Digital Scholarship in the Humanities, 2026, **00**, 1–14
https://doi.org/10.1093/llc/fqag011
Full Paper

OXFORD

Digital objects as translation: their digitization depth as a taxonomy for digital analogues

Giovanni Maria Pala[1,*]

[1]Faculty of History, Stephen A. Schwarzman Centre for the Humanities, University of Oxford, Radcliffe Observatory Quarter, Oxford, OX2 6GG, United Kingdom
*Corresponding author. Faculty of History, Stephen A. Schwarzman Centre for the Humanities, University of Oxford, Radcliffe Observatory Quarter, Oxford, OX2 6GG, United Kingdom. E-mail: giovanni.pala@history.ox.ac.uk

Abstract

Digital collections of historical objects are increasingly the first interface when setting up new academic enquiries in the Humanities. This influences how research questions are formed. The size of these collections is posited to grow even more as the costs of digitization decrease and access widens. Framing digital collections and their contents correctly is therefore of great importance. However, in many cases the conceptualization of digitized collections remains in practice tied to the mistaken notion that digital copies are surrogates for the analogue originals they represent, with serious implications for how these items are organized. This article builds from earlier seminal contributions to argue that digitized objects are a different entity altogether, neither copy nor surrogate. This difference can be captured more faithfully by the notion of 'digital translation': digital objects are a selective description of the physical original; this description is translated into digital language and written in symbols bound by rules. This structured information can be read algorithmically. From this definition, we can derive a taxonomy based on the depth of digitization, that is, how many layers of description exist about the object. These layers often include shape, colour, and recorded metadata. Adopting a by-depth taxonomy has ramifications for existing digital collections: if digital objects are individualized translations of an original, their digital collections are collections of commentaries in digital language reflecting specific ways of seeing the original. If digitization depth differentiates these sources, our archival and dissemination strategies must respond to this reality.

Keywords digital taxonomy; digital copy; digital surrogate; material collections; digital translation.

1. Introduction

Suppose a very advanced robot is designed with the capacity to precisely replicate the mobility of the human hand. The robot is programmed to hold a brush and can paint by executing a series of pre-set commands. Information on a painting is obtained from an advanced scan and chemical analysis run to the highest standard possible.[1] The robotic hand follows complex digital instructions derived from this data to paint with substantially the same progression and pressure of the original artist, using historically accurate materials. At the same time, an advanced screen is engineered: it displays in vivid detail the recorded information of the painting, including its texture, so that the recorded object is perceived to be almost like the original on the surface of the screen.[2] A close look at the screen, however, reveals enough differences to appreciate a distance between the reproduction and the original.

When used, the two devices produce two different outputs. The screen output is a quasi-equivalent object: an image on the screen, close to the original but lacking many qualities of the painting, acting in place of it. It is a surrogate of the original. The result of the other device, i.e. the robotic painter, is of such quality, flawless execution, and analogic reproduction, as to be substantially indistinguishable from the original, an almost perfect equivalent. It is a copy. However, neither the copy nor the surrogate is the digital object obtained from the original via the scanning process. At inspection, the digital object is composed of a series of symbolic descriptors, which can be used by either replication device after they are decoded into instructions. The digital object is not the screened image or the replica painting, it is rather the symbolic representation of the original, a collection of descriptors that are selectively decoded to produce the two outputs. Because they are produced *from* a digital object, the copy and surrogate are a product of it (cf. Conway 2015).[3]

This example illustrates concisely why the terms digital copy and digital surrogate, popular in much of the literature on digital objects, can fall short of adhering to the reality of digitization. It is also the sketch of an argument on the necessity of going deeper in our terminology. There are, however, good reasons for the current terminological popularity of copy and surrogate.

Firstly, when it comes to outputs (the screened image or the reproduced painting) using labels such as 'copy' or 'surrogate' captures the apparent unity between digital object and its enactment as reproduction of the original (Deegan 2006). This unity is, from the point of view of most users, deceptively self-evident. Digital materials create the illusion of equivalence but are far removed as copies from the original than any analogue copy. This issue is not lost on scholars engaged with digital collections. Definitions of digital surrogate, for example, tend to focus on the notion of 'being used in place of the original', but all seem simply to stem from the recognition that the digital object one engages with is not quite the copy it posits to be (Grycz 2006; Kropf 2016). While this is certainly true, it often falls short of recognizing the reason that makes a digital object something other than the original or a good copy: it is written in another language. Modern digital language is hidden by successful graphic interfaces, designed with ease of use in mind and to hide the workings of digital systems from our gaze: when users experience digital interfaces the relationship between digital and experienced phenomena is intentionally presented as a close one.[4]

Secondly, digital replicability is part of a long history of tools used to reproduce objects and forms, of which mechanical replicability is the immediate precedent. Because of the apparent identity between the digital and the experience we make of it, much of the existing scholarship has inherited ideas and commentaries developed for mechanical copies and applied them to the digital world. For this reason, the discussion can be traced back to an eminent twentieth century tradition prompted by the birth of photography, filming, and the rise of mass production and consumption of replicas and representations. Walter Benjamin's 'aura of the original' (Benjamin 1969; cf. Burns 2017 on surrogates and aura), Susan Sontag's essays on the politics of photography (Sontag 1977), or Jean Baudrillard's 'simulacra' (Baudrillard 1994), to name three highly influential authors from different backgrounds and periods, are all concerned with the struggle to conceptualize what is original and unique once mass production as mass reproducibility affects the nature of material objects. The three authors see in modern copiability a radical shift in attitudes and political engagement of the viewer, now turned recorder and passive receiver, and at the same time tend to construct their own critique of mass reproducibility as one where copies have the power to hollow symbols and meaning, bringing the object so close to the viewer that its cult value, its mystique, is lost.

Digitization seems to possess the same power of mechanical reproducibility: it displaces space and time by making available in the here and now details of otherwise difficult to access sources. Because digitization enables us to reproduce at low costs the same digital object for potentially an unlimited number of times, it is too easily framed as an act of copying in the sense adumbrated by Benjamin and his successors. If we follow this line of argument, digital objects are copies and relate to the authority of the original, or the symbolic meaning it is imbued with, in the same way mechanical static copies do: as a threat to their integrity for some, or a deformed reflection for others, although the act of copying is not always a source of anguish in the literature (Boon 2010: 1–40). Nevertheless, be it as a defence of the original against its copies or a critique of this setup, the entire debate rests on assuming that copying is the most important feature of digitizing. If we remove the digital object from the sphere of influence of copying and surrogating, the light the existing discourse provides on the matter is dimmer.

In what follows I will formally recall why the copy-surrogate duality is an illusion produced by our fruition of digital objects and why the notion of encoded translation is a conceptual alternative that is more faithful to their epistemological nature. This will be done without engaging too much with the debate on copies and their relationship with the original, even though our case has strong resonance with the debate on autographic works, i.e. works where the representation and content of one object are protected, and the single object matters more than its copy; and allographic works, i.e. works where the content is protected but not its representation (Goodman 1976; Zeimbekis 2012; D'Cruz and Magnus 2014). Having discussed digital translations, I will then introduce the reader to a new taxonomy based on three ways of seeing the digital source, and to a preliminary discussion of some of the consequences of this notion for our collections and research.

Importantly, this contribution joins ongoing reflections on the nature of digitization (cf. Hui 2016) and its relationship to the materiality and interpretability of digital objects (see, for example, Drucker 2011 on digitization as a non-neutral act). More recently, the *Collections as Data* movement (Padilla et al. 2019) has advocated for workflows that make archival materials computationally amenable. This contribution implicitly supports such efforts by offering a conceptual framing of digitized materials.

2. Digital acquisition is a form of description

The starting point of our argument is to acknowledge an important aspect of digital acquisition: it is always a reading of an object with sensors. Sensors here are defined broadly as the interface and filter between the information coming from the material analogue datum (e. g. an ancient vase or the page of a book) and its sensorial post-perception raw format, which is a signal or status immediately produced inside the perceiver. It is 'raw' because it has not been processed or recognized by some internal mental or machine process into a fully digital, that is symbolic and discrete, object (Blachowicz 1997). Most of the time, when we digitize something, we are attempting from a practical standpoint to capture some of the object's physical features in a fashion akin to how we read them with our senses. For example, humans perceive a painting or the pages of a book as these objects selectively reflect the ambient light, through the eye's photoreceptors; our hands might sense the texture of a statue or architectural element through their terminal nerves and thus allow us to perceive form and shape. Importantly, local changes in light, humidity, our internal mood, and so on, do change our perception of the object and, to a degree, the material properties of the object itself: higher humidity might make the oil colours of a painting interact with ambient light differently. Perception is, so to speak, about the perceiver as much as the perceived. Digital sensors try, often by mimesis, to replicate their biological equivalent, or when they do not, they are at least functionally oriented to perceive and record specific physical occurrences even when they are quite removed from biological senses. For example, a digital sensor in a camera will use a receptive field to perceive the light coming from a painting (Rudolf 2006). A simpler device such as a keyboard, which is used to input streams of distributed signals (the letters), can be equally framed as an interface between the outside world and the internal states of a machine: when we type, we input a raw signal that is sequentially sent into the digital system.

At the onset then, before giving meaning to the perceived raw datum, sensors simply receive information from the environment. This information is limited by the capturing device's own inherent constraints: we do not perceive all light-waves with our eyes, not all cameras have the same sensitivity, not all keyboards have the 'right' letters, etc. Digitizing is firstly a reading with sensors, but the language of sensors is raw: it requires codification.

After being discriminated by a sensor, the perceived raw information is introduced in some format within the receiving digitization system. It is at this stage that the process of description happens. The flow of inputs from the sensor is processed and manipulated into a specific representation of the external non-digital object. It is part of the nature of digital objects, in the words of John Haugeland, to involve 'some form of writing and then reading various tokens of various types' (Haugeland 1981: 214, 220). Inputs are therefore parsed into bits that can be read and stored. For example, the keyboard inputs, which are electric signals, are internally discretized by the machine and their symbols memorized in a file by means of some encoding process. The encoding process associates to each input in time some internal symbolic value to create the sequence, the stream of signal encodings, that compose the inputted letters or commands from the keyboard (Lewis 1971; Maley 2011). The whole process, typing, is probably one of the most basic digitizing processes for a textual source. A page-scan is not too different though: the raw input is the light-signal perceived by a camera or other device, encoded as a file which then needs to be decodified into a digital representation, an image. Importantly, the text within the page is invisible to the machine *prima facie* but is meant to be re-discovered by the user once the digital image is displayed.

In a sense, then, the interpretation of raw signals is nothing else than their description. To form a meaningful representation, we need to be able to describe what the different external inputs mean, and the couple sensor-encoder is designed precisely to achieve that. The sensor-encoder has the dual objective of filtering the existing elements outside of the digital system and then processing the sub-sample of signals it perceives by using a pre-ordained symbolic system that can differentiate the sample's elements (cf. Ciula and Eide 2017). To continue with our camera example, once the light sensors of the camera receive the light-waves they are receptive to (when the shutter opens), the sensors send some electronic signals which are then internally (in the camera's system) associated to a symbol, for example a discrete set of values, that describe the local colour value perceived by the sensor (cf. Hall 1979; Mather and Koch 2011).

It is in this passage from continuous signal to discrete symbol that lies part of the power of digitization. Notably, the process is designed to leave a sizeable amount of information out of the system, which is then lost: not all the nuances of colour might be captured by a camera, nor the odd typo, or the texture of a page when it is typed into a digital system (Drucker 2001). On the other hand, some information can become more perceivable digitally, for example when the resolution of a digital object far exceeds the perceptive capability of the human eye (Bertani and Consolandi 2006). The information is also susceptible to the environmental characteristics present at the time of the acquisition. While an original statue can react to the light in the room without changing its shape (although the emphasis about it might change), this is not necessarily the case for some different types of digitizations. A 3D scan of a statue, for example, might be constructed in a way that appears to be as neutral as possible, just the shape and texture. This choice of stylistic neutrality is far from neutral: specific lights will be chosen, shadows reduced to maximize the acquisition of shape to the detriment of other aspects, such as the emphasized volumetry of shades. It is possible to shift and change the colour palette or simulate surface shades, but these are subsequent simulated digital additions to a digital object that were not present at the time of acquisition. Any digital description is therefore something other than the original because it lacks or adds elements to the representation that the represented does not necessarily possess. Some of these elements might change at the time of acquisition. For example, and ideally, multiple captures could acquire all the possible lighting conditions, and finer sensors cover as many aspects of reality as possible. Even in this case, however, if the language remains digital, and therefore codified in a symbolic description, the analogue object it represents with its continuity of features would be lost, and a new object would emerge. The new object would however retain a connection with the original analogue. There is a clear relationship between the two, a connection that we can perceive.

It is important to acknowledge that the line between raw input and encoded symbol is increasingly blurred by computational photography pipelines and similar technologies. A modern smartphone or scanner often captures multiple exposures, performs noise reduction, and merges high-dynamic-range data into a final encoded material. In this sense, the device is performing multiple takes but flattening them into a single output (such as a single JPEG). For the purposes of the taxonomy proposed below, we classify objects based on the *accessible* layers in the final archived object. If the rich mid-way data, that is the raw data and separate layers that are then 'flattened' into a final output, are discarded in favour of a lossy format, the resulting object functions as the digital object regardless of the complexity of its creation.

So far, we have described digital acquisitions as descriptions of objects mediated by a sensor-encoder capable of perceiving and digitizing the inputs it receives. Once encoded, however, they become something quite different. Indeed, the symbolic nature of the digital object implies a more radical observation: a digital acquisition is not only a description in a conceptual sense, but rather a practical instruction to build the representation of the object it describes. This re-construction is achieved by reading the encoded symbols and reproducing them in a fashion that allows for human interpretation: the decoding step. For example, computer files storing 'images' are instructions to reproduce the image on a computer screen. However, the information in the machine, a matrix of numbers and additional instructions in machine-language (encoded), is not the information humans interpret. Each time a user 'clicks' on a computer file stored on their hard drive to see the digital acquisition of a print by Utagawa Hiroshige, the print is re-printed on the screen (decoded) and only at that point intelligible to the user. Additionally, because each digital object is a univocal set of instruction, it can be copied identically and indefinitely.[5]

3. Digital objects are instructions in symbolic language

At this point, the relevance of the instructional nature of digital objects warrants some space to clarify the terminology we are adopting in this essay. With the stated aim of achieving the greatest universality possible, we are opting in this argument for using symbolic language as the language of instructions rather than machine language. We are *de facto* assuming all digital objects, whichever their support (hardware or software) and their intricacy or simplicity, are ultimately reducible to the principle adopted here. Not all digital systems are electronic machines, but all digital systems compose (encode/decode) the symbols that the reader interprets. In fact, it is not very controversial to see machine language as a straightforward subcategory of symbolic languages.[6]

When a digital description is stored for future readings, it is stored in a way that allows for the symbols to be reproduced multiple times. Stored pictures are most often indications to compose a pixel-based description of an image, where a pixel is a small unit-tile of a defined colour. High-definition digital images have millions of pixels, creating the illusion of continuous colour when projected on the appropriate display. Audio files, similarly, register specific sonic impulses that are time related and ultimately discrete. 3D and 2D models are stored as coordinates of points and polygons or contours joining them. Each of these descriptions, and many more, are at their core (symbolic) models, and therefore abstractions, of the object that is being digitized. They possess information about how to describe it by means of basic units (pixels, sound features, polygons, etc.) that are assembled according to a specific instruction. The instruction to decode is saved together with the information (self-executing) or designed to be complementary to it in the decoder (readable). When we observe complementarity, the decoder must have the capability to recognize the encoded information and execute it by reading it correctly (for example by discerning between image formats such as JPG or PNG).

The most common symbols employed in the context of digital objects of an original analogue are numerical symbols (cf. Manovich 2001). This is a consequence of the way in which the digital description is produced. The sensors involved are measuring tools (which need some scale to measure reality) and the machines are, in almost all the cases, calculating ones, computers, which are based on numerical architectures. The description they produce is therefore one of measures, scales, addition of features which in their combination produce a specific impression of the object, and so on. The description can be loose and simplistic, for example a black and white scan is descriptively simpler (one matrix of pixels) than a colour one (multiple combined matrices), or incorporate multiple overlapping forms of information, such as ‘colours’ that are invisible to the human eye (the infrared or ultraviolet). Similarly, for digital audio descriptions, they can include inaudible sounds and background noises, or stereo sounds rather than mono. This wealth of information is then stored and organized in the descriptive ‘file’, encoded for future use. Each time that decoding happens, the organized description is re-enacted and produces the object’s representation. Interestingly, the encoding-decoding could be almost instantaneous in some instances; yet, as long as the symbolic nature is preserved, the merging of encoding and decoding into one passage is not too relevant for the representation of digitization as description: the key to the description representation is in the transition from analogue raw signal to encoded symbolic description. It is however relevant when we need to think about representations stored in larger digital systems, such as a file saved on a hard disk.

What is then the symbolic procedure that decodes an image? Let us resort to our imagination again and imagine that a digital image of a painting is acquired via high-definition photography, which at its heart has a specific sensor receptive to different light wavelengths reflected by the painting. The raw physical datum captured by the sensor is then codified in a specific format, in this example an RGB picture. The codification might be close to the original raw signal or quite distant after some elaboration; often one format can be transformed into another. RGB pictures are digital pictures whose content is saved as the combination of a red, green, and blue channel. Which means that the image is the overlay of three single colour images of variable local (pixel-level) colour intensity (e.g. dark or light blue).[7] Currently, for most applications and formats, each channel can take pixel values (related to colours) of 0–255. The numerical codification of colour is important here: it is an abstraction, and abstractions allow for further manipulation. As we have already remarked, when a user visualizes a digital picture on a computer, this person is seeing what was recorded at the time of the acquisition, which will be as close as possible or necessary to the original, but what is viewed is, in fact, a codified abstraction. Because codified signals have structure to them, we can work with the codes and create software (i.e. a rule-based system) that can elaborate the information encoded in the digital image. Codified images can thus be processed to look for specific structures (e.g. a corner) because the underlying code preserves this information in its structure. Figure 1 shows an example of this.

When a user ‘opens’ the stored digital image, the image is assembled on a screen using the RGB channels information: each pixel is associated to a little led on the screen which with enough high resolution can reproduce an image that looks as if it was formed by continuous colours. If the image is printed on paper, ink is mixed according to the pixel structure and each space of paper has that value impressed. In this latter example of printing, however, the result is an analogue object (the colours mix slightly, and the neat pixel information is lost) printed from the digital. The passage digital to analogue is therefore possible and, in some instances, legitimately part of the decoding phase. What happens in this case is a double transformation from analogue to digital, and then from digital to analogue. The second analogue is perhaps a copy (degraded, i.e. with less information) of the original analogue but might differ by such a degree in shape or format as to be a completely different object with a resemblance to the original only.

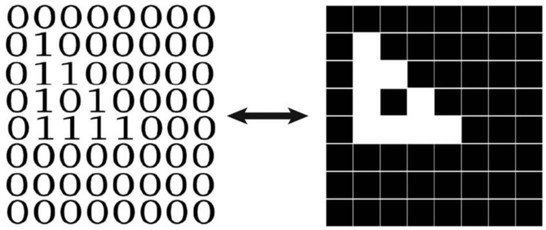

Figure 1. Binary Representation of an 8 × 8 Pixel Image. A binary image is an image with only two possible colours for each one of its local states (e.g. black, or white) which can be codified as 0s and 1s. Working on the underlying codification allows us to teach a machine to extract or find specific features, for example the number of 'vertices' of the image.

Extracting the encoded information from a digital image is another matter altogether. It requires specific software or new ways of looking at the information codified in the digital material. In the example of a digital image, different domains of computer science have developed for the precise task of correlating, transforming, or pointing to specific structures within the image. Disciplines such as Image Analysis or more recent applications such as DeepLearning, all work with the coded source to achieve specific ends, be it image manipulation for the former, or the classification of image content and features for the latter (cf. Pitas 2000; Goodfellow et al. 2016).

There are some fascinating consequences if one adopts the framing we proposed for the process of description, the triad of sensing-encoding-decoding. One immediate area of consideration has to do with the implicit legacy of positivistic readings of reality. Because most sensors involved in digitization projects today reduce reality to 'chunks' of information, selecting what to leave out by putting some blinders on while magnifying what is left in the field of vision, there are some inherent choices that are made at the point of design which are not fully appreciable at the point of fruition (cf. Bowker and Star 1999). This leads to questions concerning the participation of digital curators and collectors in the development of the acquisition technology: if passive, they work under the hard constraint of the pacing of acquisition set by their sensors and thus by others; if active, the power dynamic between the developers of acquisition tools and the curators or collectors is worthy of further enquiry. The inherent choices worth studying go beyond this and cover the soft representation preferences made at the point of description (the encoding phase), and those happening at the decoding phase. They work behind the scenes but heavily influence how the perception of the digital object is mediated. The conscious and unconscious effects of the descriptive setup that digital objects have on the end reader has yet to be fully explored. The issue of representation and encoding choices is related to the notion of 'bias' in the representation of the object: choosing what to leave out of the digital description is a choice fraught with consequences. This selective nature reflects what Thylstrup (2018) describes as the 'politics of mass digitization', where digitized cultural heritage becomes part of new networked assemblages, whether corporate, collaborative, or community-driven, each operating through its own infrastructural logics rather than traditional archival frameworks.

The presence of bias and preconceptions in digital objects' acquisition, of constraints hard and soft, is not the only consequence of digital representations. The choice of what to leave, what to emphasize, and what to pre-analyze in a digital object is one of choosing what to translate from the language of the original analogue to that of its digital description. Considering this, it seems then that *digital translation* is a more fitting term to describe what digitization is: the symbolic instructions are the new language of the translated digital object, and digital objects are a retelling through that language of the original. Previous scholars have described digitization as 'translation' in a metaphorical sense (e.g. Cameron 2007; McCarty 2007: §2), and recent contributions have concentrated either on the difficulty of qualifying digital objects against their capacity to narrate and connect between each other (Meehan 2023; Spurina 2025) or on the political and material qualities attached to the act of digitization (Thylstrup 2018; Zuanni 2021). Yet, to our knowledge, within the specific domain of digitized material collections for DH practice, a formal taxonomy based on the *mechanism* of this translation remains absent. Existing frameworks discuss *that* objects are translated or *how* to use them as data, but no work to date specifies how that translation operates at the level of signal capture and symbolic encoding and decoding, nor how one might measure its depth across collections. The framework discussed in these pages does both things.

4. Digital objects are translations—and vary by depth

Translation is preferable to the notion of copy or surrogate primarily because digital objects are supported by a completely different representation space than the original. Part of the misunderstanding around the two terms stems from the fact that digital translations have strong 'aboutness' when it comes to their relationship with the original.[8] This is a consequence of how they are designed: namely, to reproduce the experience and perception of the original. However, this is also a trick, an illusion produced by the fact that the symbolic representation of the object is hidden once the digital representation is printed on a screen or other support. The incorporated pre-made description and representation of it that the final reader sees decoded, is also one she has at that point no power upon. In fact, often readers assume that the object they see is a *copy*, one that is closer to the original than it really is if only they could peek into the secrets of encodings. This misunderstanding can be especially tricky in the Humanities because the relatedness between the original and the translation is functional and not substantial. There is, to sum up, an apparent relationship in the visual illusion of the representation which satisfies the need to reproduce the visual experience of the analogue but hides the real distance between them at a more fundamental level. One is a material object made of layered 'stuff', the other a series of commands which describe the object and then represent it for the viewer by re-enactment.

Much like with the notion of description, translation has interesting connections with debates within the Humanities and beyond.[9] If the symbolic language of digital objects is indeed a translation, and the different symbolic languages can be translated between each other, important issues around the incommensurability of language arise. These reference frames are embedded in the language and difficult to evade but need to be acknowledged and studied (Dobson 2019). Similarly, when moving from one digital language to the other (e.g. from one image format to another) the different choices made at each point might influence the result: repeated translations should increase the distance from the original. Finally, translations open the debate to issues of underdeterminacy and even meaning between languages. The critique of Quine and the responses from authors such as Kripke, Chomsky, or Kirk, remain an important reference here (Miller 2007: 141–63). Quine's argument sees a distance between languages such that any translation is indetermined: for coherent mappings of meaning ('translation manuals') different translation choices are in fact equivalent, i.e. there is indeterminacy between competing translations. If one were to apply this idea to our case, it would reinforce the notion that digital objects are something other when compared to the original.

The clarity that a language-instruction framing brings to a critical understanding of digital objects is operational. Digital objects have properties that have been captured by sensors and inscribed into symbolic language descriptions. In the case of digital objects that are of interest for the humanities, the language is machine language, and the stored objects can be read by a machine and manipulated inside its software environment. Additionally, because digital machine instructions are structured descriptions of objects, they possess a pre-digested layering of information. For example, peaks in music tracks are already described as numerical peaks in intensity, corners in images and their colour palettes are pre-codified in the RGB channels, and so on. They can all be brought into the foreground by means of algorithmic analysis, that is rule-based procedures that can emphasize or process the information layered inside the instructions as description that constitute digital translated objects. The power of the algorithms, however, is limited by the degree of malleability of the digital translation. If information is compressed or lacking, the amount of analysis that will be possible to produce will be limited and not much farther from the simple purpose of displaying the digital object as intended. On the other hand, complex acquisitions of the original analogue might produce rich pre-structured descriptions that would carry more than a single datum about the object and its context. This data, as we already stated, will still be limited by the sensors and the symbols used to encode, but will also offer to the scholar an opportunity to further describe and narrate. Once an object is translated in digital form, a hidden essay on its shape, structure, and content, lies dormant in symbols and code, awaiting the reader. Digital collections are the custodians of these essays and translations. We should think of them differently.

Even if we are at a relatively early stage in the digitization of collections, the types of sources and extant methods to achieve their digital preservation or communication is enormous (e.g. Guldi and Armitage 2014). However, any exhaustive description is elusive for two reasons. The first reason is that digital conventions change in time, and therefore what was once a technique to achieve digital permanence in a certain field, becomes obsolete relatively quickly (Brand 1999; Peters 2015; Uricchio 2015). This makes any overview by source bound to become outdated as technological innovation progresses. The second reason is that organizing digital means by kind (audio recording, 3D digital model, electronic image, etc.), while of great use for an exploration of possible solutions to archival needs, is not as useful when we need to think about the potential uses of digital sources transversally (Srinivasan and Huang 2005). Consequently, the taxonomy presented here relates digital sources to the mode of carrying information about their analogue counterpart, rather than the mode of their information codification or language. This is a conceptual approach which desires to be independent of subsequent incremental technological innovations in any one technique and as applicable as possible across digital means and their physical counterpart.

As we have seen, in the most abstract sense, an object's digitization is the carrying forward to digital format of the information it possesses as read by a machine or inputted in a machine. In a more precise sense, it is a description of the object, akin to the impression of some written or visual description by an author or artist on a book's page. The analogy is not a passing one. Computers store a stream of information bytes on their physical memories, which they can read and reinterpret each time the chunk of information describing one object is recalled from their memory. Digital machines possess the means to write, store, and read this information, and each time it is recalled they display the digital material according to some internal rule. Most of us are familiar with the process, so much so that we tend to forget the procedure generating the information we recall on our machines: a long series of commands that tell the machine how to 'build' what we ask for. Indeed, it is more correct to say that machines store instructions on how to write the information that we want to display, rather than the information itself. A 'text file' on a computer's memory contains the instructions to write the actual letters on the screen; it is the user, not the machine, then that interprets the information the machine re-assembled as a page or text.

This premise, which the first part of the essay explored in detail, is important when we discuss digital sources created from a physical analogue. Digital versions of the same object are created by acquisition and elaboration of the impression it leaves on a reader-machine and its codification of that input into structured information. Because of the variability in techniques and readers, digital versions can therefore differ in quality or even in the degree and kind of information they display about the digitized object. This latter aspect is also a consequence of the sort of technologies we possess. Better tools do describe more about their object, sometimes in areas beyond the sensing capacity of an individual. For example, X-ray scans of oil paintings can reveal hidden aspects of a painter's work and be stored in digital form (Dik et al. 2008; Twycross 2008). This digital layer of information can then coexist with a more traditional digital photography of the same painting.[10] Thinking the way we do here highlights the remarkable fact that digital materials are related and transformed representations of the originals, and the result of a precise acquisition mechanism. For the sake of our discussion, we can imagine an idealized procedure of digitization to be comprised of the following steps: (1) selection of the object, (2) acquisition of the different target 'layers' of information it possesses (e.g. image and X-ray image), (3) codification in digital language of such layers (i.e. the 'rule' we discussed), (4) long-term storage and (5) short-term efficient retrieval of the digital material. Our taxonomy focuses on the degree of removal from the original object of the connected information layers, which we will call *digitization depth*: the more abstract and synthetic the information stored about the object becomes, the higher its digitization depth. In short, our focus is on steps (2) and (3).

Using digitization depth as our compass, we can construe several degrees of abstraction with respect to their original source, which can be reconducted to roughly three modes of digitization. These three modes are introduced here for the first time with the aim of highlighting how stratified information is within digital translations. The first mode, which we could label *equalized* or *isomorphic* digitization, attempts to digitize the object in the simplest possible way and such that the basic features of its form are preserved. This is the digitization of scanning an image and encoding it as something reproducing its colour surface (e.g. JPG), or of scanning a book and saving it as an unsearchable collection of images forming document only when read (e.g. unsearchable PDF). Isomorphic digital processes are becoming less common for text but are still the norm for images and other non-textual sources.

The second mode adds a layer to the digital material by introducing additional information that is about it, but not in it. This is what we could call *descriptive* or *epimorphic* digitization, that is a digitization where additional elements about the *isomorphic* digital form are added. In its simplest form it is some added metadata about the object (author, year of production, etc.) but it can also highlight features of an image, or hide them, or add some form of functionality to navigate the material (Prescott 2008). Searchable scans of original documents are a good example of this, as the text that can be highlighted is a hidden layer applied on top of the scanned page. Epimorphic digitization is a natural choice for online digital humanities projects and often a feature of choice for collections and special projects or exhibitions. For example, it could be used to help users understand the composition of a painting or see small features that would otherwise escape them. Finally, epimorphic techniques are forms of embedding of knowledge: much like a curator that could explain or point to features in an object, overlay codifications extrapolate that information and superimpose it to the source (cf. Fig. 2).

The last mode of source digitization, and the deepest in terms of layers, is what we might label *modular* or *polymorphic* digitization (cf. Sorapure 2003). This is arguably the most laborious but rewarding approach, and

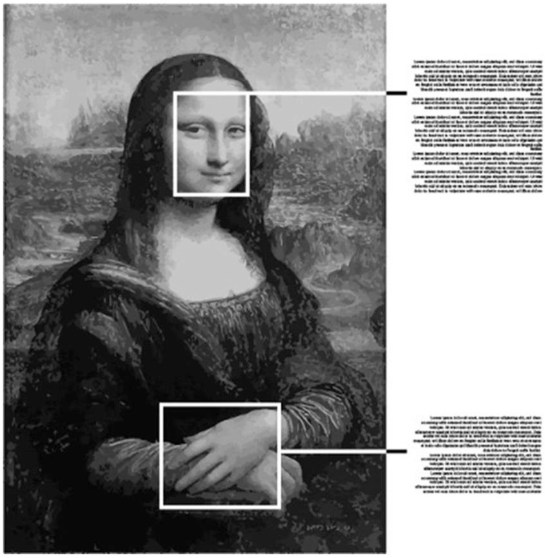

Figure 2. Example of Textual Epimorphosis of a Digital Image. An epimorphic source is a source which is complemented by external information about it that facilitate its content's comprehension or navigation within the object. This information is displayed (often interactively) with the source but can exist separately. Source of baseline image: Wikimedia Commons Digital Picture of Leonardo da Vinci's 'Portrait of Mona Lisa del Giocondo', Louvre Museum, Paris.

the least frequent in digital collections so far because of past technical limitations and the limited size of most digital projects. Polymorphic digitization is a form of digitization that takes an analytic approach to the source. The source material is 'decomposed' in its most basic components and presented with high depth of additional information (e.g. frame materials data or X-ray scans for paintings) which cannot be gauged by simply looking at the object. This approach leverages the plain fact that digital sources are not the original source; rather, they become a means to study it through 'sensorial' lenses. The overall combination of layers of information creates a rich palette of possible analysis patterns, as well as connections between materials (cf. Fig. 3). The key aspect of a modular approach is flexibility. Users and curators alike can work with greater freedom and add to the layers by post-processing the information. For example, additional features could be extracted via deep-learning and made available to the next users, or specific models could be applied to 2D or 3D scans to provide a way of studying their properties (e.g. Lecoutre et al. 2017).

Each digitization step carries forward some form of 'hidden information' which can be picked up and processed by a machine. Polymorphic processes therefore render the same object in a variety of superimposed 'forms', which in turn can be used for copy and surrogation.[11] When applied to collections, this modular approach shares many similarities with Kenderdine et al.'s (2021) concept of 'computational museology,' which advocates for replacing static 2D surrogates with deep, interactive datasets that allow for 'embodied' and multi-dimensional engagement with the source material.

The three modes of digitization are naturally connected, for they are about 'how far' one goes into digitizing something. Their taxonomy is useful but blurred at the edges. It is important then to clarify how one should discriminate between dubious cases. At the same time, as most of these classifications, they serve a navigational function and might not in every instance define an object sharply. Concisely, the distinction between the three taxonomical classes can be based on the observation that, while epimorphism is the addition of *external information about* the object, a polymorphic source is a source with different *overlapping readings of* the same object. The first is a commentary on the material, the latter a translation between analytic languages. It follows from this that an isomorphic source is simply a polymorphic source with a single translation layer (so an ideal polymorphism index would equal one) and no external epimorphic description. As a general rule, in case of concurrent elements, the chosen denomination should aim to describe the predominant feature. To move this taxonomy from theory to archival practice, we can operationalize these definitions through a classification rubric (see Table 1). Curators can determine the 'depth' of a digital object by asking: Does the file present a single visual layer? Does it contain embedded, machine-readable metadata?

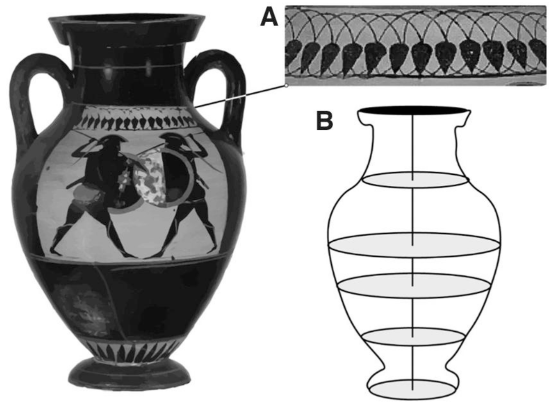

Figure 3. Modular Digital Acquisition of a Vase. This figure presents an example of modular (polymorphic) acquisition. A bare-bone digital modular approach to digitize a vase (in this case a Greek one) would decompose the object into its texture ((a) – detail of the overall pattern) and 3D shape (b). Such an approach would create a digital composite that can be analyzed in different ways. Shapes or textures can be related to known classifications or compared with specific metrics. Additional layers can be introduced. For example, porosity of the terracotta, ultrasound scan of the content if sealed, chemical map of the coating, and so on. Source of baseline image: Terracotta amphora (jar) ca. 530 BC, The Metropolitan Museum of Art, New York City, Accession Number: 56.171.14. Retrieved online at https://www.metmuseum.org/art/collection/search/254872 in May 2021.

Table 1. Taxonomy of Digitization Depth. This table presents an overview of the three proposed modes of digital translation (Isomorphic, Epimorphic, and Polymorphic). Their categorization is based on the level of descriptive capacity carried forward by the digitization process they underwent.

Taxonomic class	Translation depth	Key characteristics	Common examples
Isomorphic	Surface	Single layer of description; prioritizes fidelity to sensorial experience; content is unreadable by machine.	Flat JPEG/TIFF; Unsearchable PDF; Standard Audio file.
Epimorphic	Descriptive	Visual layer combined with an Information layer that contains external data about the object or searchable text within it.	Searchable PDF (OCR); IIIF Image with Annotations; Georeferenced Map.
Polymorphic	Analytic	Multiple overlapping layers of description; often captures data invisible to the senses across physical channels (e.g. X-ray scan or heat cameras).	3D Mesh with texture maps; Multispectral imaging set; RTI.

Does it allow modular analysis? This type of questions can quickly facilitate the positioning of an object within either class.

As a way of closing the taxonomical discussion outlined so far, we can tie it with the significance of interpreting digital sources as a 'translation'. In everyday language, a translation is the process of rewriting a source into another language in as close a fashion as possible to the original. Translations preserve meaning and structure, but translators often need to choose where to 'betray' the original in favour of readability in the new language. Like any other translation, in digitization there is an element of betrayal of the original as well as an attempt at the preservation of meaning and the structure about it. To digitize is to translate for the machine, and to translate for the machine is to make something countable. The notion of translation is rich in consequences: if our digital copies are indeed digital translations, we should not keep reading them as if written for human eyes only, rather scholars can exploit their symbolic-qualities and leverage on the possibilities unlocked by symbolic and machine language. One such possibility is the ability of machine-language sources to be connected by feature, rather than by their external labels imposed by third parties, and this connection can be explored, for example, in its change across time (cf. Jockers 2013). The array of digital applications developed in the past decades shows that the machine-readability is fecund and already clear to the digital scholar. Yet, its presence beyond the premises of pioneering projects is still limited. This sometimes might happen because of disciplinary resistance, but more frequently mundane cost barriers and data-organization bottlenecks are the real offenders. Once digital sources obtained from hard originals are conceptualized as a new source in themselves, digital collections acquire a different role in the ecosystem of knowledge production: they can become the centres sparking a new wave of studies. It is to this possibility that the closing parts of the article are dedicated.

5. Conclusion: digital collections are collections of translatons

Digital objects are used today in a myriad of applications, from deep learning to textual analysis, and their application has now fully come into the humanities beyond a basic operational use. Digital repositories however are still mostly organized around the notion of digital objects as copies, and do not permit for the leveraging of their translational nature: download bottlenecks make it hard to compare large groups of sources on a researcher's machine, and the sources cannot be downloaded or accessed in significant amounts; alternative file formats and technical metadata on the acquisition are not always available, making it difficult to understand what encoding and choices have been made at the point of acquisition; the quality of the digital material is often mixed, mainly because the collection is seen as a catalogue of the original analogue equivalent, and the online interfaces might be designed with an eye to keeping the information within the institution, rendering any further elaboration very difficult if not impossible.[12] If digital collections are collections of copies, isomorphic and epimorphic sources dominate, as the deeper level of analysis is not perceived and polymorphic sources are not deemed relevant if seen through the lenses of functional equivalence or direct educational purposes. If digital collections are translational, the best digital translations are modular, polymorphic ones, for they carry the highest amount of descriptive content on the source and therefore are akin to richer descriptions of the original.

In this context, a collection of translations can become a space for new scholarship. It does so by linking the descriptive elements of the sources between themselves and to other areas of knowledge in a way not dissimilar to what has been done so far in other fields. One such field is the study of book bindings, where the content of the book is linked to descriptive information about the binding materials. This linking creates new narratives and deepens the engagement with the object, strengthening the understanding of its value. It is not hard to imagine a similar process applied to modular digital translations: once the layers become evident, the extracted information can be linked to other layers or the original to form a rich palette of stories.

Collections worldwide are not oblivious to the challenges or opportunities of using digital objects. In the past decades much has been produced in terms of proposed standards, strategies to avoid some of the bottlenecks we described, and overall improved access (Jantz and Giarlo 2005; Meehan 2022). Reconceptualizing digital objects as translations aligns with the 'postdigital' reality described by Meehan (2023), where digital objects act as nodes in vast, transnational networks. If we treat digital objects merely as surrogates, we limit them to a subservient role. However, if we treat them as modular translations (such as polymorphic ones), they acquire the agency to be linked, re-interpreted, and 'networked' across boundaries: connecting a cloth pattern scanned in a European museum to trade routes described in a Bengali archive through shared data layers rather than just visual similarity. A collection of deep translations, therefore, becomes a generative space for new histories, rather than a static archive of copies.

This shift also maps onto existing technical infrastructures. The 'depth' of a digital translation dictates its compatibility with standards like the International Image Interoperability Framework (IIIF), which implicitly supports epimorphic and polymorphic depth through its canvas and annotation capabilities.[13] Similarly, provenance standards like PREMIS can be utilized to record the 'translation event' itself by documenting the specific sensor settings and encoding choices that shaped the resulting digital object.[14]

To conclude: digital objects, when produced by means of digital acquisition, are translations of the original analogue. The translations maintain some features of the original, shifting the emphasis on the object's structure and encodable characteristics. They are not a copy or surrogate per se but can produce either depending on the decoding choices. What is more, translations are also instructions to represent the object. They make it possible to leverage the properties of digital systems to process digital language and analyze the descriptions they carry within them. The notion of digital translation captures, then, aspects of the process of digitization that can be missed when we opt for other framings of the procedure.

When seen through the lens of translation, digital collections can be evaluated in terms of their capacity to provide spaces for the fruition of the descriptive elements encoded in digital objects, as well as the ability to allow users to enrich such descriptions with new digital acquisitions. We do not yet know all the possibilities that our digital collections possess. What we know however is that each digital source is a unique description of the original, while new acquisitions of the same source can provide new details. We are also aware of novel technologies that can be employed to explore the connections made manifest by digital translation. Digital collections are a great library of essays on material objects written in symbols. While we increase its volumes, we should not disregard the possibilities offered by their existence: they house more knowledge than might appear at first sight.

Author contributions

Giovanni Maria Pala (Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing—original draft, Writing—review & editing)

Conflicts of interest. None declared.

Funding

This article is based on research that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 883758).

Notes

1. Cf., for example: https://www.factum-arte.com/pag/recording-facsimile-reina-maria-luisa.
2. *The Factum Foundation for Digital Technology in Conservation*, and its commercial twin, *Factum Arte*, are both heavily involved with cutting edge technology for the acquisition, preservation, and high-fidelity replication of historical spaces or objects. Our thought experiment is not too distant from the projects covered in their reports. See: https://www.factumfoundation.org/ and https://www.factum-arte.com/.
3. Important to note that according to Zuanni (2021), born-digital heritage objects enact their own materiality determined by their software/hardware infrastructure, file formats, platform dependencies and shifting preservation contexts. This materiality means their ontology is not simply that of physical artefacts transformed, but constitutes a different kind of heritage object altogether.
4. See Kirschenbaum (2008), Ernst (2012), and Parikka (2015) for arguments from a media-archeology point of view that the storage substrate and code component shape what a digital 'object' can be. These look deeper than the surface and anticipate some of the arguments of this article.
5. Crypto art is trying to work around this by creating a token, an identifier, which is an enforcement of uniqueness by *fiat* rather than inherent to the object, which is what they miss as reproducible digital instructions.
6. A proof could go along the lines of showing the symbolic nature of modern mathematics, then introduce Turing machines as an instance of such symbolic use and link them to the machines we are discussing (cf. Klein 1968; Copeland and Proudfoot 2011).
7. A pixel is the unit of colour that comprises an image. They are small tiles that, like in a mosaic, together form an image. Modern images usually found online might be comprised of as little as two hundred thousand pixels to several millions.
8. 'Aboutness is the relation that meaningful items bear to whatever it is that they are *on* or *of* or that they *address* or *concern*' (Yablo 2014).
9. The Digital Humanities themselves, as a discipline, have been framed by some authors as part of implicit agendas whose ethos is potentially subversive with respect to the Humanities as a discipline (cf. Allington et al. 2016; Brennan 2017).
10. Important to note that a description is a digital element of an object, and so are the metadata. There is here an interesting case of an analogue with a digital layer.
11. As another example, 3D printing is a form of data-production, although hardware, which follows the same process and should therefore be assimilated to any other analogue printing procedure we discussed.
12. This issue becomes even more relevant when we observe a certain degree of dissatisfaction with the current usage of digital collections. See for example the work of Gibbs and Owens (2012).
13. URL to IIIF: https://iiif.io/
14. URL to PREMIS: https://www.loc.gov/standards/premis/

References

Allington, D., Brouillette, S., and Golumbia, D. (2016) 'Neoliberal Tools (and Archives): A Political History of

Digital Humanities', *LA Review of Books*, available at: https://lareviewofbooks.org/article/neoliberal-tools-archives-political-history-digital-humanities/.

Baudrillard, J. (1994) *Simulacra and Simulation*. Ann Arbor, MI: University of Michigan Press.

Benjamin, W. (1969) 'The Work of Art in the Age of Mechanical Reproduction', in H. Arendt (ed.) *Illuminations*, pp. 217–251. New York: Schocken Books.

Bertani, D., and Consolandi, L. (2006) 'High Resolution Imaging in the Near Infrared', in L. MacDonald (ed.) *Digital Heritage*, pp. 211–38. London: Routledge.

Blachowicz, J. (1997) 'Analog Representation Beyond Mental Imagery', *The Journal of Philosophy*, 94: 55–84.

Boon, M. (2010) *In Praise of Copying*. Cambridge, MA: Harvard University Press.

Bowker, G. C., and Star, S. L. (1999) *Sorting Things Out*. Cambridge, MA: MIT Press.

Brand, S. (1999) 'Escaping the Digital Dark Age', *Library Journal*, 124: 46–8.

Brennan, T. (2017) 'The Digital-Humanities Bust', *The Chronicle of Higher Education*. https://www.chronicle.com/article/the-digital-humanities-bust/

Burns, J. E. (2017) 'The Aura of Materiality: Digital Surrogacy and the Preservation of Photographic Archives', *Art Documentation: Journal of the Art Libraries Society of North America*, 36: 1–8.

Cameron, F. (2007) 'Beyond the Cult of the Replicant: Museums and Historical Digital Objects—traditional Concerns, New Discourses', in F. Cameron and S. Kenderdine (eds) *Theorizing Digital Cultural Heritage: A Critical Discourse*, pp. 49–75. Cambridge, MA: MIT Press.

Ciula, A., and Eide, Ø. (2017) 'Modelling in Digital Humanities: Signs in Context', *Digital Scholarship in the Humanities*, 32: fqw045.

Conway, P. (2015) 'Digital Transformations and the Archival Nature of Surrogates', *Archival Science*, 15: 51–69.

Copeland, J., and Proudfoot, D. (2011) 'Alan Turing: Father of the Modern Computer', *The Rutherford Journal*, 4. Available at: https://www.rutherfordjournal.org/article040101.html

D'Cruz, J., and Magnus, P. D. (2014) 'Are Digital Images Allographic?', *The Journal of Aesthetics and Art Criticism*, 72: 417–27.

Deegan, M. (2006) 'Surrogacy and the Artefact', in G. E. Gorman and S. J. Shep (eds.) *Preservation Management for Libraries, Archives and Museums*, pp. 54–63. London: Facet.

Dik, J. *et al.* (2008) 'Visualisation of a Lost Painting by Vincent van Gogh Using Synchrotron Radiation Based X-Ray Fluorescence Elemental Mapping', *Analytical Chemistry*, 80: 6436–42.

Dobson, J. E. (2019) 'Can an Algorithm Be Disturbed? Machine Learning, Intrinsic Criticism, and the Digital Humanities', in *Critical Digital Humanities: The Search for a Methodology*, pp. 1–28. Urbana, IL: University of Illinois Press.

Drucker, J. (2001) 'Digital Ontologies: The Ideality of Form in/and Code Storage—or—Can Graphesis Challenge Mathesis?', *Leonardo*, 34: 141–5.

Drucker, J. (2011) 'Humanities Approaches to Graphical Display', *Digital Humanities Quarterly*, 5. Available at: http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html

Ernst, W. (2012) *Digital Memory and the Archive*. Minneapolis, MN: University of Minnesota Press.

Gibbs, F., and Owens, T. (2012) 'Building Better Digital Humanities Tools', *Digital Humanities Quarterly*, 6. Available at: http://www.digitalhumanities.org/dhq/vol/6/2/000136/000136.html

Goodfellow, I., Courville, A., and Bengio, Y. (2016) *Deep Learning*. Cambridge, MA: MIT Press.

Goodman, N. (1976) *Languages of Art: An Approach to a Theory of Symbols*. Indianapolis, IN: Bobbs-Merrill.

Grycz, C. J. (2006) 'Digitising Rare Books and Manuscripts', in L. MacDonald (ed.) *Digital Heritage*, pp. 33–68. London: Routledge.

Guldi, J., and Armitage, D. (2014) 'Big Questions, Big Data', In: *The History Manifesto*, pp. 88–116. Cambridge: Cambridge University Press.

Hall, E. (1979) *Computer Image Processing and Recognition*. New York: Elsevier.

Haugeland, J, University of Arkansas Press (1981) 'Analog and Analog', *Philosophical Topics*, 12: 213–25.

Hui, Y. (2016) *On the Existence of Digital Objects*. Minneapolis, MN: University of Minnesota Press.

Jantz, R., and Giarlo, M. J. (2005) 'Digital Preservation: Architecture and Technology for Trusted Digital Repositories', *D-Lib Magazine*, 11. https://doi.org/10.1045/june2005-jantz

Jockers, M. L. (2013) *Macroanalysis: Digital Methods and Literary History*. Urbana, IL: University of Illinois Press.

Kenderdine, S., Hibberd, L., and Shaw, J. (2021) 'Radical Intangibles: Materializing the Ephemeral', *Museum and Society*, 19: 252–72.

Kirschenbaum, M. G. (2008) *Mechanisms: New Media and the Forensic Imagination*. Cambridge, MA: MIT Press.

Klein, J. (1968) *Greek Mathematical Thought and the Origin of Algebra*. Cambridge, MA: MIT Press.

Kropf, E. (2016) 'Will That Surrogate Do? Reflections on Material Manuscript Literacy in the Digital Environment from Islamic Manuscripts at the University of Michigan Library', *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies*, 1: 52–70.

Lecoutre, A., Negrevergne, B., and Yger, F. (2017) 'Recognizing Art Style Automatically in Painting with Deep Learning'. in *Asian Conference on Machine Learning (PMLR)*, pp. 327–42.

Lewis, D. (1971) 'Analog and Digital', *Noûs*, 5: 321–7.

Maley, C. J. (2011) 'Analog and Digital, Continuous and Discrete', *Philosophical Studies*, 155: 117–31.

Manovich, L. (2001) *The Language of New Media*. Cambridge, MA: MIT Press.

Mather, P. M., and Koch, M. (2011) *Computer Processing of Remotely Sensed Images: An Introduction*. Hoboken: John Wiley & Sons.

McCarty, W. (2007) 'What's Going On?'. Plenary address for Countries, Cultures, Communication: Digital Innovation at UCLA, Institute for Digital Research and Education, University of California at Los Angeles, 10 May.

Meehan, N. (2022) 'Digital Museum Objects and Memory: Postdigital Materiality, Aura and Value', *Curator: The Museum Journal*, 65: 417–34.https://doi.org/10.1111/cura.12361

Meehan, N. (2023) 'Digital Museum Objects and Transnational Histories', in E. Bond and M. Morris (eds) *Scotland's Transnational Heritage: Legacies of Empire and Slavery*, pp. 171–84. Edinburgh: Edinburgh University Press.

Miller, A. (2007) *Philosophy of Language*, pp. 141–63. London: Taylor & Francis.

Padilla, T. *et al.* (2019) *Final Report–Always Already Computational: Collections as Data*. Zenodo. 10.5281/zenodo.3152935

Parikka, J. (2015) *A Geology of Media*. Minneapolis: University of Minnesota Press.

Peters, J. D. (2015) 'Proliferation and Obsolescence of the Historical Record in the Digital Era', in B. B.Tischleder and SWasserman. (eds.) *Cultures of Obsolescence*, pp. 79–96. New York: Palgrave Macmillan.

Pitas, I. (2000) *Digital Image Processing Algorithms and Applications*. Hoboken, NJ: John Wiley & Sons.

Prescott, A. (2008) 'The Imaging of Historical Documents', in M. Greengrass and L. M. Hughes. (eds.) *The Virtual Representation of the Past*, 1: pp. 7–22. Farnham: Ashgate Publishing.

Rudolf, P. (2006) 'Principles and Evolution of Digital Cameras', in L. MacDonald (ed.) *Digital Heritage*, pp. 177–209. London: Routledge.

Sontag, S. (1977) *On Photography*. London: Penguin.

Sorapure, M. (2003) 'Five Principles of New Media: Or, Playing Lev Manovich', *Kairos*, 8: 26–48.

Spurina, M. (2025) 'Digitization of Museum Objects and the Semantic Gap', *Heritage*, 8: 369. https://doi.org/10.3390/heritage8090369

Srinivasan, R., and Huang, J. (2005) 'Fluid Ontologies for Digital Museums', *International Journal on Digital Libraries*, 5: 193–204.

Thylstrup, N. B. (2018) *The Politics of Mass Digitization*. Cambridge, MA: MIT Press.

Twycross, M. (2008) 'Virtual Restoration and Manuscript Archaeology', in M. Greengrass and L. M. Hughes (eds.) *The Virtual Representation of the Past*, vol. 1, pp. 23–48. Farnham: Ashgate Publishing.

Uricchio, W. (2015) 'Replacement, Displacement, and Obsolescence in the Digital Age', in B. B. Tischleder and S. Wasserman (eds) *Cultures of Obsolescence*, pp. 97–109. New York: Palgrave Macmillan.

Yablo, S. (2014) 'Introduction', in *Aboutness*. Princeton, NJ: Princeton University Press. pp. 1–14

Zeimbekis, J. (2012) 'Digital Pictures, Sampling, and Vagueness: The Ontology of Digital Pictures', *The Journal of Aesthetics and Art Criticism*, 70: 43–53.

Zuanni, C. (2021) 'Theorizing Born Digital Objects: Museums and Contemporary Materialities', *Museum and Society*, 19: 184–98.