The preservation of Digital Heritage: Epistemological and Legal Reflections
Melanie Dulong de Rosnay, Francesca Musiani

To cite this version:

HAL Id: hal-00771856
https://minesparis-psl.hal.science/hal-00771856
Submitted on 9 Jan 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The preservation of digital heritage: epistemological and legal reflections

Mélanie DULONG de ROSNAY
Chargée de recherche, Institute for Communication Sciences, CNRS, Paris, FRANCE
melanieddr@gmail.com

Francesca MUSIANI
Researcher, Institute for the Study of Diplomacy, Georgetown University, Washington DC, USA
Centre for the Sociology of Innovation, MINES ParisTech, Paris, FRANCE
francesca.musiani@gmail.com

Abstract: Different disciplines and fields of study seem to be heralding the rise of an interdisciplinary scientific and intellectual movement focused on digital heritage, operationally defined as the ensemble of documents and information created in digital formats and subjected to preservation policies developed by individuals, companies and institutions. This article seeks to address some of the methodological challenges that – notwithstanding a diverse, thriving body of work that is currently contributing to the establishment of the scholarship on digital heritage – are currently facing scholarly attempts to consider digital heritage in its plurality. At the present, exploratory stage of the digital heritage scientific/intellectual movement, contributions to a reflection on the very foundations of this movement are needed, so as to refine the possible approaches of future digital heritage-related studies. This
article is meant to provide such a contribution, drawing on the authors’ experience with interdisciplinary approaches to subjects of study such as alternative, decentralized infrastructures for Internet services, or the techno-legal governance of data, the commons and the public domain. The article reflects on practical tools, and epistemological/theoretical foundations, allowing to define and to include in the analysis all the facets of digital heritage – its archives, traces and instruments.

**Keywords:** preservation, epistemology, copyright, infrastructure, typology

***

La conservation du patrimoine numérique: réflexions juridiques et épistémologiques

**Résumé :** Différentes disciplines et champs de recherche semblent être précurseurs de la montée d’un mouvement scientifique et intellectuel interdisciplinaire axé sur le patrimoine numérique, défini opérationnellement comme l’ensemble des documents et informations créés dans des formats numériques, et soumis à des politiques de préservation développées par des individus, des entreprises et des institutions. Cet article cherche à interroger certains des défis méthodologiques que – malgré le développement d’une littérature florissante autour du patrimoine numérique – se rencontrent actuellement pour considérer le patrimoine numérique dans sa pluralité. Au stade exploratoire actuel du mouvement scientifique et intellectuel du patrimoine numérique, des contributions à une réflexion sur les fondements mêmes de ce mouvement semblent nécessaires, dans le but d’affiner les approches possibles des recherches dans le domaine. Cet article est destiné à fournir une telle contribution, en s’appuyant sur l’expérience des auteurs avec des approches interdisciplinaires à des sujets comme les infrastructures décentralisées pour les services Internet, ou la gouvernance techno-juridique des données, les biens communs et le domaine public. L’article s’engage dans une réflexion sur les outils pratiques et les fondations théoriques et épistémologiques permettant de définir et d’inclure dans l’analyse l’ensemble des facettes du patrimoine numérique, ses archives, ses traces et ses instruments.

**Mots-clés :** préservation, épistémologie, droit d’auteur, infrastructure, typologie

***

**Introduction**

Digital heritage may be operationally defined as the ensemble of documents and information created in digital formats and subjected to preservation policies developed by individuals, companies and institutions. Its conditions of emergence,
usage, analysis, production, preservation, reproducibility and reusability should be studied in depth in order to keep track and preserve the memory of these intangible objects, and support their promotion, visibility and re-mobilization.

Drawing on complementary approaches, grounded in science and technology studies and aimed at embedding information and communication technologies in their production settings, the architecture subtending them, and their social and human contexts, this article intends to contribute to the ongoing reflection on the epistemological and legal foundations underlying the definition of digital heritage as a concept, and on the conditions of its constitution and preservation.

While a diverse, thriving body of work is currently contributing to the establishment of the scholarship on digital heritage, there are still a number of methodological challenges for any attempt to consider digital heritage in its plurality. For instance, the borders of the Internet are surpassing the mere Web and should not exclude content that is produced outside of it – software, languages or protocols. Besides, it is necessary to single out the different stages and actors of production and preservation, and to take into account the legal and technical restrictions for accessibility and reusability. This article discusses and proposes possible methodological approaches for (1) the definition of this complex object and (2) its legal, technical and political framework.

1. Digital heritage, a matter of interdisciplinary definition

The international, ever-growing community of Web historians (Abbate, 2012); the nascent Digital Humanities discipline (Schreibman, Siemensa and Unsworth, 2004); the crossing bodies of work of scholars in history, sociology of innovation, information and communication sciences, law and computer science (e.g. van Schewick, 2010): when intersecting with an interest in the preservation and promotion of digital archives and data, these disciplines and fields of study seem to be heralding the rise of an interdisciplinary scientific/intellectual movement (Frickel and Gross, 2005) focused on digital heritage. Interestingly, this current of study is taking shape as an “object-oriented” scientific movement: one that acknowledges the complexity and diversity of its subject of study as a starting point of its approach; equips itself from the beginning with an interdisciplinary toolbox; and reflects both the orientation and the methodology in its name, which is derived from its subject of inquiry, rather than from a discipline in particular – not unlike what the field of Internet studies has done in a recent past (Jouët, 2011).

1.1. What this article is (and is not) meant to do

Notwithstanding a lively, diverse body of work that is currently contributing to the establishment of the scholarship on digital heritage (Kalay, Kvan and Affleck, 2008), researchers in this field are facing a number of methodological challenges: the high level of technical skills needed to address many of the sources; those
sources being scattered in a number of databases and their opacity vis-à-vis particular audiences; and the lack of accessibility and appropriation of several digital archives; in brief, challenges in methods that may entail going back to the very definition of digital heritage: its perimeter, frontiers and boundaries, and the epistemological issues underlying it.

In France, the country that is at the core of the authors’ research, a number of State-driven, nationwide endeavors have been initiated. Among them, the seminar cycle on topics relevant to the “legal deposit” of Web sites (dépôt légal du Web), organised by the National Audiovisual Institute (Institut national de l’audiovisuel, INA) and the meetings promoted by the National Library of France (Bibliothèque nationale de France, BnF) stand out. However, these pioneer projects have only partially succeeded, so far, in addressing digital heritage in its most comprehensive meaning. In the case of the Internet, for example, it may be argued that digital heritage encompasses far more than the World Wide Web, to include elements as diverse as the memory of relevant actors (founding fathers and “great witnesses”, grands témoins); the exchange and sharing practices (email exchange, social networking and database shaping) that happen daily outside the Web, relying e.g. on peer-to-peer software that is increasingly private (Musiani, 2011); and the software applications developed by individuals and firms (Net Art, social media, videogames).

As the introduction has already suggested, this article certainly does not have the ambition of addressing directly, by means of case studies or fieldwork analysis, this articulate ensemble of socio-technical objects or to provide any definite answer to all of the methodological and epistemological challenges briefly introduced above. Nonetheless, we believe that at the present, exploratory stage of the digital heritage scientific/intellectual movement, contributions to a reflection on the very foundations of this movement are needed, so as to refine the possible approaches of future digital heritage-related studies. This article is meant to provide such a contribution, drawing on the authors’ experience with interdisciplinary approaches to subjects of study such as alternative, decentralized infrastructures for Internet services (Musiani, 2012) or the techno-legal governance of data, the commons and the public domain (Dulong de Rosnay, 2010) to reflect on practical tools, and epistemological/theoretical foundations allowing to define and include all the facets of digital heritage – its archives, traces and instruments.

1.2. Towards a typology of relevant venues, sources and actors

A preliminary condition for understanding the issues relevant to digital heritage is a further exploration and detailing of a typology of the relevant actors involved in its preservation and promotion, the material sources from which they draw – and from which scholars may draw, in turn. This implies an attempt to define the object by taking into account the distinction between digital and digitized heritage (Pernoo, 2009) and identifying the places, material and intangible, where it is preserved, as
well as the conditions of such preservation. Scholars of digital heritage may benefit from mapping efforts that would compare international and supra-national initiatives with their national counterparts or equivalents, and possibly show their complementarity, eventually going so far as to provide a global cartography or a first typology of relevant venues for the preservation of digital heritage. French core actors and venues, such as the above-mentioned BnF and INA, may be studied in a comparative perspective with initiatives such as the International Internet Preservation Consortium (IIPC), the Internet Archive, or even the Library of Congress in the United States, whose recent agreement to host billions of Twitter-generated messages has been described as a “unique record of our time” (Eversley, 2011).

A comprehensive inventory of what is to be included in the perimeter of digital heritage as a subject of inquiry should have as a precondition a reflection on sources. Interestingly, some of them may not be the target of a preservation endeavour yet, but the “history in the making” of the network of networks increasingly suggests that they cannot be neglected in an inclusive approach to the display and visualization first, and the study next, of digital heritage: audio-visual formats (Chiariglione, 2012), code, protocols, languages, off-Web and peer-to-peer contents, videogames, and databases.

Thinking digital heritage as a global socio-technical system equally entails the need for an inventory of the different methods of collection and digitization employed, and the different approaches to the “granting of heritage status” (patrimonialisation: e.g. Bortolotto, 2006; Skounti, 2009) specific to each venue of heritage collection. These collection methods and tools vary not only according to the diversity of objects that can possibly be included, but also to apprehend the same object: it is the case, for example, of the attempts to take a full inventory of the World Wide Web, by both the Internet Archive (Howell, 2006) and the Dépôt légal du Web initiatives. Furthermore, the actors in charge of the collection of “intangible” heritage make specific technical and legal choices, thereby shaping and influencing the accessibility of such data by scholars: for example, the modalities of appropriation and study of a videogame will vary greatly if it is run on its original support, or on a more recent version of it, if it is accompanied or not by technical instructions, whether or not its perusal is mediated by an emulator. A complete inventory of the tools, as well as the content, is likely to greatly benefit not only from a careful exploration of the available institutional and technical documentation, but also from in-depth qualitative interviews with the actors originating such initiatives and involved in preservation processes.

1.3. Digital infrastructure as heritage

A working, extended definition of digital heritage may at this point include the collection of all documents and information stored in digital formats, within the frame of preservation policies put in place by individuals, companies, and
institutions. However, we wish to suggest that a relevant part of this digital heritage cannot be analytically addressed without taking into account the technical infrastructures and the human environment that underlie its emergence and its different ways of shaping – in a nutshell, without reintroducing materiality in the comprehension of this “intangible” or “virtual” heritage. Code, programming languages, protocols, file formats, interfaces, network architectures, need to be taken into account for digital heritage to “speak” as an object of scientific inquiry. An epistemological reflection on the articulation between material and intangible heritage, between digital heritage and the sciences shaping our world of digital tools and supports, need to be initiated. To put it in the words of Susan Leigh Star, “Study an information system and neglect its standards, wires, and settings, and you miss equally essential aspects of aesthetics, justice, and change” (Star, 1999: 339).

Some venues and projects appear in this regard particularly useful to analyse, as case studies that may allow scholars to apprehend infrastructure as an integral part of digital heritage. Initiatives such as the KEEP European project, aimed at developing a mobile, universal emulator, or the debates carried on within the IIPC, can help shed new light on the technical constraints and opportunities underlying the preservation of digital heritage, through an exploration of questions such as: how can digital heritage be exposed or displayed with little or no risk of damage or degradation? To what extent can it freely circulate, and re-appropriation or manipulation by actors such as the public or the scholarly community may be permitted?

Finally, going deeper into the lower layers of information systems infrastructure – Internet “plumbing” (Musiani, 2012) – is an occasion to link the study of digital heritage to the understanding of the Internet as a commons (Massit-Folléa, 2008). Over the years, the constitution of the body of Requests of Comments (RFCs) documents by the Internet Engineering Task Force (IETF), and their subsequent online availability to the public, have increasingly suggested that code and protocols may be conceived as a commons, and by extension, as part of humanity’s heritage. However, this issue calls for a more detailed exploration: can the “instability by design” (Braman, 2011) principle, that guides the conception of Internet protocols in organizations such as the World Wide Web Consortium (W3C) and the International Standardization Organization (ISO), be qualified as digital heritage? Can “alternative” networking models, such as peer-to-peer, be studied as part of digital heritage, as an infrastructure of circulation, distribution and exchange able to facilitate its free re-appropriation?

Such questions cannot be addressed without taking interdisciplinarity seriously. A number of innovative perspectives on the digital world are opening up with the study of the intersection between information technology and heritage preservation: for example, the establishment of “network archaeology”, or an approach that considers code, programming and “material” digital infrastructure as a core
component of evolutions, appropriations and usability of technical artefacts. Research situated at this intersection is likely to require a methodological perspective that selectively draws elements from disciplines such as history, sociology of innovation, information and communication sciences, political sciences and law; and builds on computer science in order to clarify the management and transformations of formats, the database design process, and evolutions in network architecture.

These disciplines, complementing one another, are able to provide an analytical framework which embeds information and communication technologies in their contexts of production, their technical architecture, and enables to think these technologies in the interplay of technology and human interaction. Historians are accustomed – while maintaining them as separate objects of inquiry – to enable dialogue between time, memory and heritage; history of innovation, in particular, opens up to the social co-construction of technology and may, together with communication sciences, shed light on issues such as intermediation, formation of digital identities, reception, appropriation and acculturation, that contribute to the shaping of digital heritage. Sociologists of innovation can complete this perspective, drawing on recent advances in fields such as software, code and cyberinfrastructure studies (Fuller, 2008; Marino, 2006; Ribes and Lee, 2010), in order to explore the social and political qualities of infrastructures (e.g. Star, 1999), and balance “the deployment of critical terms like “virtuality” […] with] a commitment to […] recover and stabilize the material traces of new media” (Kirschenbaum, 2003).

Finally, the contribution of legal and political sciences to this interdisciplinary and architectural approach to the theme of digital heritage is able to provide a crucial contribution to its study, by addressing facets such as the relationship between access and reuse, copyright and format, heritage and the public domain. The next section delves into these issues in some more detail.

2. The legal conditions of appropriation of digital heritage

A reflection on the definition and the preservation of digital heritage needs to address copyright-related questions. The law is often used as a red herring to prevent a document to be reused, shared or even copied, with three main reasons. First, it is difficult to identify right-holders and manage rights that are at stake, especially in the case of multiform, complex objects. Second, the legal availability of intangible documents and information must not be jeopardized by a walled garden, a closed platform or a technical obstacle to export, copy or otherwise manipulate data. Third, the policy environment should better accompany the mission of digital heritage preservation, to maintain access to sources that exist only digitally (or “born-digital”, e.g. Unesco, 2003), therefore lacking the materiality that would not require the making of a copy. Unlike artefacts such as paintings, this “born-digital” state automatically triggers copyright and sometimes database trademark, privacy and publicity rights, leading to “a second enclosure movement of the intangible
commons” (Boyle, 2003) and the risk of not having a digital heritage at all until elements become part of the public domain.

2.1. The constitution of digital heritage collections

Guidelines produced by national or international institutions of preservation, for the benefit of local archives, libraries or thematic projects, are usually very clear (for instance, National Library of Australia, 2003). They explain the necessity to identify, on the one hand, rights pertaining to objects before integrating them in a database, and on the other hand, rights that will be required for users to make effective use of the preservation endeavour. The bundle of rights should be broad enough to allow consultation beyond the physical premises of the institution, and reproduction with the possibility of modification and treatments such as data mining, and other technical processing, which are necessary for research and porting to future formats.

Once inbound and outbound rights have been identified, then the institution may rely on an exception to copyright, a system provided by the law to limit right-holders’ monopoly and grant some prerogatives to users, such as libraries. But at least “twenty-one countries have no library exception in their copyright law”, which will cause a major problem of absence of cultural diversity for future generations accessing world heritage. Similarly, many exceptions are not broad enough and do not address the harvesting of online resources (Crews, 2008), creating legal uncertainty and chilling effects. Besides, some preservation projects are initiated by parties that are not libraries, for instance private individuals, projects, foundations or companies that would be out of the scope of the legal exception in the case it was provided by copyright legislation for libraries only. Alternatives may even be considered illegal, as it is the case of some file sharing services (Tushnet, 2006) even if they are providing a comparable service to the public. Sometimes, peer-to-peer services are of higher quality and fill the gap in the market failure for niche cultural items, as no other alternative is available commercially or in the library in that case (Bodo and Lakatos, 2012). The option to rely on unauthorized community efforts to preserve the digital heritage in the absence of legal solutions might well be the only workable solution in light of the challenges presented below.

If no legal exception can be used, the preservation institution will have to negotiate with the right-holders to be a priori identified and located. It can be particularly challenging to identify all rights applicable to a single web page, and impossible to ever clear all rights. For instance, a forum will most likely not have imposed a transfer of rights to its contributors or a perennial identification in order to request the permission to reproduce their personal information, which will not be covered by copyright exceptions. Also, a video game producer will probably have cleared rights to use the external elements incorporated in the design of its product for the purpose of selling the game, but not for future preservation by a memory
institution or any other party. Therefore, best legal practices developed by and for libraries may be of relative inapplicability; indeed, identifying right-holders may be an endless task and the absence of fair use or broad exceptions and limitations for preservation may jeopardize the legal security of the institutions. General counsels may oppose innovative projects coming from digital departments. The risk of liability is enhanced in the case of digital complex objects, which incorporate a plurality of external contributions and different rights.

The approach described above is particularly unsatisfactory and a strict application of current positive law is not the way to move forward in order to tackle preservation needs. The World Intellectual Property Organization is currently discussing Exceptions and Limitations for Libraries and Archives (WIPO, 2012). However, the publishing industry and developed countries, namely the United States and the European Union, may well oppose the full and effective deployment of this instrument, as negotiations on the exceptions for Visually Impaired Persons/Persons with Print Disabilities revealed in July 2012 (Love, 2012a). But even with an optimistic estimation, a workable exception might not reach the stage of an international instrument before 2015 (Love, 2012b), and should then be transposed into national legislation to be enforceable.

2.2. The technical accessibility of digital heritage

In addition to public regulation through national laws deriving from European directives or international treaties, digital heritage may be managed by private governance. Terms of use and end-user license agreements are unilateral contracts imposed by platforms, applications, services and websites to their users. These sets of conditions are fragmenting web and non-web content, to the point of excluding contributors from reusing their own data that they created in another platform.

“Walled-garden” centralized services, such as Facebook, maintain a large amount of digital data hidden and impossible to export (Dulong de Rosnay, 2011a). A centralised technical architecture will favour the implementation of centralised corporate governance, with a high level of legal excludability of users from access and usufruct of their own creation. It will be even more the case for external parties, such as libraries, who would need to crawl and archive this digital information for data mining or preservation purposes. Hidden technical restrictions, such as registration, create barriers to automated back-up and preservation, and will prevent access by memory and archive institutions in the same way in which the chilling effect of technical protection measures on archival and other legitimate uses has been denounced by scholars and library advocacy groups (Coyle, 2004). The question of acceptable formats, which should be sustainable and interoperable, and the role of technical open formats, has been underlined by libraries while producing recommendations (Library and Archive Canada, 2012).
Stand-alone objects, which can be downloaded from a webpage, are not exempted from technical hidden barriers to effective preservation. For instance, the preservation of a Net Art game, or installation, will require the compliance with instructions that should also be available, and the preservation of a videogame will require emulation (Esposito, 2011). In the same vein, documentation is necessary to install and use a software programme. To that aim, the developers of free software licenses also created a Free Documentation Licence (Free Software Foundation, 2008) to ensure the “effective freedom to copy and redistribute”.

2.3. Policy recommendations

Technical obsolescence and “the lack of supportive [sic] legislation” have been noted by Unesco (2003) as preventing the archival of digital heritage to an even greater extent than physical heritage or digitized heritage because no physical copy, albeit degradable and non-movable, is available for posterity. The quasi-epistemological distinction, stated by copyright, between access and copy must be abandoned, and several other legal categories need to be urgently revisited. In the realm of copyright, legal deposit and exceptions must be extended to reflect the need of preserving an increasingly complex digital heritage, and making it available to educators, researchers, innovators and the public in general. A realistically applicable legislation related to orphan works must be adopted, in order to cope with works whose right-holders cannot be identified; it should not only focus on public cultural preservation institutions but also include individuals and non-profit initiatives such as Wikipedia (Communia, 2011).

Consumers need to be protected from unfair terms of use imposed by platforms, and not only from commercial, centralized services. Libraries themselves can also be responsible for the partial enclosure of common heritage, as their own terms of use sometimes restrict access to or reuse of the digital version of public domain works they curate (Dulong de Rosnay, 2011b).

The relevant digital information may be difficult to single out in the large amount of available data, in the absence of a search result. The appropriateness of this result may heavily rely on metadata, whose accessibility is then crucial. The development of open policies concerning their dissemination to the public for reuse will contribute to remove obstacles for further reuse by preservation actors; these may themselves, subsequently, act as intermediaries between digital heritage and developers of future applications for the reuse of our common heritage. For instance, under the Europeana Data Exchange Agreement (Europeana, 2011), data providers accept to release their metadata under a Creative Commons CC0 Public Domain Dedication, meaning “that all metadata provided to Europeana can be re-used by third parties without any restrictions.”

International efforts are urgently required in order to avoid a vacuum in the preservation of currently-developing digital heritage. A combination of legal and
policy measures can prevent the establishment of an important gap in the preservation of a generation’s historical digital traces. However, this will only happen through the combined efforts of creators, in order to release their creations in accessible formats; of non-profit and for-profit institutions, to guarantee access and reuse; and of publishing industry lobbyists and states to accept a balance between copyright protection and the preservation of the general interest.

Conclusions

Different disciplines and fields of study seem to be heralding the rise of an interdisciplinary scientific and intellectual movement focused on digital heritage. In order to contribute to a reflection on the very foundations of this movement, this article has sought to address some of the methodological challenges that the scholarship on digital heritage is currently facing, in order to analyze this interdisciplinary phenomenon in its plurality.

Section 1 has discussed how a preliminary condition to the understanding of the issues relevant to digital heritage is a further exploration and detailing of a typology of the relevant stakeholders in its preservation and promotion, the material sources from which they draw – and from which scholars may draw, in turn. A comprehensive inventory of what is to be included in the perimeter of digital heritage as a subject of inquiry should have as a precondition a reflection on sources, and on the different collection methods and tools that vary not only according to the diversity of objects that can possibly be included, but also to apprehend the same object. Finally, the section has discussed how a relevant part of digital heritage cannot be analytically addressed without taking into account the technical infrastructures and the human environment that underlie its emergence – without reintroducing the material in the comprehension of “intangible,” “virtual” heritage.

Section 2 has proposed to overtake the limitations imposed by applicable law, which are challenging the constitution of digital heritage collections. After demonstrating how negotiations of rights can lead to a deadlock, all the more with complex, intangible, born-digital objects, several paths of reform are proposed in order to enable the development of preservation initiatives. The development of strong exceptions benefiting to libraries and other preservation initiatives, flexible schemes to manage orphan works, balanced terms of use and the use of open formats and open licensing schemes to release accompanying data, are crucially needed in order to remove obstacles to the constitution of usable archives for future generations.

Acknowledgments

This article draws on prior conversations and joint work with Valérie Schafer, Institute for Communication Sciences, CNRS, and Fanny Georges, Université Sorbonne Nouvelle-Paris III, Paris, France.
References


