requirements of its users. Secondly, concerning information value: information with a higher quality needs to be preserved and propagated with higher priority. IM measures quality pragmatically, in terms of relevance, i.e. fitness for purpose: how well the information supports the tasks of its users. In addition to pragmatic information quality, IM is also keen on inherent information quality: how well the information reflects the real-world entities it represents.

In both senses, information quality is determined within each application environment. IM offers a tactical, operational and technical framework but does not provide answers to domain questions. These answers have to be supplied by the application environment in order for IM to know which information to preserve, disseminate or prioritize. It should be noted that IM is not passive with regard to information quality. It can also improve it both at meta-levels (e.g. by systematically applying tags) and with respect to content (e.g. through condensation).

Information quality concerns the paradigmatic dimension: the symbols of a representation and their relations. As this dimension tends to be quite structured in symbolic representations, one can go beyond the pragmatic level of IM and utilize the underlying semantic level to understand better how information quality is determined.

The first advantage of utilizing the semantic level lies in the definition of acceptable data as being well-formed and meaningful. This determines the fundamental quality of data: their acceptability within each representation. A coffee stain cannot be part of a building representation but neither can a line segment be part of a model in BIM: it has to be a symbol that has the appearance of a line segment (i.e. uses the line segment as implementation mechanism), e.g. a room separation line in Revit, the most abstract of bounding elements. By the same token, a colour is not acceptable as a description of the material of a wall and a floor cannot be host to a door (except for a trapdoor). In conclusion, any data that do not fit the specifications of a symbol, a property or a relation cannot be well-formed or meaningful in BIM. Therefore, they have low quality, which requires attention. If quality cannot be improved, these data should be ignored as noise.

Data that pass the fundamental semantic test must then be evaluated concerning *relevance* for the particular building or project and its tasks. To judge relevance, one needs additional criteria, e.g. concerning specificity: it is unlikely that a model comprising generic building elements is satisfactory for a task like the acoustic analysis of a classroom because the property values of generic elements tend to be too vague regarding factors that influence acoustic performance.

The semantic level also helps to determine information value beyond utility: prioritizing which information should be preserved and propagated relates to semantic type. As derivative data can be produced from primary data when needed, they do not have to be prioritized – in many cases, they do not have to be preserved at all. Operational data and metadata tend to change little and infrequently in BIM, so these too have a lower priority relative to primary data. Finally, anti-data have a high priority, both because they necessitate interpretation and action, and because such action often aims at producing missing primary data.

Parsimonious IM concerning information value in a symbolic representation like BIM can be summarized as follows:

- Preservation and completion of primary data
- Establishing transparent and efficient procedures for producing derivative data when needed
- Identification and interpretation of anti-data, including specification of relevant actions
- Preservation of stable operational and metadata

The priority of primary data apparently conflicts with IM improvement of information quality through condensation, i.e. operations that return pragmatically superior derivative data and metadata. Such operations belong to the second point above: if the primary data serve as input for certain procedures, then these procedures have to be established as a dynamic view or similar output in BIM. If users need to know the floor areas of spaces, one should not just give them the space dimensions and let them work out the calculations but supply instead transparent calculations, ordered and clustered in a meaningful way. This does not mean that the results of these calculations should be preserved next to the space dimensions from which they derive.

Moving from the semantic to the pragmatic level, *veracity* is a key criterion of quality: fitness for purpose obviously requires that the information is true. In addition to user feedback, veracity can be established on the basis of additional data, e.g. laser scanning to verify that a model represents faithfully, accurately and precisely the geometry of a particular building.

Before relevance or veracity, however, one should evaluate the structural characteristics of primary information: a model that is not complete, coherent and consistent is a poor basis for any use. *Completeness* in a building representation means that all parts and aspects are present, i.e. that there are no missing symbols for building elements or spaces in a model. BIM software uses *deficiency detection* to identify missing symbols. Missing aspects refer to symbol properties or relations: the definition of symbols should include all that is necessary to describe their structure, composition, behaviour and performance.

Completeness is about the presence of all puzzle pieces; *coherence* is about how well these pieces fit together to produce a seamless overall picture. In a building representation this primarily concerns the interfacing of elements, including possible conflicts in space or time. *Clash detection* in BIM aims at identifying such conflicts, particularly in space. Relations between symbols are of obvious significance for coherence, so these should be made explicit and manageable. In BIM, there are examples of this in the way some symbols attach themselves to others, e.g. co-terminating walls to each other, spaces to their bounding walls and floors, windows and doors to hosting walls. Parameterization can extend such relations further into a network that automatically ensures coherence.

Finally, *consistency* is about all parts and aspects being represented in the same or compatible ways. In a symbolic representation, this refers to the properties and relations of symbols. If these are described in the same units and present in all relevant symbol types, then consistency is also guaranteed in information use. Colour, for example, should be a property of the outer layer of all building elements. In all cases, the colour should be derived from the materials of this layer. This means that any paint applied to an element should be explicit as material with properties that include colour. Moreover, any colour data attached to this material layer should follow a standard like the RAL or Pantone colour matching systems.

Allowing users to enter any textual description of colour does not promote consistency.

It is important to evaluate completeness, coherence and consistency only after clarifying the semantic types in a representation. This allows one to concentrate on the data that really matter, in particular primary and anti-data, and the procedures that produce derivative data. This allows higher focus in IM and reduces the amount of data to be processed.

Key Takeaways

- *IM is more than a technical necessity: it is also a means of improving performance in a project or enterprise and therefore a key component of overall management.*
- *IM is inclusive and accepts all kinds of information, from structured, semi-structured and unstructured sources. As a structured information system, BIM simplifies IM.*
- *IM has two main goals: regulate information flow and safeguard or improve information quality.*
- *Custodianship of information is critical for information control.*
- *Information flow relates to the syntagmatic dimension of a representation and draws from the sequence of tasks in a process, as well as from extrinsic constraints.*
- *In managing information flow one needs to make explicit what, who, how and when.*
- *The I-P-O scheme helps translate tasks into information actions.*
- *Even before a design takes shape, there are substantial amounts of information that should be made explicit in a model as feedforward.*
- *Information quality concerns the paradigmatic dimension and can therefore build on the semantic typology of data.*
- *In addition to semantic and pragmatic criteria, information quality also depends on completeness, coherence and consistency.*

Exercises

1. Use the I-P-O scheme to explain how one decides on the width of an internal door in a design. Cluster the input by origin (general, specific, project) and describe the relations between input items.
2. Use the I-P-O scheme to explain what, who, how and when in deciding the layout of an office landscape, particularly:
    1. Which workstation types are to be included, including dimensions and other requirements.
    2. How instances of these types are to be arranged to achieve maximum capacity.
3. In a BIM editor of your choice make the permissible building envelope for a building in a location of your choice. Describe the process in terms of input, information instances produced and resulting constraints for various kinds of symbols in the model.
4. Evaluate the completeness, coherence and consistency of the permissible building envelope model you have made.
5. Analyse how one should constrain types of building elements in relation to performance expectations from the use type of building: compare a hotel bedroom to a hospital ward on the basis of a building code of your choice. Explain which symbol properties are involved and how.

Notes

---

1. The views on DM derive primarily from: Richards, M., 2010. *Building Information Management - a standard framework and guide to BS 1192*. London: BSI; Eynon, J., 2013. *The design manager's handbook*. Southern Gate, Chichester, West Sussex, UK: CIOB, John Wiley & Sons; Emmitt, S., 2014. *Design management for architects* (2nd ed.). Hoboken NJ: Wiley·
2. The presentation of IM is based on: Bytheway, A., 2014. *Investing in information*. New York: Springer; Detlor, B., 2010. Information management. *International Journal of Information Management, 30*(2), 103-108. doi:10.1016/j.ijinfomgt.2009.12.001; Flett, A., 2011. Information management possible?:Why is information management so difficult? *Business Information Review, 28*(2), 92-100. doi:10.1177/0266382111411066; Rosenfeld, L., Morville, P., & Arango, J., 2015. *Information architecture :for the web and beyond* (4th ed.). Sebastopol CA: O'Reilly Media.

3. IM definitions of information quality derive from: Wang, R.Y., & Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33. doi:10.1080/07421222.1996.11518099; English, L.P., 1999. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits.* New York: Wiley.

7. Process and information

This chapter describes how process and information diagrams can describe tasks in a process and the information actions relating to these tasks in a comprehensive and coherent manner.

Flow charts

As we have seen in the previous chapter, there is some correspondence between the sequence of tasks in a process and the sequence of information actions: process management and IM overlap. The main difference is that IM goes beyond the actions and transactions in a task, in order to identify, structure and connect information in a way that supports and anticipates the needs of the process. Therefore, the first step towards effective IM in any process is understanding the process itself: what people actually do and how their actions, decisions, interactions and transactions relate to the production, dissemination and utilization of information. Starting IM by analysing the process also has advantages for the deployment of IM measures: most people and organizations are more process-oriented than information-oriented and may have difficulty identifying and organizing information actions without a clear operational context. Using a process model as background makes clearer why and how one should manage information.

A process can be described as a sequence of tasks towards a specific outcome. Representing processes diagrammatically is particularly useful in our case because of the abstraction and consistency afforded by diagrams. Of the many kinds of diagrams available for this purpose, basic flow charts suffice in practically all cases. These diagrams are directed graphs, in which objects are represented by nodes of various kinds corresponding to different kinds of objects, while relations are described by arcs (Figure 1). The direction of the arcs indicates the direction of flow in the process. Bidirectional arcs should be avoided because they usually obscure separate tasks, e.g. evaluation followed by feedback. Explicit representation of such tasks is essential for both the process and IM.

To make an unambiguous and useful flow chart of a process, one should adhere to another basic rule: that each object should appear only once as a node in the diagram. This allows us to make feedback explicit and to measure the degree of a node, its closeness and all other graph-theoretic measures that can be used in analysing the process.

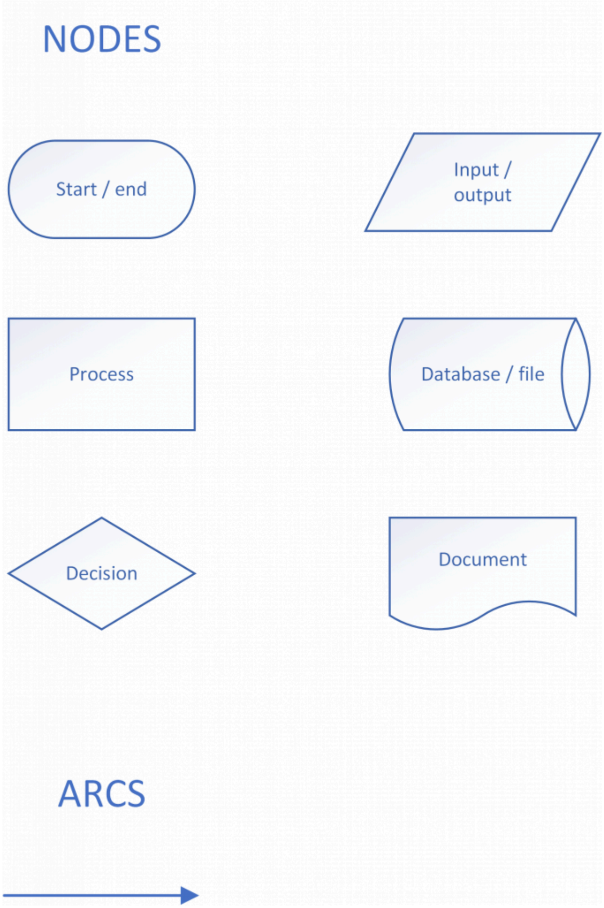


Figure 1. Nodes and arcs in a flow chart

Let us consider a simple example of a process in building design: the estimation of construction cost in early design, on the basis of gross floor area. The process involves three actors: the client, the architect and the cost specialist. These are responsible for the budget, the design, the cost estimation and the evaluation of the estimate, which leads to either feedback to the design (usually to lower the cost) or acceptance of the design as it is (Figure 2).

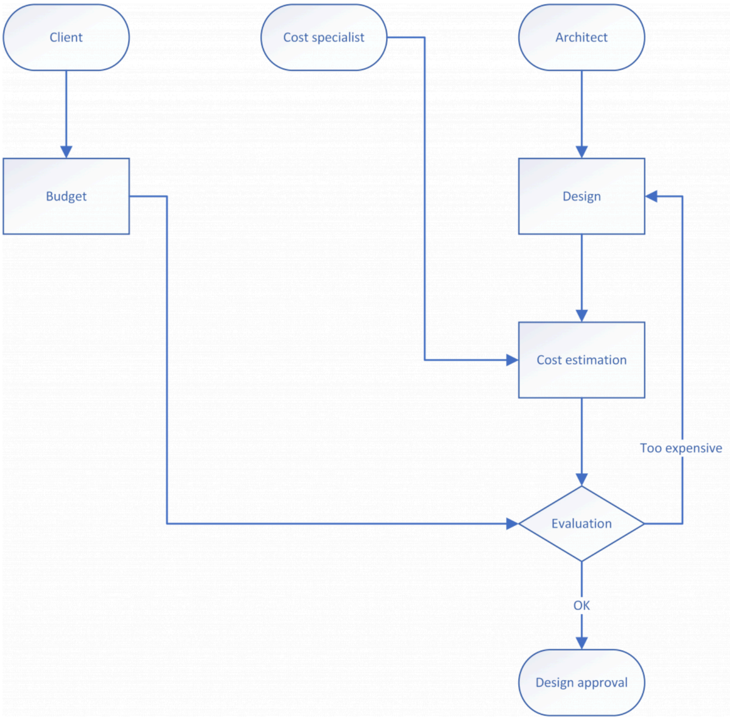


Figure 2. Cost estimation process diagram

Information diagram

The process diagram is clear about who does what but the actual information they produce and consume remains obscure. This generic depiction may be useful for process management but is too abstract for IM. Using the process diagram as a foundation, one can develop an information diagram that makes information and its flow explicit (Figure 3). Actor nodes may be abstracted from an information diagram in order to focus on the analysis of process-related and data-related nodes into information instances. Ultimately, however, in IM one should always identify who does what in unambiguous terms.

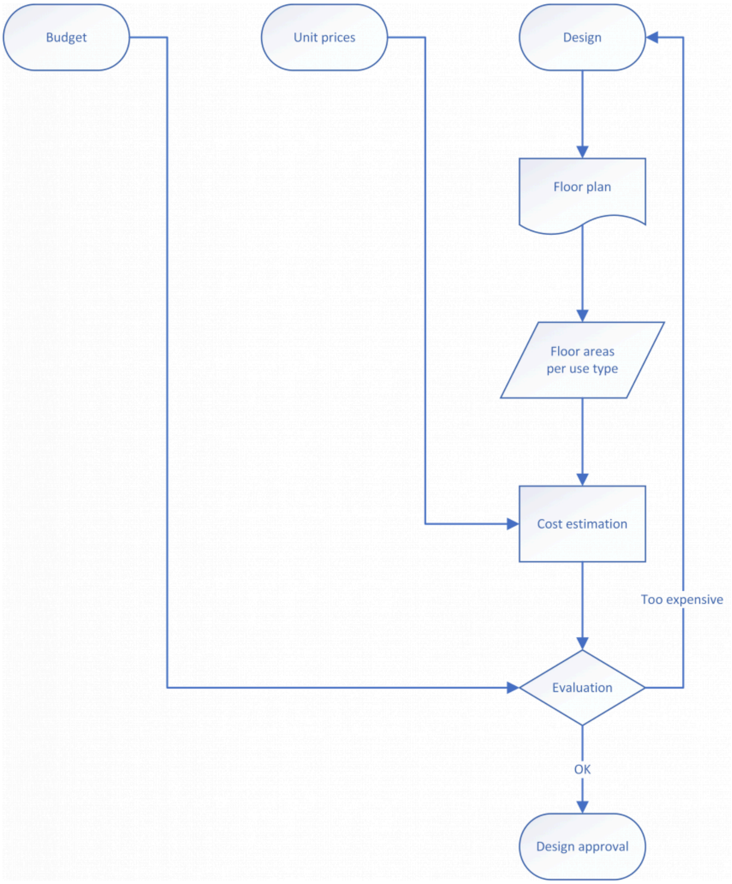


Figure 3. Cost estimation information diagram

In this diagram, too, each object should appear only once, as a single node. So, if the floor plan has to be redrawn because the design is deemed too expensive, the diagram should contain feedback loops to the floor plan of the same (albeit modified) design, so as to make the process cycles explicit. If the cost evaluation leads to a radically new design, requiring a new node in the diagram, then this should be made clear by means of unambiguous node labelling (e.g. *Design 1* and *Design 2*). Such new versions of the same nodes should be used cautiously and sparingly, only when absolutely necessary, e.g. when a process involves design alternatives.

The information diagram should reveal what actually takes place in terms of input, processing and output in the process. For example, a cost specialist contributes to the cost estimation by supplying a list of unit prices, i.e. cost of gross floor area per m^2 for different categories of building use. One m^2 of storage area costs significantly less than one m^2 of office space, which in turn costs less than one m^2 of an operating theatre in a hospital. This also means that one has to extract matching data from the design. It is not enough to calculate the total gross floor area of a hospital design; one has to know the use of every space, so as to be able to calculate the subtotals for each category. The subtotals are then multiplied by the unit prices to arrive at a correct estimate and ascertain which category may be too big or too costly.

The above illustrates an important difference between process and information diagrams: the former can be abstract about what each task entails but the latter has to be specific regarding information sources (e.g. which drawings are used), the information instances these sources accommodate and the actions through which these instances are processed. The higher specificity of the information diagram leads to a finer grain in the analysis of the process into nodes and arcs that allow one to trace the flow of information instances. In general, it may be assumed that the flow is the same in both diagrams but the finer grain of the information diagram may lead to new insights and local elaborations or changes.

I-P-O and primary versus derivative

Transforming a process diagram into an information diagram involves the I-P-O scheme. Examining each node in the process diagram with respect to this scheme reveals which information is used as input and produced as output. The *Design* node, for example, is expected to contribute to a cost estimation involving gross floor areas. This means that the design cannot exist solely in the architect's mind; we need some external representation as input, on the basis of which we can measure floor areas, moreover by use category. The obvious candidate is a floor plan and, more precisely, one where all spaces are indicated and labelled by their use. This floor plan rather than some abstract notion of a design is the appropriate input for the processing we require (calculation of gross floor areas). In the same manner, one can establish that these areas are the local output of a task: what has to be passed on to the next processing step (cost estimation) as input.

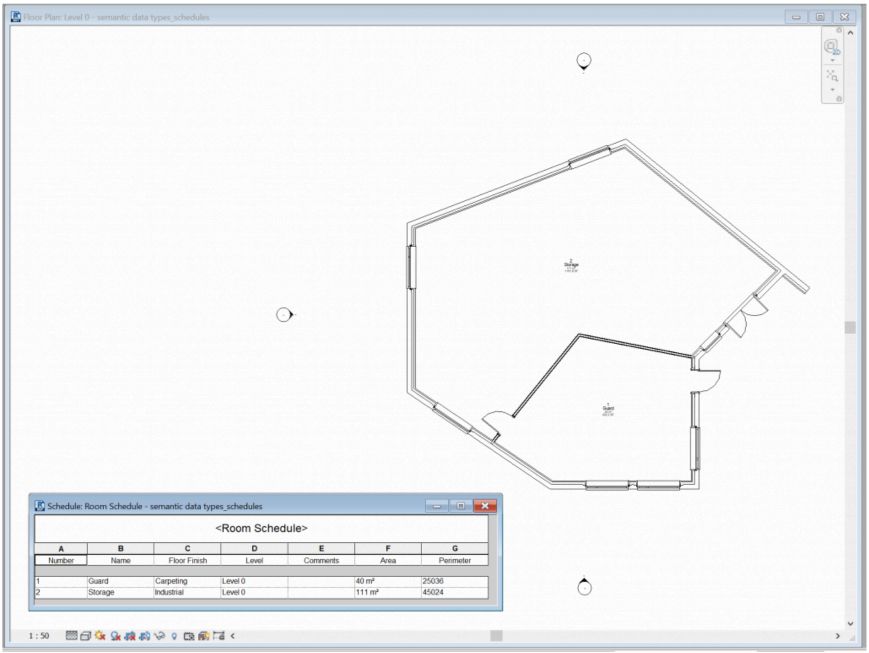
Equally important for the development of a complete and specific information diagram is the semantic type of information used as input: if it is derivative, one has to trace it back to the primary data from which it is produced. Floor areas are derivative, so one needs to identify the primary data from which they derive, as well as the representations that accommodate these primary data. Consequently, one should not just require a table of all spaces, their areas and use type from the design but also specify that for making this table one needs floor plans that describe the spaces and their uses. These floor plans should be present as sources in the information diagram.

Information diagrams for BIM

As explained in previous chapters, implementation mechanisms may affect the structure and use of a representation in non-trivial ways. Consequently, one should take implementation environments into account in an information diagram, including adapting the diagram following any change in implementation environment: what applies to analogue information processes may be significantly different to what should take place in a computer.

The information diagram we have considered so far for cost estimation could be called generic, although it primarily reflects analogue practices and media. Adapting it to BIM means first of all that the model (the central information system) should be explicitly present as a source. This information system contains the symbols and relations in which primary data are found. Derivative data like floor area calculations are produced from the model in views like schedules. These schedules are typically predefined in various formats, including room schedules that list spaces and their properties, including floor area calculations (Figure 4).

They can be used to verify that the model contains all the primary data needed for the cost estimation.

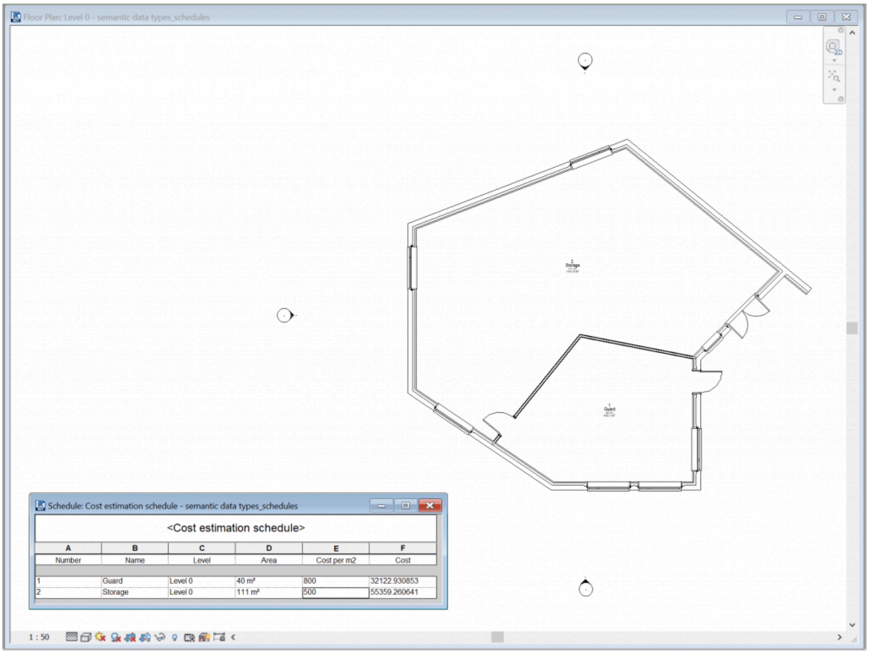


<Room Schedule>

A	B	C	D	E	F	G
Number	Name	Floor Finish	Level	Comments	Area	Perimeter
1	Guard	Carpeting	Level 0		40 m²	25036
2	Storage	Industrial	Level 0		111 m²	45024

Figure 4. Room schedule in a BIM editor

Unit prices can be added to room schedules, thus integrating cost estimation in BIM in a straightforward and transparent manner (Figure 5).



<Cost estimation schedule>

A	B	C	D	E	F
Number	Name	Level	Area	Cost per m2	Cost
1	Guard	Level 0	40 m²	800	32122.930853
2	Storage	Level 0	111 m²	500	55359.260641

Figure 5. Room schedule with integrated cost estimation

Integrating cost estimation in BIM also means that feedback to the model should be similarly direct and straightforward, e.g. through annotations to spaces, especially large or expensive ones that should be prioritized when improving the design to match the constraints of the budget (Figure 6). Note that feedback in this example is abstract with regard to the particular symbols or relations that are affected. One may choose to be quite specific about this, so as to guide information actions with precision and certainty. Reversely, if one chooses abstract actions, then this would involve interpretation by certain actors. These actors should therefore be explicit in the diagram as authors or custodians of specific information .

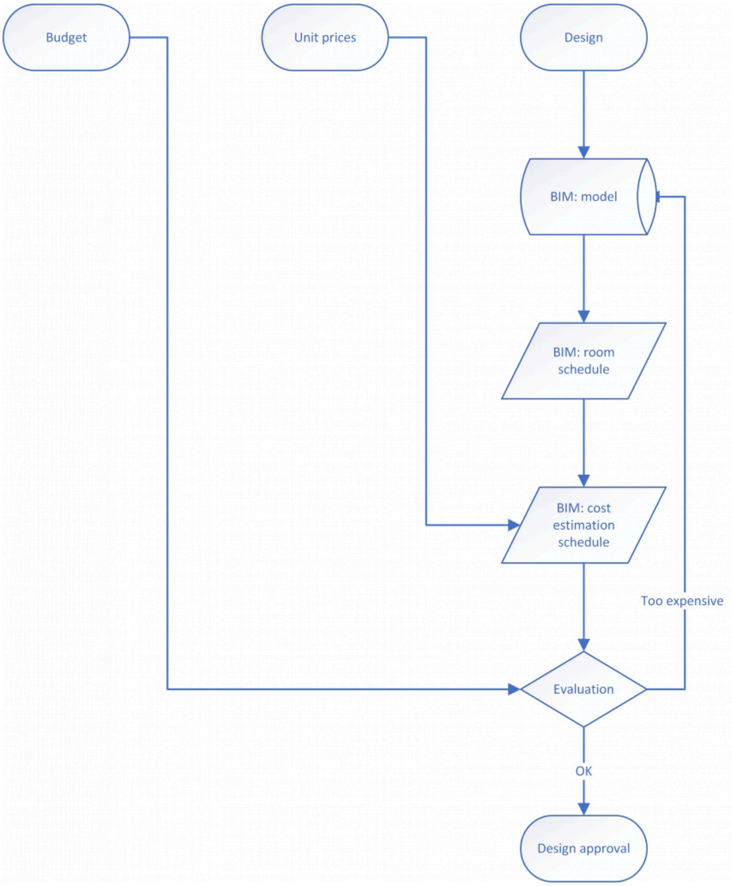


Figure 6. Information diagram for cost estimation in BIM

An integrated information environment like BIM also makes automation of various information controls possible, e.g. concerning the presence of essential primary data. These too can be included in an information diagram for BIM (Figure 7).

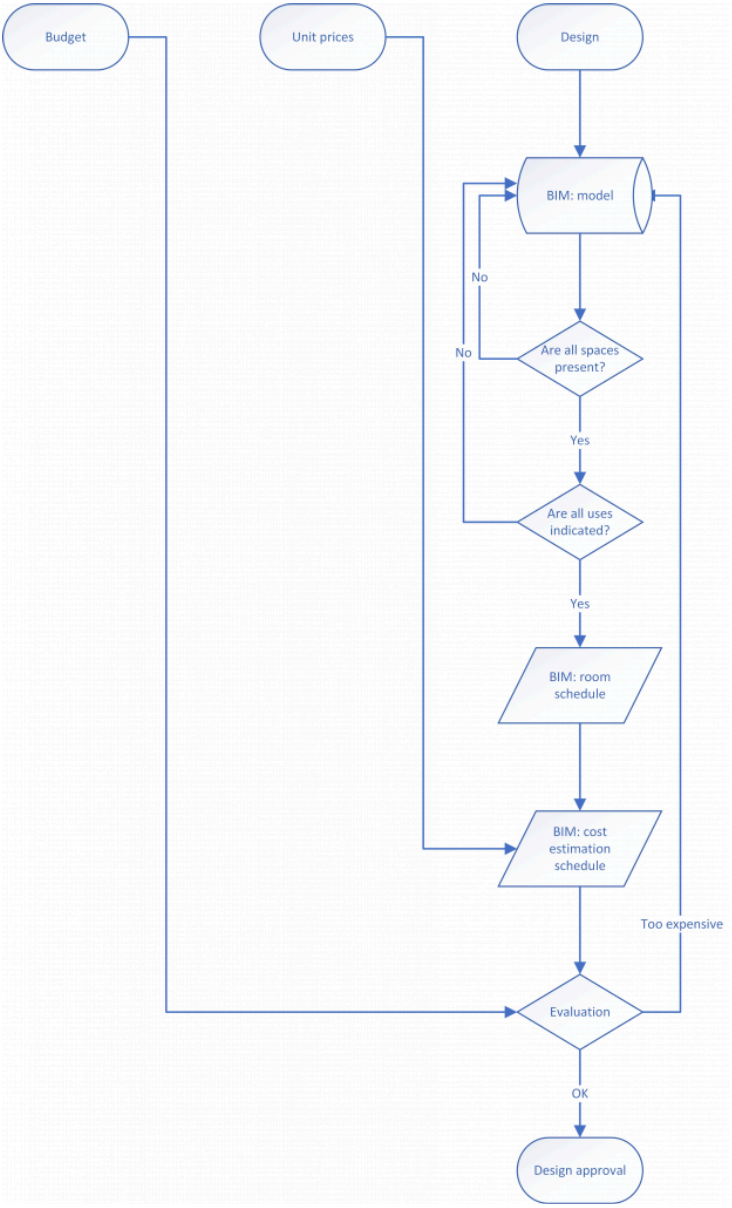


Figure 7. Information diagram for cost estimation in BIM including quality controls

Using information diagrams in information management

An information diagram that captures both the needs of a process and the capacities of BIM can make IM clear and unambiguous to both managers and actors in the process. Information flow can be explicitly depicted in the diagram, especially concerning what, who and when. Managers can use the information diagram to guide and control the process at any moment, while actors have a clear picture of the scope and significance of their actions. Addressing *how* questions depends on the fineness of the grain in the description of information instances: the finer it is, the more specific answers one can draw from the diagram. As such specificity affects interpretation, one should be careful about the balance between the two: many actors in a building project are knowledgeable professionals who may not take kindly to IM approaches that overconstrain them.

On the other hand, IM has to be strict about matters of authorship and custodianship because not everybody is yet accustomed to the possibilities and responsibilities of digital information processing. By linking actors to information with accordingly labelled arcs in the information diagram, one can indicate responsibilities and actions throughout a process. Note that roles can be variable: an actor who authors some information in one task may become custodian of other information in another task.

Concerning information quality, the information diagram forms a usable background for pragmatic value: applying the I-P-O scheme at any node is a critical part of measuring pragmatic value, i.e. establishing what users need to process and must produce in a task. Similarly, the information diagram is essential for the evaluation of completeness, coherence and consistency: it reveals the moments when one should return to the representation and analyse it. Such moments occur after critical information instances or after a multitude of information instances, i.e. when the model changes substantially.

The information diagram is also a necessity for our parsimonious approach to information value. This approach focuses on primary data and their propagation; both can be traced with accuracy in the diagram, including explicit, manageable connections to derivative data, enabling managers and users to know what should be preserved or prioritized. Finally, in the same manner one can identify anti-data, on the basis of expectations (e.g. knowing when information from different disciplines comes together in a process) and interpretation (e.g. that a space without a door is a shaft). This leads to directed action (e.g. requiring that two disciplines work together to solve interfacing problems), which should be present in an information diagram of appropriate specificity.

Key Takeaways

- *Flow charts are directed graphs that can be used to describe a sequence of tasks (process diagram) or a sequence of information actions (information diagram).*
- *In both process and information diagrams, each object should be represented by only one node and each arc should be unidirectional.*
- *The I-P-O scheme helps translate a process diagram into an information diagram.*
- *Tracking the primary data needed for a process makes the information diagram complete and specific.*
- *Information diagrams should take into account the implementation environment of BIM: the symbols and relations that contain the primary data and the views that present derivative data, as well as the possibilities for quality control.*
- *Information diagrams make flows explicit and manageable; they also support analyses of information quality by identifying significant actions and moments.*

Exercises

1. Measure the degree and eccentricity of nodes, and the eccentricity, diameter and radius of the graph in the process diagram of Figure 2. Do these measurements suggest critical parts in the process?
2. Measure the same in the information diagram of Figure 3. Do you observe any differences with Figure 2, also in terms of critical parts?
3. Add symbols, properties and relations to the information diagram of Figure 7. Does the increased specificity make IM easier or more reliable?
4. Add actors to the information diagram of Figure 7. How does the result compare to the diagram of the previous exercise?

PART IV

EXERCISES

Key concepts

The following list of key concepts from the previous chapters is a reminder or checklist of what can be used in solving information problems, e.g. in the following exercises.

- Symbolic representation
    - Symbols, properties and relations
    - Symbols and things
    - Symbols versus implementation mechanisms

- Graphs: objects and relations represented respectively by vertices (nodes) and edges
- Directed graphs (digraphs): graphs consisting of nodes and arcs (edges with a direction)

- Abstraction: visual versus mnemonic
- Solids and voids in building representations
- Paradigmatic and syntagmatic dimensions

- Data and information instances
- Semantic data types: primary, anti-data, derivative, operational, metadata
- Information instances by scope: single symbol versus multiple symbols

- Structured, semi-structured and unstructured information sources
- Information flow: what, who, how, when

- Completeness, coherence and consistency
- Information authorship versus information custodianship
- Process diagram: sequence of tasks in a digraph representation
- Information diagram: information instances and flows
- I-P-O: transition from process to information management

Exercise I: maintenance

Organize the process of repainting all walls of a large lecture hall at a university. The walls are in good condition, so a single coat of pain suffices. The process therefore can be reduced to the following tasks:

- Make a model of the lecture hall in BIM using direct measurements and photographs
- Classify wall surfaces and their parts with respect to:
    - Labour (e.g. painting parts narrower than 30 cm are more time consuming)
    - Equipment (e.g. parts higher than 220 cm require scaffolding)
    - Accessibility (e.g. parts behind radiators or other fixed obstacles are hard to reach and therefore also time consuming)
- Measure the wall surfaces
- Make cost estimates
- Make a time schedule in 4D BIM

Deliverables

1. Process and information diagrams, accompanied by short explanatory comments
2. Basic model of the lecture hall in a BIM editor
3. Schedules for classification, measurement, estimates and scheduling in BIM

Roles

If the exercise is a group assignment, consider roles for the following aspects:

- Process management
- Information management
- BIM modelling (two or more people)

- Analyses in BIM (using schedules – two or more people)

Exercise II: change management

Organize how changes to a design in the development and realization stages can be registered and processed in BIM. These changes may refer to:

- Change to a property of a symbol (e.g. lengthening of a wall)
- Change of the type of a symbol (e.g. change of family for a door)
- Change in a relation between symbols (e.g. relocation of a door in a wall)
- Change in a time property of a symbol (e.g. as a result of a scheduling change)

Organize the process of change management in both stages as a series of tasks that reflect the above types of changes and take into account possible causes of change, such as:

- Changes in the brief (e.g. new activities added)
- Changes in the budget (e.g. increase of façade cost necessitating reduction of cost elsewhere)
- Changes in an aspect of the design (e.g. change in the heating solution or the fire rating of internal doors and ensuing interfacing issues – not just clash detection)
- Changes in the construction schedules (e.g. due to delays in the delivery of components or to bad weather)
- Errors in construction (e.g. wrong dimensioning or specifications of an element)

Deliverables

1. Process and information diagrams, accompanied by short explanatory comments
2. Basic model in a BIM editor demonstrating the way changes can be implemented
3. Short overview of findings (two A4 sheets)

Roles

If the exercise is a group assignment, consider roles for the following aspects:

- Process management
- Information management
- BIM modelling
- Case analyses (for finding realistic examples)

Exercise III: circular energy transition

The planned energy transition in the Netherlands means that most buildings have to undergo an expensive renovation to meet new standards. To reduce costs, one can adopt a circular approach to both components or materials released from existing buildings and the new components and subsystems that will be added to the buildings. Organize the following tasks for a typical Dutch single-family house:

- Document the existing situation in a model appropriate for renovation, i.e. including realization phases, distinction between existing and planned, what should remain and what should be removed
- Identify in the model components and materials that should be extracted (e.g. radiators: the house will have underfloor heating), explaining how identification takes place (preferably automatically) in the model
- Estimate the expected circularity form for these components and materials (recycle, remanufacture, repurpose, re-use etc.), explaining which factors play a role (weathering, wear, interfacing with other elements etc.) and how these factors can be detected in the model
- Identify which elements should be upgraded and specify what this entails in the model (paying attention to phasing and element type changes)
- Specify how new elements (for the renovation) should be added to the model to support the above in the remaining lifecycle of the house
- Make a time schedule for the renovation in 4D BIM

Deliverables

1. Process and information diagrams, accompanied by short explanatory comments
2. Incomplete model in a BIM editor containing demonstrations of your solutions
3. Schedules for circularity analyses in BIM
4. Short overview and table of contents (two A4 sheets)

Roles

If the exercise is a group assignment, consider roles for the following aspects:

- Process management
- Information management
- BIM modelling
- Analyses in BIM (using schedules – probably more than one group member)
- Legal and technical aspects of the energy transition
- Building documentation (emphasis on how to deal with incompleteness and uncertainty)
- Subsystem integration
- Circularity in design (technical aspects)

References

Attneave, F., 1959. *Applications of information theory to psychology; a summary of basic concepts, methods, and results.* New York: Holt.

Bytheway, A., 2014. *Investing in information*. New York: Springer.

Cosgrove, D., 2003. Ptolemy and Vitruvius: spatial representation in the sixteenth-century texts and commentaries. A. Picon & A. Ponte (eds) *Architecture and the sciences: exchanging metaphors*. Princeton NJ: Princeton University Press.

Cover, T.M., & Thomas, J.A., 2006. *Elements of information theory* (2nd ed.). Hoboken NJ: Wiley-Interscience.

Detlor, B., 2010. Information management. *International Journal of Information Management, 30*(2), 103-108. doi:10.1016/j.ijinfomgt.2009.12.001

Eastman, C., Teicholz, P.M., Sacks, R., & Lee, G., 2018. *BIM handbook* (3rd ed.). Hoboken NJ: Wiley.

Emmitt, S., 2014. *Design management for architects* (2nd ed.). Hoboken NJ: Wiley.

English, L.P., 1999. *Improving data warehouse and business information quality: methods for reducing costs and increasing profits.* New York: Wiley.

Evans, R., 1995. *The Projective Cast: Architecture and Its Three Geometries.* Cambridge MA: MIT Press.

Eynon, J., 2013. *The design manager's handbook*. Southern Gate, Chichester, West Sussex, UK: CIOB, John Wiley & Sons.

Flett, A., 2011. Information management possible?:Why is information management so difficult? *Business Information Review, 28*(2), 92-100. doi:10.1177/0266382111411066

Floridi, L., 2008. Trends in the philosophy of information. P. Adriaans & J. v. Benthem (eds), *Philosophy of information*. Amsterdam: North-Holland.

Floridi, L., 2009. Philosophical conceptions of information. G. Sommaruga (ed), *Formal Theories of Information: From Shannon to semantic information theory and general concepts of information*. Berlin, Heidelberg: Springer.

Floridi, L., 2016. Semantic conceptions of information. *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2016/entries/information-semantic/

Gantz, J. & Reinsel, D., 2011, "Extracting value from chaos." 2011, https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf

Goodman, N., 1976. *Languages of art; an approach to a theory of symbols* (2nd ed.). Indianapolis IN: Hackett.

Kanizsa, G., 1979. *Organization in vision: essays on Gestalt perception.* New York: Praeger.

Lyman, P. & Varian, H.P. 2003, "How much information." http://groups.ischool.berkeley.edu/archive/how-much-info/

Pierce, J.R., 1980. *An introduction to information theory : symbols, signals & noise* (2nd, rev. ed.). New York: Dover.

Richards, M., 2010. *Building Information Management – a standard framework and guide to BS 1192*. London: BSI.

Rosenfeld, L., Morville, P., & Arango, J., 2015. *Information architecture :for the web and beyond* (4th ed.). Sebastopol CA: O'Reilly Media.

Shannon, C., 1948. A mathematical theory of communication. *Bell System Technical Journal, 27*(July, October), 379-423, 623-656.

Shannon, C.E., & Weaver, W., 1998. *The mathematical theory of communication*. Urbana IL: University of Illinois Press.

Simonite, T., 2016. "Moore's law Is dead. Now what?" Technology Review https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/

Sommaruga, G., 2009. Introduction. G. Sommaruga (ed), *Formal Theories of Information: From Shannon to semantic information theory and general concepts of information*. Berlin, Heidelberg: Springer.

Steadman, P., 1983. *Architectural morphology: an introduction to the geometry of building plans.* London: Pion.

Toffler, A., 1970. *Future shock*. New York: Random House.

Turner, V., Reinsel D., Gantz J. F., & Minton S., 2014. "The Digital Universe of Opportunities" https://www.emc.com/leadership/digital-universe/2014iview/digital-universe-of-opportunities-vernon-turner.htm

Van Sommers, P., 1984. *Drawing and cognition: descriptive and experimental studies of graphic production processes.* Cambridge: Cambridge University Press.

Wang, R.Y., & Strong, D.M., 1996. Beyond accuracy: what data quality means to data consumers. *Journal of Management Information Systems, 12*(4), 5-33. doi:10.1080/07421222.1996.11518099

Waltz, D., 1975. Understanding line drawings of scenes with shadows. P.H. Winston (ed) *The psychology of computer vision.* New York: McGraw-Hill.

Summary and Author Biography

The book presents a coherent theory of building information, focusing on its representation and management in the digital era. It addresses issues such as the information explosion and the structure of analogue building representations to propose a parsimonious approach to the deployment and utilization of symbolic digital technologies like BIM.

Alexander Koutamanis has studied architecture at Aristotle University of Thessaloniki, Greece, and received his PhD from Delft University of Technology, the Netherlands, where he is currently associate professor of computational design.

Building Information - Representation and Management: Fundamentals and Principles

Alexander Koutamanis

The book presents a coherent theory of building information, focusing on its representation and management in the digital era. It addresses issues such as the information explosion and the structure of analogue building representations to propose a parsimonious approach to the deployment and utilization of symbolic digital technologies like BIM.

Dr.ir. Alexander Koutamanis
TU Delft | Architecture

Alexander Koutamanis has studied architecture at Aristotle University of Thessaloniki, Greece, and received his PhD from Delft University of Technology, the Netherlands, where he is currently associate professor of computational design.

© 2019 TU Delft Open
ISBN 978-94-6366-160-7
DOI: https://doi.org/10.5074/T.2019.003

textbooks.open.tudelft.nl