

# **DNA in the Archive**

Material and Metaphor of a New Data Storage Format

by

Madeleine Claire Mendell

A thesis submitted in partial fulfillment

of the requirements for the degree of

Master of Arts

Moving Image Archiving and Preservation

Department of Cinema Studies

New York University

22 July 2021

Advisor: Mél Hogan

Deposit Copy

## **Table of Contents**

<i>Abstract</i>	2
<i>List of Figures</i>	3
<i>List of Tables</i>	3
<i>Acknowledgements</i>	4
<i>Designated Community</i>	5
<b>Introduction</b>	7
<b>Chapter One: How DNA Works</b>	15
<i>The Technology of DNA-Based Data Storage</i>	18
<i>Why DNA-Based Data Storage?</i>	41
<i>The State of the Field</i>	45
<b>Chapter Two: DNA in the Archive</b>	47
<i>“Care and Feeding”</i>	53
<i>Archival Requirements</i>	59
<i>Containers</i>	63
<b>Chapter Three: A Case Study for Audiovisual Archives</b>	70
<i>Cathy Freeman and Australia’s DNA</i>	72
<i>Indigenous DNA, Reconciliation, and Repatriation</i>	81
<i>Indigenous Collections and the Protocols</i>	90
<b>Conclusion</b>	98
<b>Works Cited</b>	101

## Abstract

As the amount of digital information in the world increases exponentially and current data storage formats deplete the Earth of memory-grade raw materials, DNA has increasingly become a contender as a viable solution for offline digital data storage. Marketed as a cleaner, denser, and longer lasting format, DNA will usher biotechnology and bioethics into the field of data storage and by proxy digital preservation. Memory institutions must confront the implications this new format will have on our relationships to the living world, to our collections, and to our data. Moreover, as DNA-based data storage companies and consortia see archival storage as the format's first use case, archivists are uniquely positioned to imbue archival standards, ethics, and sensibilities into this new format before its standardization.

This thesis explores the technical specifications of DNA-based data storage, describes the necessary advances for DNA to become an appropriate and appealing preservation format, and surveys the field to identify areas where archival knowledge could be crucial to the development of this new format. This thesis also grapples with the ways in which DNA's perpetuation as 'the code of life' in the wake of the Human Genome Project and in the public imaginary affects institutional preservation strategies and reveals underlying ideologies at work in the archive. The final chapter's case study of the Australian National Film and Sound Archive's DNA-based data storage pilot project dissects the slippage of material quality with metaphoric language at play. DNA makes readily apparent that the archive is not a metaphor, that its technological materialities must be taken seriously in order to contend with the racial and colonial projects that have reappeared as genomics and cultural heritage institutions collide at the site of data storage.

**List of Figures**

Figure 1. A DNA-based data storage life-cycle.

Figure 2. Different methods of encoding binary data onto DNA (Ping, et al.).

Figure 3. Column-based oligonucleotide synthesis and microarray-based oligonucleotide synthesis (Hughes and Ellington).

Figure 4. The Open Archival Information Systems (OAIS) Reference Model (CCSDS).

Figure 5. FRBR Group 1 Entities (IFLA Study Group on the Functional Requirements for Bibliographic Records).

**List of Tables**

Table 1. Summary of sequencing platforms as of 2018 (Ping, et al.).

## Acknowledgements

This work would not exist without the patience, support, and encouragement of my professors, internship supervisors, classmates, friends, family, community. Thank you.

I am most grateful to my thesis advisor, Mél Hogan, whose own work is an inspiration, whose feedback has been invaluable, and whose support has been overwhelming.

Additionally:

To my 2021 MIAP cohort – we’ve been through it! – David Griess, Robert Hoffman, Shiyang Jiang, Ryan Lakin, Aparna Subramanian, Klavier Wong, Zoe Yang. <3

For their support: the MIAP faculty, administration, and lab.

For their time: Linda Tadic and Nick Gold.

For reading and editing my work: Madeleine Collier, Claire Fox, and the University of Calgary Grad Chat; for mandating easter eggs: Dave Rice. :) )

For showing me an academic community: the Environmental Media Lab (in particular, Crystal Chokshi), Elaine Gan, the Elements of Media seminar, and especially Nicole Starosielski and Nicole Martin, who were the first to encourage me in pursuing this research.

For feeding me: Kate, Zoë, Cody, Mark; for keeping me warm, MT. (=^.^=)

In solidarity with the Union for Graduate Employees at New York University (GSOC-UAW), my comrades and co-workers, who have won a fair contract after a successful, month-long strike.

Lastly, thank you to my parents, Randee Trabitza and Henry Mendell, for bearing with me.

## Designated Community

Borrowing terminology from the Open Archival Information System reference model,<sup>1</sup> here, I define my Designated Community, or the various persons who “should be able to understand [this] particular set of information.”<sup>2</sup> Not only do many view long term archiving as the most viable use case for DNA-based data storage, but many organizations, public and private, are pursuing cultural heritage institutions and archives for investment, pilot projects, and partnerships. At this early point in the format’s development, the necessarily interdisciplinary nature of DNA-based data storage signifies that we, archivists, have the unique opportunity to position ourselves as primary collaborators in setting and defining standards and uses. This inclusion from this stage means that archivists can influence the technological makeup of this new medium as well as advocate for the incorporation of archival ethics and standards that may not be considered by other stakeholders.

For these reasons, the designated community for this thesis project primarily consists of archivists and practitioners who may be involved in setting standards for DNA-based data storage or who may, in the near future, come into contact with the format in their work of maintaining collections. As the development of DNA-based data storage encourages transdisciplinarity and interdisciplinary collaboration from the technical perspective, I include those working on bioethics and postgenomics as well as on the implications any use of DNA – living or synthetic – has on constructions of race, gender, policy, and yes, archives. Positioning archival science as component of science and technology studies (STS) necessitates the

---

<sup>1</sup> The idea of using the OAIS Designated Community as an Audience statement comes from Claire Fox’s 2020 MIAP thesis “Not Normalized: Born-Digital Camera Original Video Formats in the Archives.” Doing so bridges our pragmatic, archival work with the academic labor of a thesis project.

<sup>2</sup> *Reference Model for an Open Archival Information System (OAIS)* (Washington, DC: The Consultative Committee for Space Data Systems (CCSDS), June 2012), <https://public.ccsds.org/pubs/650x0m2.pdf>, 1-11.

simultaneous inclusion of archivists in the theoretical realm of the academy and vice versa. By including discussions of the use of DNA as a biological component and ‘code for life’ as well as the applications of DNA-based data storage, I aim to contribute to a conversation influenced by a multitude of perspectives such that this inchoate format can be handled with appropriate care and attention.

## Introduction

In 1988, artist and scientist Joe Davis created the scientific artwork *Microvenus*. Davis, along with molecular geneticists and scientists at the Massachusetts Institute of Technology (MIT), converted a line drawing into a DNA sequence and placed it into an *Escherichia coli* bacterium. By translating the binary code representing the image into the nucleotide components of DNA – represented by the letters ATCG – they were able to insert it, like a virus, into the bacterium’s own genetic material. The string in question encoded a graphical representation of “an ancient Germanic rune...used to represent *life* and the female *earth*.”<sup>3</sup> A Y-shaped diagram with a yonic line extended through the center of the fork, this graphic “can also be taken as a representation of the female human genitalia.”<sup>4</sup> With *Microvenus*, Davis became the first to paint his own artificially created human origin story of life and reproduction onto this new canvas of biological, genetic material. Two years later, Davis successfully retrieved the rune from the bacterium with a lab at Harvard, proving the storage of digital data onto DNA was possible.

Writing about the project eight years later, Davis states that *Microvenus* was conceptualized first as a response to the messages sent into the unknown – to extraterrestrial life, perhaps – aboard the NASA *Voyager* and *Pioneer* spacecraft. Davis asserts that biological storage would persist further into the future than any tape or golden disc. With respect to the content of *Microvenus*, he chose the rune because of the *Voyager* Golden Record’s omission of female genitalia in the depictions of *Homo sapiens* due to censorship. Moreover, says Davis, sending messages to extraterrestrials is necessarily an ontological inquiry into humanity itself. He argues his “Venus figurine,” representing life and sex at once, is constitutive of this search

---

<sup>3</sup> Joe Davis, “Microvenus,” *Art Journal* 55, no. 1 (Spring 1996): 70, [www.jstor.org/stable/777811](http://www.jstor.org/stable/777811).

<sup>4</sup> Davis, “Microvenus,” 70.



for what it is to be human, and to be alive. Davis explicitly embeds DNA with an aesthetic of “reproductive biology [that] is also reflected in various romantic concepts about beauty and nature that have been securely attached to the discussion of living things.”<sup>5</sup> Davis perpetuates this aesthetic by participating in the long tradition of male artists depicting nude female bodies, and he does it in a 35-bit string.<sup>6</sup> Davis’ depiction of life is represented doubly: by the symbol of female genitalia and by the DNA strand on which the image resides. The ethos of *Microvenus*, expressed in the metaphors of life, its gendered metonymic representation in the rune, and in the so-called language of life itself, continues to pervade DNA discourse today.

I begin this thesis with a discussion of *Microvenus* because an analysis of the artwork opens up the discourse of DNA-based data storage beyond mere description of technical specifications or comparisons to existing storage formats. Fulfilling those goals is necessary to understanding how the format operates, but a work like *Microvenus* implores us to extend an inquiry beyond those questions alone. The technological issues Davis describes in his documentation of *Microvenus* – especially the encoding scheme and the attention to DNA’s physical and biological structure and limitations – continue to be raised thirty years later as DNA-based data storage technology becomes more advanced. Similarly, his enthusiasm for DNA as a resilient storage format with a much better chance of reaching extraterrestrial life than the gold-plated records aboard the NASA *Voyager* probes continues to inform the industry’s view of long term archiving as the format’s primary use case. Moreover it perpetuates that false conflation between long term archiving and the jettison of cultural heritage materials into space.

---

<sup>5</sup> Davis, “Microvenus,” 74.

<sup>6</sup> See Guerrilla Girls’ ongoing “Do women have to be naked to get into the Met. Museum?” billboards and pamphlets, first published a year after Davis’ experiment in 1989: <https://www.tate.org.uk/art/artworks/guerrilla-girls-do-women-have-to-be-naked-to-get-into-the-met-museum-p78793>.

As an aside, while a proof-of-concept for DNA-based data storage, *Microvenus* itself does not seem to have been archived beyond its paper documentation. *Microvenus* is certainly not the last DNA strand to enter an arts institution – recently, artist Lynn Hershman Leeson exhibited her own synthesized DNA strand at a retrospective of her career held at the New Museum in New York City.<sup>7</sup> Lastly, Davis’ reliance on the imagery of life, the metaphor of reproductive origins, and explanation of DNA as a “language” that in turn makes the narrative of data storage compelling persist in today’s descriptions of the technology. These three areas of inquiry have guided the research for this thesis project; that they are so present in the first endeavor to store data onto DNA only makes exploring them that much more urgent.

*Microvenus* is an origin story for DNA-based data storage, not only in its technical achievement but also in its vision for the future of data storage and in its expression of humanity and life. Like other technological firsts, it will receive consistent attention as DNA-based data storage endures. Consider how film scholars repeatedly turn to Eadweard Muybridge’s pioneering horse photography as a form of early cinema in order to analyze the medium.<sup>8</sup> While there is certainly a critique of dwelling on the “firsts” of technology,<sup>9</sup> *Microvenus*, like its fellow firsts, offers in that short string of numbers much to mine from, to analyze, and to discuss. As a template for the thesis work that follows, analysis of *Microvenus* reveals, at the bit and nucleotide levels of information, how using the technology of DNA encodes ideologies of gender, humanity, exploration, and aesthetics. Throughout this work, I will argue that archivists

---

<sup>7</sup> Jillian Steinhauer, “Lynn Hershman Leeson: The Artist Is Prescient,” *New York Times*, 8 July 2021: <https://www.nytimes.com/2021/07/08/arts/design/hershman-leeson-review-art-museum.html>.

<sup>8</sup> Coincidentally, these images were stored onto an *E. coli* bacteria colony as a moving image in 2017: <https://news.harvard.edu/gazette/story/2017/07/taking-cells-out-to-the-movies-with-new-crispr-technology>.

<sup>9</sup> Max Liboiron, writing for the Environmental Media Lab’s Heliotrope journal in 2021, gives, in their “Firsting in Research,” a substantial critique of this phenomenon in the field of research and the academy: <https://www.environmentalmedialab.com/heliotrope/firsting-in-research>.

are uniquely positioned to investigate what DNA does as a technology. Science and technology studies (STS) scholar Sheila Jasanoff's *co-production* – the analysis of how “science and society, in a word, are *co-produced*, each underwriting the other's existence” – offers a useful framework to situate what an archival intervention will look like in exploring DNA-based data storage.<sup>10</sup> A *co-productionist* framework does not deny “the reality or the power of science,” but instead facilitates “explaining why the products of science and technology acquire such deep holds on people's normative instincts as well as their cognitive faculties.”<sup>11</sup> Library sciences is a field in which the convergence of making identities, institutions, representations, and discourse – the four “most common instruments of co-production [that] operate at the nexus of natural and social order” and “stabilize both what we know and how we know it” – with technological artifacts appears through administration, access, presentation, and preservation.<sup>12</sup> By relocating it from STS to archival practice, I aim to use this co-productionist framework as a methodology for understanding how the ideologies manifest in *Microvenus* reappear and reconfigure themselves at a macroscale as cultural heritage institutions consider DNA-based data storage.

Davis' idea of an archive – one rooted in media sent into space never to be seen again by humans – is not representative of archives on Earth. However, this reductionist view of archival storage reappears as those proffering DNA-based data storage see the technology catering to large archives with long term cold storage needs. I see this development as an opportunity for archivists to get involved in the format at the earliest stage. Where film, tape, and most digital media were first intended for a production environment, DNA is a format whose intended primary use is archival. If so, archivists have a responsibility to ensure that the standards set in

---

<sup>10</sup> Sheila Jasanoff, *States of Knowledge: The Co-Production of Science and Social Order* (New York: Routledge, 2006), 17.

<sup>11</sup> Jasanoff, *States of Knowledge*, 38.

<sup>12</sup> Jasanoff, *States of Knowledge*, 39.

place by these consortia of companies, academic institutions, and government organizations cater to archives. The development of DNA-based data storage will necessarily be transdisciplinary and interdisciplinary; it will involve biomolecular engineers, molecular geneticists, computer scientists, information technologists, and archivists. Archivists have the opportunity to influence this format with the ethics, standards, and best practices that continue to inform the field and ensure that those sensibilities make their way into the fabric of this new technology.

In this thesis, I argue that archivists must advocate for their collections, methodologies, users, and staff in the development of DNA-based data storage. In order to do so, especially as DNA-based data storage departs from the understood lineage of storage media found in film, tape, and digital, I use this space to insist upon including perspectives on DNA that come from science and technology studies, race and ethnicity studies, and gender and sexuality studies. Drawing on the work of scholars in these fields will help archivists to explore how these various epistemologies of DNA are co-produced alongside biological understandings of DNA, genetics, and genomics when we begin to store cultural heritage onto DNA. Additionally, archivists must not only consider the data density of DNA or any other technological novelty; we must also investigate what DNA's technological qualities are, aesthetic and programmatic. What does a format do to its contents? How can we continue to care and maintain for information stored in this new container? How will DNA change our preservation methodologies and our perceptions of our collections? In considering that DNA-based data storage will become a viable option for long term preservation and will soon be adopted by cultural heritage institutions, this thesis offers an analysis of the new format through an archival lens. I hope to encourage archivists to think about the effects any format has on the objects it stores, from technical integrity to the way we perform access, to how we describe objects in light of their containers.

A discussion of DNA-based data storage reveals the many different social structures and legacies at play within the new format. Archivists are those who care for objects in order to maintain their accessibility into the future; our decisions – how we describe, how we manage, and how we provide access for those objects under our care – affect how those objects persist. DNA-based data storage, while proven in concept, has not yet coalesced into a standardized format. Still in development and having designated its destined use for archives, DNA-based data storage has yet to “accrue phenomenological and aesthetic value” reserved for commercially available and widely used formats.<sup>13</sup> As Jonathan Sterne writes of MP3s and formats in general, “A characteristic that might first appear as the result of numb technological imperatives is actually revealed as something that had an aesthetic and cultural formation.”<sup>14</sup> Following Joe Davis’ assertion that DNA-based data storage has an aesthetic, I use this thesis to discuss how archivists maintain not only aesthetic objects in storage but also an aesthetic in which they reside. I posit that DNA’s aesthetic and cultural formation, co-produced with scientific and social perception affects, just as the MP3 format does to the audio it compresses, the data it will store.

*Chapter One* lays a foundation for archivists interested in DNA-based data storage. Beginning with a short history of this format in the making, I offer a straightforward, introductory description of how the new technology works, what makes DNA-based data storage appealing (or not), and the various developments required to make it a viable option for archival storage. This chapter considers the context of increasing data storage needs, exponentially expanding amounts of data, and the growing lack of memory-grade rare earth metals and sand that are used for current data storage formats. I conduct a brief survey of the field in order to

---

<sup>13</sup> Pierre Bourdieu, quoted in Jonathan Sterne, *MP3: The Meaning of a Format* (Durham: Duke University Press, 2012), 15.

<sup>14</sup> Sterne, *MP3*, 15.

identify main sectors where DNA-based data storage is gaining attention. Identifying the backgrounds certain companies have – in biotechnology industries like sequencing and synthesis or in information technology and computing – contextualizes the development of DNA-based data storage. I point to ongoing workshops and consortia where discussions of developing a standard take place to reveal the key issues different stakeholders have identified for the format’s development. Of particular interest are issues of bioethics, sustainability, and the physically specific limitations DNA has that affect encoding schemes. Lastly, this chapter examines the necessary developments in DNA-based data storage that will make the new format a competitor for current industry standard long term offline storage, namely Linear Tape-Open (LTO).

Using the Open Archival Information System (OAIS) reference model, the *second chapter* considers how DNA-based data storage could be integrated into archival storage. Importantly, I use the OAIS descriptions of archival storage that dictate the necessary work of providing access and maintaining integrity in order to ask how DNA-based data storage will account for the work of preservationists. Incorporating the work of STS scholars alongside archival ethics, I primarily refer to Zoë Sofia’s work on container technologies in order to examine how archives can use existing practices to reconcile the changes DNA-based data storage will have on containing collections. This chapter weaves together three claims: that DNA has been coded as informative, linguistic, and as “the code for life” through metaphors will affect our perceptions of actual data stored onto DNA; that container technologies holding objects – namely cultural heritage – commingle with those materials; and that archives, especially national settler-state archives which have been custodians of cultural heritage, must engage with the socio-political consequences that the language of DNA has engendered as genetics and genomics

have gained powerful cultural capital. These discussions come together in order to place archives and archives staff in a position of power and responsibility to the technology we use to preserve.

The *third chapter* continues this discussion of archival responsibility to the ways our technologies interact with our collections through a lengthy analysis of a case study: the recent 2020 pilot project by the National Film and Sound Archive of Australia (NFSA) and Olympic Foundation for Culture and Heritage (OFCH). This project explored the use of DNA-based data storage to store footage of Aboriginal-Australian athlete Cathy Freeman's gold medal win at the 2000 Sydney Olympics. I discuss the NFSA's own curatorial and preservation policies as well as corporate strategies alongside an in-depth critique of the way Cathy Freeman's legacy has been constructed in light of the contemporary Aboriginal-Australian movement for reconciliation. Relying on work by Kim TallBear (2013) and Reardon (2012) amongst others that discuss the co-option of Indigenous DNA by scientific researchers as a racial project, I situate the NFSA's embrace of DNA through Freeman within a lineage of archives and scientific research institutions using Indigenous genetic material and imagery in order to reaffirm colonial power structures. This chapter, while offering a critique of the case study, sees understanding these implications as a great opportunity for archival intervention. Influenced heavily by the work of the First Archivist's Circle and the various Indigenous Protocols for handling Indigenous materials at non-Native archives, I argue that these protocols offer tools archivists can use to work through the issues that DNA-based data storage will raise for our collections.

## Chapter One: How DNA Works

The idea for storing information onto deoxyribonucleic acid (DNA) came soon after scientists had revealed the double helical structure of DNA and confirmed the macromolecule's role in genetics in 1953.<sup>15</sup> In the mid-1960s, Russian scientist Mikhail Samoilovich Neiman wrote several papers describing how DNA's retention of hereditary information functioned analogously to memory storage devices containing numeric data.<sup>16</sup> Scientist Katrin Weigmann attributes the association made between genetic material and information to physicist Erwin Schrödinger, who "probably coined the term 'code' when he described living organisms in terms of their molecular and atomic structure" in 1944.<sup>17</sup> Invoking code, and therefore malleable information, influenced the way John Watson and Francis Crick wrote about DNA as they described the structure ten years later, and has continued to inform the way we speak about DNA.<sup>18</sup> While this connection may seem obvious today, it is important to acknowledge the ways in which scientific language becomes part of the popular vernacular. Moreover, this development demonstrates the ways the language of computer science, physics, and mathematics have always been embedded in our understandings of DNA. These associations paved the way for Neiman's observations, and later for Davis' *Microvenus*.

---

<sup>15</sup> Namely Francis Crick, Maurice Wilkins, Rosalind Franklin, and James Watson (whose racist and sexist beliefs the United States National Institute of Health denounced as "unsupported by science" and therefore counter to their mission; that "science" can and cannot support socio-political structures itself points to the "scientific" construction of such systems): <https://www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet>.

<sup>16</sup> Mikhail Samoilovich Neiman, "Some fundamental issues of microminiaturization," *Radiotekhnika* 19, no.1 (1964), [https://2a008ed5-a-62cb3a1a-s-sites.googlegroups.com/site/msneiman1905/Neiman-1964\\_Micromini.pdf](https://2a008ed5-a-62cb3a1a-s-sites.googlegroups.com/site/msneiman1905/Neiman-1964_Micromini.pdf).

<sup>17</sup> Katrin Weigmann, "The code, the text and the language of God," *European Molecular Biology Organization Reports* 5, no. 2 (2004): 116.

<sup>18</sup> Weigmann, "The code, the text and the language of God," 116.



Davis' 1988 experiment showed that the storage of digital information could propagate in a living cell, but the use of DNA as a data storage format on a large scale would not be made possible without the ensuing advances made in synthesizing and sequencing technologies over the next thirty years. These advances, which lowered the cost of reading and writing DNA as well as increased speed and capacity, were spurred on by the Human Genome Project (HGP) during the 1990s and with its conclusion two years ahead of schedule in 2003.<sup>19</sup> Since then, the fields of genomics and genetics have continued to expand into a growing, competitive industry. At the time of the Human Genome Project's completion, it was seen as a great success story. Bill Clinton said of the achievement:

Today, we are learning the language in which God created life. With this profound new knowledge, humankind is on the verge of gaining immense, new power to heal.<sup>20</sup>

While there may be some debate as to the success of the Human Genome Project in how much the research has led to these bold claims, the HGP catapulted the sequencing industry from the niche and expensive to the commercial and cheap. The field of genomics and genetics has continued to expand such that Neiman's and Davis' theoretical musings are becoming plausible realities at a large scale.

The field of postgenomics has attempted to reconcile this metaphor of code with ongoing research that reveals how complex these processes are. Of the way this language perpetuates in actual genetic research, Eva M. Neumann-Held and Christoph Rehmann-Sutter write: "The more deeply scientists have searched for the mechanisms by which individual genes control development, and the more details they have learned about the functioning of *genetic programs*

---

<sup>19</sup> National Human Genome Research Institute, "The Cost of Sequencing a Human Genome," accessed 18 April 2021, <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.

<sup>20</sup> "President Clinton: Announcing the Completion of the First Survey of the Entire Human Genome," The White House at Work, 26 June 2000, <https://clintonwhitehouse3.archives.gov/WH/Work/062600.html>.

and *genetic blueprints*, the less plausible the metaphors of genetic control, programs, and blueprints seem.”<sup>21</sup> Neumann-Held and Rehmann-Sutter are challenging the “molecular paradigm,” which “explains wholes through molecular parts and which often identifies DNA as the central determinant.”<sup>22</sup> In the era after the Human Genome Project, which was often referenced as the first step in deciphering that master code and therefore reaffirming its import, the task of interpreting the results of the HGP have not gained the same level of fame as its predecessor. All this is to say that when the language of code appears as computer scientists attempt to work with DNA as if it were, we should be cognizant of the ways in which further research has undermined or complicated that claim as well as where such claims emerged. As a new group of people approaching genetic research, archivists should be able to understand how a technology works as well as its history and how it came to be a technology.

In the years since Davis’ one-off experiment in inserting digital information onto a cell’s DNA, numerous pilot projects have expanded efforts in the storage of digital media onto DNA. As recent as 2020, Eadweard Muybridge’s horse photography, excerpts from “It’s a Small World,” biblical text, the Svalbard Global Seed Vault database, Shakespearean sonnets, and a variety of other image, text, software, audio, and video data have been encoded onto DNA.<sup>23</sup> Moreover the size of files encoded has increased from *Microvenus*’ 35 bits in 1988 to over 400 MB in 2018 and is only increasing.<sup>24</sup> These projects, for the most part endeavors initiated by private companies, universities, and government institutions as proofs-of-concept, represent unstandardized experiments in the encoding, synthesizing, and sequencing of DNA. My third

---

<sup>21</sup> Eva M. Neumann-Held and Christoph Rehmann-Sutter, “Introduction,” in *Genes in Development: Re-Reading the Molecular Paradigm* (Durham: Duke University Press, 2006), 4.

<sup>22</sup> Neumann-Held and Rehmann-Sutter, 3.

<sup>23</sup> Potomac Institute for Policy Studies, “The Future of DNA Data Storage,” September 2018, 12, [https://potomac institute.org/images/studies/Future\\_of\\_DNA\\_Data\\_Storage.pdf](https://potomac institute.org/images/studies/Future_of_DNA_Data_Storage.pdf).

<sup>24</sup> Potomac Institute for Policy Studies, “The Future of DNA Data Storage,” 12.

chapter examines one such archival case study from the National Film and Sound Archive of Australia and Twist Bioscience. However, the storage of audiovisual media and cultural heritage objects as experiments demonstrates the industry's interest in archives as a primary use for this technology. Within this context, this chapter discusses how DNA data storage works, the rudimentary application of DNA as an offline storage format, and the technical qualities of DNA that make it appealing as such.

### **The Technology of DNA-Based Data Storage**

Audiovisual archivists have had to work with diverse technologies in order to accomplish the task of preservation. From the chemistry of film to the electromagnetic physics of video tape to computer science in digital preservation, archivists have had to engage with a variety of different sciences, archivists have had to engage with a variety of different sciences in order to perform our jobs. Familiar with these fields, when DNA becomes an option for archivists, we will have to understand its technology as intimately as we do the path of an electron beam on a cathode-ray tube monitor. In order to make the language of DNA-based data storage accessible to as wide a variety of people and archivists possible, this section offers some basic descriptions of the biology of DNA. More specifically, I outline the way in which DNA's biological qualities have been harnessed by the DNA-based data storage technology; in order to do so, I offer a necessary remedial lesson in biology for archivists. Subsequently, I dive into the technology of DNA-based with further specificity.<sup>25</sup>

---

<sup>25</sup> As an aside, learning about DNA-based data storage has been a steep learning curve. Prior to this research, I had a high school level understanding of biology. Reading many scientific articles to learn how DNA-based data storage works has been daunting. However, in the case that archives employ DNA, archivists must have at least a basic understanding of the biology at play, if not a deeper understanding of the storage format's architecture. This chapter represents my best efforts in doing so.

DNA is a macromolecule that consists of individual parts called nucleotides. Each nucleotide has three main components: a phosphate group, a sugar group, and a nitrogen base.<sup>26</sup> There are four naturally occurring nitrogen bases that determine if the nucleotide is adenine (A), thymine (T), guanine (G), or cytosine (C). These nucleotides bind with each other (A to T, G to C, and vice versa) in order to form a sequence in the double helical shape that makes DNA such a stable molecule. In living cells, DNA is what the National Institute for Health (NIH) describes as a set of “biological instructions.”<sup>27</sup> In replication, a protein called DNA polymerase, facilitates the ‘write’ of DNA by matching each nucleotide to its component, forming the double helix.<sup>28</sup> In order to express a gene, or an individual set of those ‘instructions,’ the DNA must be ‘transcribed’ to RNA, a single-stranded, smaller segment of DNA. Transcription is facilitated by an enzyme called transcriptase. Once transcribed, the messenger RNA (mRNA) strand<sup>29</sup> will be used to synthesize proteins as each set of three nucleotides forms an amino acid, of which there are at least 20 variants. This process is called translation – or the translation of genes to actionable proteins. These amino acids form various proteins that continue to carry out functions within cells. This fundamental biology process describes the way DNA functions within our bodies and within the bodies of others, from eukaryotic bacteria to plant to animal cells.

### *In Vivo and In Vitro*

This basic description – of DNA as a master code that informs the subprocesses of transcription and translation into these active proteins – is common in biology textbooks and basic descriptions. It is also a description that deeply informs the way we think about the storage

---

<sup>26</sup> National Human Genome Research Institute, “Deoxyribonucleic Acid (DNA) Fact Sheet,” accessed 17 April 2021, <https://www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet>.

<sup>27</sup> National Human Genome Research Institute, “Deoxyribonucleic Acid (DNA) Fact Sheet.”

<sup>28</sup> Pasadena Community College, *General Chemistry and Chemical Analysis*, Chemistry LibreTexts, 10 June 2020, 1197, <https://batch.libretexts.org/print/Finished/chem-219108/Full.pdf>.

<sup>29</sup> As an example, I use mRNA, but there exist also rRNA and tRNA molecules that have other functions.

of digital information onto DNA. In rudimentary terms, as nucleotides function similarly to binary data in that they code for something else and are essentially a string of data, so could biological methods be used to insert or create sequences of nucleotides that would code for digital information instead of biological information. There are currently two schemes for DNA-based data storage. The first, which is what *Microvenus* employed and what many other experiments have used, is *in vivo* storage. This method relies on a living cellular substrate, such as the *E. coli* bacteria. Until 2012, with experiments at the Wyss Institute at Harvard University, only one DNA-based data storage project did not rely on an *in vivo* step or storage substrate.<sup>30</sup> Recently, *in vivo* data storage projects have used a combination of synthetic DNA to increase data density while using these living cells as substrates. Many of the papers advocating for more research into using living cells identify using these bacteria or spores' natural physical resistance to heat or radiation as a great advantage.<sup>31</sup> Additionally, many recent *in vivo* projects aim to take advantage of the recently discovered CRISPR-Cas system. This system uses a variety of Cas proteins – namely Cas9, Cas1, and Cas2 – to make small edits in a sequence that are most applicable for recording and tracking changes in the DNA over time.<sup>32</sup> However, many of these applications are not yet scalable, and the mutation rate in living cells has been a physical and mental barrier many in the data storage business have yet to overcome.

While most of the focus of industry reports have focused on *in vivo*'s counterpart, *in vitro*, for reasons I address below, papers researching *in vivo* storage continue to argue that *in vivo* storage will continue to show promise with more attention. As Bhattarai-Kline, Lear, and

---

<sup>30</sup> Luis Ceze, Jeff Nivala, and Karin Strauss, "Molecular digital data storage using DNA," *Nature Review Genetics* 20 (2019): 457, <https://doi.org/10.1038/s41576-019-0125-3>.

<sup>31</sup> Sung Sun Yim, Ross M. McBee, Alan M. Song, Yiming Huang, Ravi U. Sheth, and Harris H. Wang, "Robust direct digital-to-biological data storage in living cells," *Nature Chemical Biology* 17, (11 January 2021): 246, <https://doi.org/10.1038/s41589-020-00711-4>.

<sup>32</sup> Ceze, Nivala, and Strauss, "Molecular digital data storage using DNA," 465.

Shipman write in their analysis of the Yim (2021) paper’s research on direct digital-to-biological data storage, “Unlike the semiconductors and magnetic media that we currently use for most data storage, DNA data storage will likely benefit from other genetic research that drives technologies forward at a pace faster than would be expected for a less universally relevant media.”<sup>33</sup> For that reason, while this paper focuses more specifically on in vitro technology, it is important that we acknowledge past and concurrent research in in vivo storage, which will undoubtedly have repercussions on synthetic technology. On that same note, in their comparison of synthetic and in vivo data storage applications, Ceze et al. state that “As the field of synthetic biology continues to mature, in vivo data storage may yet provide answers to lingering drawbacks of in vitro storage methods.”<sup>34</sup> Additionally, many in vivo projects rely heavily on synthesis technology for living cell substrates as well. Perhaps for these reasons, as well as the relative instability and lower data density of in vivo data storage, in vivo data storage has not been given the same attention as its synthetic counterpart.<sup>35</sup> More probably, many of the commercial companies branching out into DNA-based data storage have emerged from the DNA synthesis industry and are using DNA-based data storage as an extension of that business.<sup>36</sup> For that reason, primarily, this thesis work focuses on in vitro DNA-based data storage.

In vitro DNA-based data storage is the synthesis of DNA outside of a living cell. This method of storage rose mostly out of the DNA synthesis industry, which has grown in the post-HGP era as increased interest in genetic research testing has required the need for mass-produced synthetic sequences of DNA. Synthetic DNA, like in vivo DNA, can also undergo biological

---

<sup>33</sup> Santi Bhattarai-Kline, Sierra K. Lear, and Seth L. Shipman, “One-step data storage in cellular DNA,” *Nature Chemical Biology* 17 (26 January 2021): 233, <https://doi.org/10.1038/s41589-021-00737-2>.

<sup>34</sup> Ceze, Nivala, and Strauss, “Molecular digital data storage using DNA,” 465.

<sup>35</sup> Ceze, Nivala, and Strauss, “Molecular digital data storage using DNA,” 457.

<sup>36</sup> Potomac Institute for Policy Studies, “The Future of DNA Data Storage,” 16.

processes. Using various enzymes, these processes can amplify or replicate entire fragments or segments. Outside of a living vector on which to store the DNA, synthetic DNA is often stored on a silicon gel or chip. DNA is synthesized as oligonucleotides, or individual DNA fragments, which are much shorter than an entire genome, usually 100-300 base pairs long. These strands are often referred to by their shorthand, as oligos. These oligonucleotides can be connected into longer sequences, allowing for more base pairs per segment. In the field of medical technology, oligonucleotides are used to identify specific fragments of mRNA, as antisense RNAs (RNA segments that bind to mRNA in order to negate their function), or a variety of other factors.<sup>37</sup> In data storage, the further development of technology for *de novo* production of oligos, or the “uninterrupted synthesis of large sequences,” as opposed to the altering or modification of pre-existing DNA, will be crucial to reducing the cost and increasing the speed of DNA-based data storage.<sup>38</sup> This technology will also show to be the greatest limitation in making DNA-based data storage a viable and competitive format.

Unlike the basic process of biological DNA, which follows the basic steps of replication, transcription, and translation into proteins, DNA-based data storage can be broken down into seven steps (see fig. 1). *Encoding* (1) the data, *synthesizing* (2) the DNA molecules, *encapsulating* (3) the molecules onto a substrate, and *storing* (4) the DNA form the process of writing, creating, and storing the information. The retrieval of information is comprised of *releasing* (5) the DNA from the substrate, *sequencing* (6) the DNA in order to extract the sequences of nucleotides, and *decoding* (7) the information from nucleotides to binary (or digital) data. We can refer to these two directions of transferring data as the *write* and *read* of

---

<sup>37</sup> Kazuki Takakura, Atsushi Kawamura, Yuichi Torisu, Shigeo Koido, Naohisa Yahagi, and Masayuki Saruta, “The Clinical Potential of Oligonucleotide Therapeutics against Pancreatic Cancer,” *International Journal of Molecular Sciences* 20, no. 3331 (2019): 1, doi:10.3390/ijms20133331.

<sup>38</sup> Potomac Institute for Policy Studies, “The Future of DNA Data Storage,” 24.

information. In the most basic sense, the write and read power of DNA – the ability to store a lot of information, to store it quickly, to store it for a long time, and to retrieve it quickly – will determine the viability of DNA-based data storage into the future. In the following section, I will work through each of the seven steps of DNA-based data storage, where current research is pointing the field, and what developments are being made. Due to the availability of current research, and with the intent of making these descriptions relevant to the archiving field, discussions will focus on the encoding, synthesis, and sequencing stages.

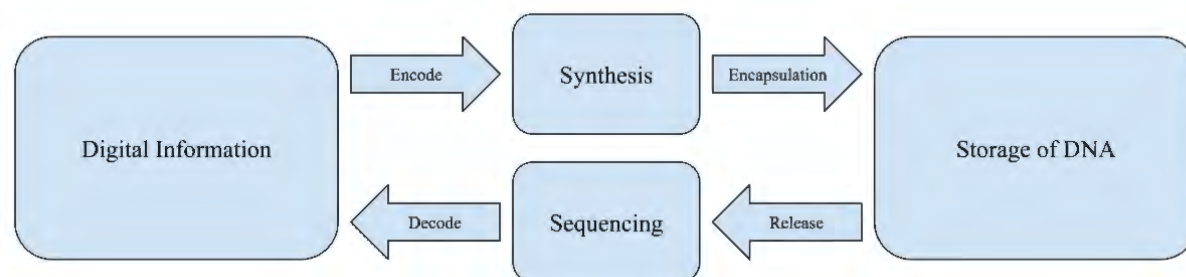


Figure 1. Workflow for DNA-Based Data Storage.

### Encoding

With either in vivo or in vitro, encoding – or in basic terms, the conversion from binary data to nucleotides – is the first stage for storing digital information onto DNA. As aforementioned, DNA-based data storage has yet to be standardized by any industry. In a workshop organized by the JPEG DNA-based Media Storage working group on 9 April 2021, both Sergey Yekhanin of Microsoft Research and Reinhard Heckel of the Technical University of Information Studies agreed that DNA-based data storage will become commercially viable within the next ten years.<sup>39</sup> In anticipation of that public availability, DNA will need to become

<sup>39</sup> Touradj Ebrahimi, Fernando Pereira, Linda Tadic, Reinhard Heckel, Sergey Yekhanin, “3<sup>rd</sup> JPEG DNA Workshop” (Zoom webinar, JPEG, 9 April 2021), [https://jpeg.org/items/20210316\\_jpeg\\_dna\\_3rd\\_workshop\\_announcement.html](https://jpeg.org/items/20210316_jpeg_dna_3rd_workshop_announcement.html).



standardized. A discussion of the various consortia led by industry players as well as emerging leaders will follow later in this section; however, it is important to note how the stage of encoding will be absolutely crucial in building that standard. As with all media – the grain density or flexibility of photochemical film; the voltage difference between North America and Europe for video tape – setting a standard will require contending with the physical limitations of that media. For example, with synthetic DNA, encoding must be done in segments in order to accommodate the short length of oligonucleotides. In the encoding of digital data onto synthetic DNA, setting a standard must contend with both the biological constraints of the macromolecule as well as the computing constraints affected by those physical limitations.

The JPEG DNA-based Media Storage working group, has held three workshops between October 2020 and May 2021. Working with a number of biotechnology industry professionals and researchers as well as those working with media storage in other formats, they have been compiling resources outlining how these different biological and computational constraints affect each other. The working group has identified three main encoding architectures that DNA-based data storage researchers and companies have been investigating to ensure that data will be stable on DNA: a simple transcoding-based architecture, constrained coding-based architecture, and sample-based architecture.<sup>40</sup> These three different encoding architectures approach the new medium of DNA differently. All three have advantages and disadvantages with respect to compression, how they approach biological and computational constraints, and how portable they are (which will directly impact viability into the future). Setting an encoding scheme will

---

<sup>40</sup> Marc Antonini, Luis Cruz, Eduardo da Silva, Melpomeni Dimopoulou, Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, António Pinheiro, and Mohamad Raad, “DNA-based Media Storage: State-of-the-Art, Challenges, Use Cases and Requirements version 3.0,” ed. Touradj Ebrahimi and Fernando Pereira, *JPEG DNA Exploration* (Lausanne: ISO/IEC JTC 1/SC29/WG1N90023, 22 January 2021): 6, [http://ds.jpeg.org/documents/jpegdna/wg1n90023-REQ-DNA-based\\_Media\\_Storage\\_State-of-the-Art\\_Challenges\\_Use\\_Cases\\_and\\_Requirements.pdf](http://ds.jpeg.org/documents/jpegdna/wg1n90023-REQ-DNA-based_Media_Storage_State-of-the-Art_Challenges_Use_Cases_and_Requirements.pdf).

affect not only how data is decoded into the future, but also the error rate of files, the speed and cost of writing and reading files, and, in many ways, how the DNA strands are synthesized. Because sample-based architecture is most advantageous to JPEG and other images encoded similarly (and therefore not appropriate for other digital data), I will focus more specifically on the other two. After first considering the computational constraints with an overview of transcoding- and constrained coding-based architectures, it becomes apparent that the biological and physical constraints of DNA may become the determining factor in selecting an appropriate encoding scheme.

Currently, several different schemes for encoding have been used in a variety of different experiments, but the most popular by far has been the simple transcoding scheme used to encode a book, images, and software at the Wyss Institute.<sup>41</sup> Simple transcoding approaches digital data as a set of binary information in base-2 and maps those 1's and 0's directly onto the 4-base nucleotides AGCT. There are several different ways to convert binary to quaternary, either in a 'one to two' encoding scheme where a 0 becomes either an A or T and a 1 becomes either a C or G or where each nucleotide corresponds to two digits, i.e., A becomes 00, T becomes 01 and so on (see fig. 2). To note of this simple transcoding scheme is how this architecture assumes that data has been pre-coded into binary. This facet will become important when considering the biological constraints; in effect, the data itself is agnostic to its storage substrate. The JPEG working group points out that this encoding structure could cause a lack of control over compression eliminate the possibility for lossy compression.<sup>42</sup> The lack of lossy compression may not appear as an immediate concern for archivists, but this fact of simple transcoding can

---

<sup>41</sup> George M. Church, Yuan Gao, and Sriram Kosuri, "Next-Generation Digital Information Storage in DNA," *Science* 337, no. 6102 (28 September 2012): 1628, <https://www.jstor.org/stable/41703606>.

<sup>42</sup> Antonini, et al., "DNA-based Media Storage," 6.

become a problem with larger file sizes that become more unwieldy as they grow larger due to the redundancy required to prevent errors.

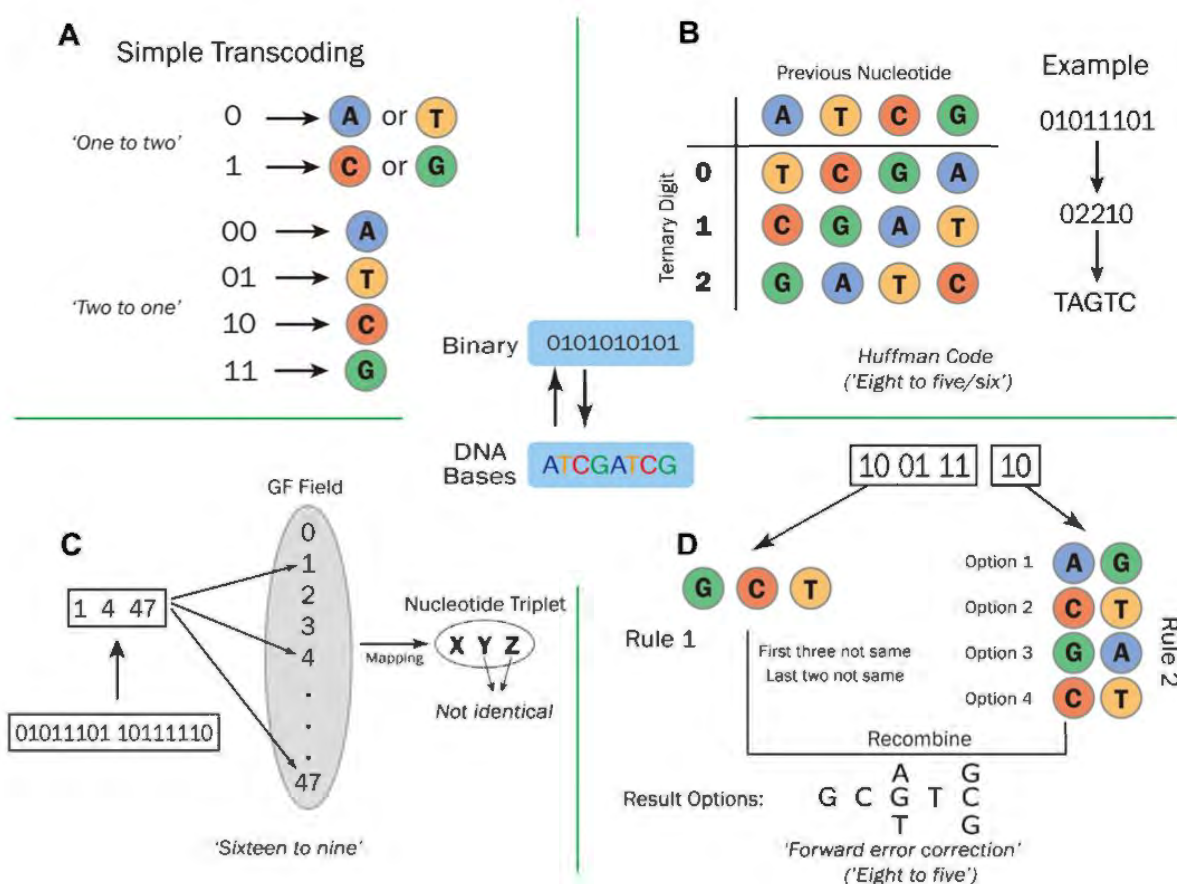


Figure 2. Different methods of encoding binary data onto DNA, from simple transcoding (A) to more complex mapping schemes.<sup>43</sup>

Constrained coding-based architecture, as the name suggests, considers the actual constraints of the encoded final product – for example, a fixed length – and uses statistics to generate the code.<sup>44</sup> More specifically, as Ping et al. describe, a constrained coding-based

<sup>43</sup> Zhi Ping, Dongzhao Ma, Xiaoluo Huang, Shihong Chen, Longying Liu, Fei Guo, Sha Joe Zhu, and Yue Shen, "Carbon-based archiving: current progress and future prospects of DNA-based data storage," *GigaScience* 8 (2019): Figure 1, doi: 10.1093/gigascience/giz075.

<sup>44</sup> Antonini, et al., "DNA-based Media Storage," 6.

architecture can use more complex coding as a form of lossless data compression through use of Huffman coding.<sup>45</sup> Huffman coding is a form of entropy coding that employs probability and matrices in order to represent information spatially into bins and thus compress it using an algorithm.<sup>46</sup> In effect, using Huffman coding allowed researchers to improve “the coding potential to 1.58 bits/nt.”<sup>47</sup> Importantly, quantization cannot be reversed, and while it does allow for error detection, it does not correct errors. In their search for a more effective coding scheme, Ping et al. submit that while Nick Goldman’s use of Huffman coding improved their efficiency and took into account some biological constraints, it still has room for improvement. In their study, they continue to research encoding schemes that can apply error detection and correction based on other entropy coding. Moving from entropy coding and quantization, Ping et al. turn to fountain coding schemes. Fountain coding, also called erasure coding, is used primarily for communication systems. It segments information into packets and, still relying on a simple 2-base to 4-base transcoding table, organizes the oligonucleotides into grids.<sup>48</sup>

Constrained coding-based architectures that employ additional methods of coding beyond simple transcoding, require more computational power and more redundancy in order to ensure error prevention, slowing down decoding. Ping et al. see fountain coding as a method that could help encoders avoid the high cost of decoding. Still, any encoding scheme will require more experimentation and research before a standard is set. Most of the math and computer science of these encoding schemes is ultra-specific and not completely necessary for archivists to be concerned with at this stage. However, digital preservationists often must understand how data

---

<sup>45</sup> Ping, et al., “Carbon-based archiving,” 3.

<sup>46</sup> Wing-Kuen Ling, *Nonlinear Digital Filters: Analysis and Applications* (Academic Press, 2007), 1, <https://doi.org/10.1016/B978-012372536-3/50001-6>.

<sup>47</sup> Ping, et al., “Carbon-based archiving,” 3.

<sup>48</sup> Ping, et al., “Carbon-based archiving,” 5.

persists on its physical substrate, be that a hard drive, a data tape, or a floppy disk. Therefore, learning about these encoding schemes helps us to understand how a standard is constructed and how the actual information will interact with the macromolecule on which it persists.

Understanding the biological constraints involved in setting an encoding scheme also informs archivists how the structure of DNA will affect information and therefore affect our work of maintaining data fixity. Archivists have dealt with vinegary film, sticky tapes, and flipped bits; in the very near future, double helixes will unravel and nucleotides will mutate.

The biological constraints of encoding can be divided into three different aspects of DNA: oligonucleotides, DNA structure, and bioethics. Oligonucleotide length is restricted in *de novo* synthesis – the synthesis of new nucleotides, as opposed to replication of existing ones – especially synthesis with a high through-put that reduces cost by creating large numbers of oligonucleotides at once.<sup>49</sup> The longer the oligo length, the more likely it is that the oligo will degrade; for these reasons, oligos are usually between 100 and 300 base pairs. Above 300 base pairs, the oligo degrades relatively quickly and is prone to errors.<sup>50</sup> Microsoft Research, in their application of DNA-based data storage, has identified the range between 120 and 150 nucleotides for maximum efficiency through a ternary encoding process (base-2 to base-3).<sup>51</sup> In many cases, oligos are synthesized and then mixed together at a high through-put. Each is assigned an index; of course, the longer the index sequence, the shorter the length of the oligo that contains the payload of information. The less space to include information, of course, the more oligos one needs, and so on.<sup>52</sup> These issues are currently being addressed by advancements

---

<sup>49</sup> Ping, et al., “Carbon-based archiving,” 5.

<sup>50</sup> Antonini, et al., “DNA-based Media Storage,” 9.

<sup>51</sup> James Bornholt, Randolph Lopez, Douglas M. Carmean, Luis Ceze, Georg Seelig, and Karin Strauss, “Toward a DNA-Based Archival Storage System,” *IEEE Micro* 37, no. 3 (2017): 99, 10.1109/MM.2017.70.

<sup>52</sup> Ping, et al., “Carbon-based archiving,” 7.

in several areas: more compression and encoding schemes, gaining the ability to synthesize longer oligos, and using libraries of pre-existing oligos in order to construct sequences.

Like other media that have their own specific constraints, DNA's will stem from its physical qualities that make it both appealing and difficult to manage. There are several biological constraints that arise specifically from the chemical structure of DNA. The size difference between the four nucleotides means that high repeats of G-C bonded nucleotides can result in an unstable double helix. Because guanine nucleotides are that much larger, if they are repeated often, the double helix becomes unbalanced and can lose structural integrity. For these reasons, the "percentage of G and C in the oligos should be lower or equal to one of A and T."<sup>53</sup> Additionally, homopolymers, or strands of the same nucleotide – as well as any small, repeated sequences – can cause structural issues. For example, these would look like long repeats of a single nucleotide like AGCTTTTACGT or repeats of a short sequence like ATATATATAT. To put this into perspective, these small, repeated sequences also cause problems in human DNA; for example, symptoms caused by Huntington disease appear because of a mutation that causes a CAG trinucleotide repeat that exceeds 28.<sup>54</sup> An elongated sequence of repeats causes unstable replication and produce even more repeats upon replication. As synthetic DNA can also undergo biological processes, such as the polymerase chain reaction (PCR) that amplifies or replicates sequences, these long repeats in oligonucleotides will also cause an unstable DNA or even secondary structure formation.<sup>55</sup> Any encoding architecture, simple or constrained or

---

<sup>53</sup> Antonini, et al., "DNA-based Media Storage," 7.

<sup>54</sup> Walker, Francis, "Huntington's disease," *The Lancet* 369, no. 9557 (20 January 2007): 218, [https://doi.org/10.1016/S0140-6736\(07\)60111-1](https://doi.org/10.1016/S0140-6736(07)60111-1).

<sup>55</sup> Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, Christopher N. Takahashi, Sharon Newman, Hsing-Yeh Parker, Cyrus Rashtchian, Kendall Stewart, Gagan Gupta, Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze, and Karin Strauss, "Random

fountain, must account for these peculiar qualities of DNA in order to avoid the degradation or error production of DNA in storage.

The last consideration encoding schemes must make is that of bioethics and biosafety. As stated earlier, synthetic DNA can undergo biological processes just like living DNA. For example, recently, synthetic DNA is what allowed scientists to create the COVID-19 mRNA vaccine. While it is unclear at this point how much of a threat this aspect of DNA synthesis may be to health and safety, the JPEG working group has noted:

Because dealing with the code of life, there are challenges related to the creation of dangerous DNA sequences, notably corresponding to known or unknown viruses or other dangerous pathogens. All the safety and ethical issues related to this possibility need to be carefully addressed, clarified and avoided.<sup>56</sup>

The JPEG paper later clarifies that there must be a way to ensure that the encoding output “be unambiguously recognized as artificial DNA.”<sup>57</sup> Additionally, they propose that any encoding scheme and later, the DNA-based data storage standards, must ensure that any outputs do not constitute danger with respect to biosafety. Scientists have used synthetic DNA to turn relatively harmless bacteria into toxin-producing ones, and any production of synthetic DNA has the same potential. Currently, most companies manufacturing synthetic DNA provide those genes to researchers who order specific genes and sequences. For example, Twist Bioscience, does privately determine whether a synthetic order is too dangerous, occasionally refusing an order if it can be used to construct a dangerous pathogen.<sup>58</sup> As the US Department of Health and Human

---

access in large-scale DNA data storage,” *Nature Biology* 36, no. 3 (29 February 2018): 243, <https://doi.org/10.1038/nbt.4079>.

<sup>56</sup> Antonini, et al., “DNA-based Media Storage,” 17.

<sup>57</sup> Antonini, et al., “DNA-based Media Storage,” 19.

<sup>58</sup> Nell Greenfieldboyce, “As Made-To-Order DNA Gets Cheaper, Keeping It Out of the Wrong Hands Gets Harder,” *National Public Radio*, 24 September 2019, accessed 23 April 2021,

Services works to update the government's guidelines for production and sale of synthetic genes, encoding structures will have to work against a pre-existing database of viral genes so as not to reproduce them accidentally. Regardless, this issue of bioethics ushers in a whole new realm of ethical quandaries archivists will have to contend with on a technical level when handling the storage of our cultural heritage.

### *Synthesizing*

The second step in a DNA-based data storage workflow, after binary has been mapped into a nucleotide sequence, is to synthesize the DNA. DNA-based data storage as a consumer technology will rely heavily on the development of synthesis technology as well as sequencing technology in order to become a cost-effective and viable storage format. Where the sequencing industry has grown massively since the start of the Human Genome Project, the synthesis industry is slightly more recent and has yet to become similarly cost-effective.<sup>59</sup> However, with the recent success of biopharmaceutical research relying on synthetic DNA and the growth of these companies, many in the industry expect the cost to decrease significantly.<sup>60</sup> DNA-based data storage as a commercial enterprise has primarily emerged from DNA synthesis companies that have seen developing storage technology as a natural extension of their current business in generating DNA sequences for medical research. As with sequencing, the greatest constraint with synthesis will be speed and cost: the cost of each per base and how long it takes to create or read back each base. Unlike encoding, which will be the most crucial step of setting a standard, synthesis relies more on the current biotechnology industry in order to make DNA-based data

---

<https://www.npr.org/sections/health-shots/2019/09/24/762834987/as-made-to-order-dna-gets-cheaper-keeping-it-out-of-the-wrong-hands-gets-harder>.

<sup>59</sup> Randall A. Hughes and Andrew D. Ellington, "Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology," *Cold Spring Harbor Perspectives in Biology* 9, no. 1 (January 2017): 2, <https://doi.org/10.1101/cshperspect.a023812>.

<sup>60</sup> Potomac Institute for Policy Studies, "The Future of DNA Data Storage," 27.



storage appealing in the first place. However, neither exists in a vacuum; as with *in vivo* research affecting *in vitro* synthesis, the co-option of new enzymatic processes for synthesis will affect encoding schemes. For these reasons, a brief description of synthesis technology merits discussion.

Scientists have been successfully synthesizing genetic material since at least the mid-1960s when Gobind Khorana and his colleagues successfully synthesized oligonucleotides and ligated (or attached) them together to create a 14-amino-acid hormone and insert it into *E. coli*.<sup>61</sup> Since then, many developments have been made, but synthesis still uses oligonucleotides ligated together. Today, DNA synthesis primarily relies on “variations of the phosphoramidite chemistry methods either on traditional column-based synthesizers or microarray-based synthesizers.”<sup>62</sup> In short, these are chemical reactions that stitch together an oligo one nucleotide at a time. In these processes, certain chemical reactions such as chain elongation cycles reduce the error rate in oligos.<sup>63</sup> Column-based synthesis traditionally uses a plate with 96 columns where the sequence is built on controlled-porosity glass beads (see fig. 3).<sup>64</sup> On the other hand, microarray-based synthesis employs chips “containing tens of thousands of distinct features” in order to synthesize tens of thousands of oligos at once.<sup>65</sup> Microarray-based synthesis essentially dumps the oligonucleotides into a pool from which they must be retrieved. While early chip-based synthesis also used phosphoramidite synthesis, others have employed light-activated chemistries with micromirrors.<sup>66</sup> More recently, use of ink-jet printing technology combined with chips have

---

<sup>61</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 4.

<sup>62</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 4.

<sup>63</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 4.

<sup>64</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 6.

<sup>65</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 6.

<sup>66</sup> Hughes and Ellington, “Synthetic DNA Synthesis and Assembly,” 7.

made phosphoramidite processes less expensive and more efficient and accurate.<sup>67</sup> Because of the thousands of available features on a chip and the ability to synthesize more than one at once, microarray-based synthesis is cheaper than column based in its first stages, but requires more labor after the fact to reduce synthesis-related errors.<sup>68</sup>

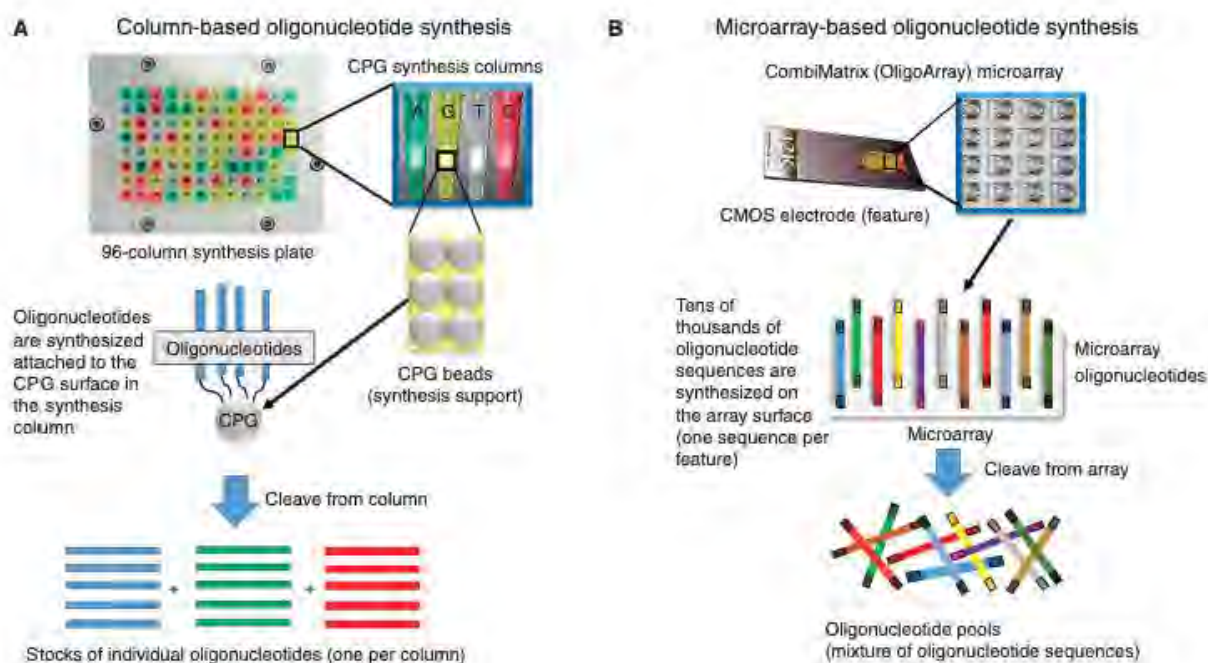


Figure 3. Column-based and microarray-based oligonucleotide synthesis.<sup>69</sup>

After being the backbone for DNA synthesis for over thirty years, the needs for synthesis are outpacing the limitations of phosphoramidite; molecular biologists are turning not just to larger plates or more miniaturization but also to researching more enzymatic processes that can speed up the process. The turn towards more enzymatic processes is a theme in all of DNA-based data storage. Instead of relying on outside technology, molecular biologists have stressed that the power of DNA will come from the molecule itself and all the enzymatic processes it

<sup>67</sup> Hughes and Ellington, "Synthetic DNA Synthesis and Assembly," 7.

<sup>68</sup> Hughes and Ellington, "Synthetic DNA Synthesis and Assembly," 7.

<sup>69</sup> Hughes and Ellington, "Synthetic DNA Synthesis and Assembly," Figure 3, 6.

already undergoes in cells.<sup>70</sup> Examples of these ventures can be found in the increased interest in enzymes that add nucleotides to the terminus of sequences, or even with the CRISPR-Cas system (which at this point, is only useful for small edits on template DNA). With the growing market for synthetic DNA and now with its potential application as a data storage format, a variety of companies and researchers are now using the opportunity to develop a variety of new applications and enzymatic processes. For example, Twist Bioscience uses a silicon synthesis chip the same physical size as a column-based oligonucleotide synthesis plate that, by miniaturizing the process, has increased throughput a thousand-fold.<sup>71</sup> DNA Script is in the process of commercializing a benchtop printer with a highly reduced error rate.<sup>72</sup> It should be noted that once a sequence is synthesized *de novo*, it is very fast and simple to replicate it through polymerase chain reaction (PCR), which is why Catalog DNA – another boutique DNA synthesis company based in Massachusetts – uses pre-synthesized DNA molecules as well as *de novo* synthesis.<sup>73</sup>

### *Encapsulating, Storing, and Releasing*

Once the DNA has been synthesized and removed from the column or chip, the molecules must be transferred to a substrate in order to be preserved. Without a substrate, DNA

---

<sup>70</sup> Michael Eisenstein, “Enzymatic DNA synthesis enters a new phase,” *Nature Biotechnology* 38 (October 2020): 1113, <https://doi.org/10.1038/s41587-020-0695-9>.

<sup>71</sup> Twist Bioscience, “Reducing Our Chemical Footprint for a Sustainable Future in Biotechnology,” Company News & Updates, 22 April 2021, accessed 23 April 2021, <https://www.twistbioscience.com/blog/company-news-updates/reducing-our-chemical-footprint-sustainable-future-biotechnology>.

<sup>72</sup> Eisenstein, “Enzymatic DNA synthesis enters a new phase,” 1114.

<sup>73</sup> “DNA-based Data Storage and Computation Provider CATALOG, Founded by MIT Scientists, Raises \$10 Million and Names IBM Executive as CTO,” Catalog DNA, 8 September 2020, accessed 24 April 2021, <https://www.catalogdna.com/blog/dna-based-data-storage-and-computation-provider-catalog-founded-by-mit-scientists-raises-10-million-and-names-ibm-executive-as-cto>.

is fragile and can degrade from hydrolysis or oxidation.<sup>74</sup> Some may be surprised to hear how vulnerable DNA actually is, especially when those proffering DNA often refer to the extraction of ancient DNA (aDNA) from fossils that are thousands of years old. However, those fossils and bones encase aDNA within a “collagen/calcium phosphate matrix of bones,” rendering it incredibly stable.<sup>75</sup> In some recent experiments using aging simulation through exposure to higher heat and relative humidity, researchers compare DNA in bone with solution, nanoparticle, and salt. In the tradeoff between stability, loading DNA, and handling simplicity, alkaline salts – such as calcium chloride or magnesium chloride – proved to be comparable if not better in preventing degradation.<sup>76</sup> Other techniques for encapsulating the DNA onto a substrate include encapsulating DNA into amorphous glass spheres that also mimic the way aDNA persists in bone.<sup>77</sup> Work by Organick et al. also moves this research forward with a comprehensive study on the preservation of DNA and working towards solutions for room temperature storage making scalable data storage possible.<sup>78</sup> These methods of encapsulation also force us to ask how sustainable these materials will be compared to current data storage formats. Currently little research has been done to compare how silicon used in encapsulation compare to the depletion of memory-grade silicon used for computing.

---

<sup>74</sup> A. Xavier Kohll, Philipp L. Antkowiak, Weida D. Chen, Bichlien H. Nguyen, Wendelin J. Stark, Luis Ceze, Karin Strauss, and Robert N. Grass, “Stabilizing synthetic DNA for long term data storage with earth alkaline salts,” *Chemical Communications* 56 (24 February 2020): 3613, <https://doi.org/10.1039/D0CC00222D>.

<sup>75</sup> Kohll, et al., “Stabilizing synthetic DNA,” 3613.

<sup>76</sup> Kohll, et al., “Stabilizing synthetic DNA,” 3615-6.

<sup>77</sup> Daniela Paunescu, Michela Puddu, Justus O. B. Soellner, Philipp R. Stoessel, and Robert N. Grass, “Reversible DNA encapsulation in silica to produce ROS-resistant and heat-resistant synthetic DNA ‘fossils’,” *Nature Protocols* 8 (7 November 2013): 2440, <https://doi.org/10.1038/nprot.2013.154>.

<sup>78</sup> Organick, Lee, Bichlien H. Nguyen, Rachel McAmis, Weida D. Chen, A. Xavier Kohll, Siena Dumas Ang, Robert N. Grass, Luis Ceze, and Karin Strauss, “An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage,” *Small Methods* (25 January 2021), <https://doi.org/10.1002/smt.202001094>.

Research is still being done to pinpoint the best conditions for storing DNA and ensuring that those conditions will mirror archival storage conditions for current data storage media or be better and more energy efficient. For example, current optimal DNA storage requires low temperatures (-80° C) or storage in dry, anoxic (no oxygen) environments, which are difficult to maintain at a large scale.<sup>79</sup> Additionally, these conditions far exceed the capabilities of most archival storage. For Linear Tape-Open (LTO), the preferred storage environment can range from 55° F to 74° F and 25-50% relative humidity.<sup>80</sup> Storing DNA must be able to match the offline cold storage formats it intends to replace. These encapsulation techniques make storing DNA much less difficult. They also indicate that the storage of DNA will require much more than simple comparisons to aDNA recovery after thousands of years. Additionally, releasing DNA from the substrate in order to prepare it for sequencing (or reading back the DNA) must be simple and accessible in order to speed up the process. Lastly, archivists must be able to handle the DNA with relative ease themselves in order to maintain physical control.

### *Sequencing*

In short, sequencing is the process of ‘reading’ back DNA, or uncovering the sequence of nucleotides. The industry of sequencing technology has expanded exponentially since the start of the Human Genome Project in 1990. With much more commercial real estate bought up by sequencing companies, this technology has progressed rapidly and continues to do so. As mentioned earlier, the cost of synthesis (in 2019, \$.00001-.001 USD per nucleotide)<sup>81</sup> is far more costly than that of sequencing (in 2020, \$0.01 per million base pairs).<sup>82</sup> The HGP precipitated the

---

<sup>79</sup> Kohll, et al., “Stabilizing synthetic DNA,” 3613.

<sup>80</sup> Linda Tadic, “Audiovisual digital preservation and DNA storage requirements,” (Zoom slide presentation, Third JPEG DNA Workshop, 9 April 2021).

<sup>81</sup> Ping, et al., “Carbon-based archiving,” 6.

<sup>82</sup> National Human Genome Research Institute, “The Cost of Sequencing a Human Genome.”

commercialization of Sanger sequencing, the first automated sequencing technology. Today, most sequencing technology use Next-Generation Sequencing (NGS) technology, which was first proposed in 2000. NGS can have a high error rate with longer sequences, but because of the many short sequences used in data storage (due to ease of synthesis and lowered costs), it is appropriate for sequencing for data storage.<sup>83</sup> NGS requires a lab to generate a template library – essentially indexing the fragments by adding adapters to the sequences – and then, the fragments are amplified (replicated), and in the replication, chemically modified nucleotides with fluorescent tags bind to those fragments.<sup>84</sup> From there, software can identify the accuracy of the read and employ error correction in order to extract the information appropriately.

NGS is the dominant technology for sequencing; however the equipment is large and requires specialized skill sets to operate. The Oxford Nanopore MinION technology provides a promising alternative as a handheld device. However, the MinION reads back sequences incredibly slowly at only 75 base pairs per second.<sup>85</sup> The relative success of the MinION, however, shows that the need for small devices with low error rates and high read times is still a necessity; moreover, it encourages inspiration from biology in order to increase through-put. The current state of DNA data sequencing is still far behind that of LTO data tape in terms of time and cost however (see table 1). LTO data tape, for example, has a performance latency of minutes, as opposed to DNA's hours or even days.<sup>86</sup> Additionally, this discussion has up until now only concerned the sequencing of the entire data stored. With many stating DNA's data density has a theoretical limit in the exabytes, any sequencing must consider that clients will

---

<sup>83</sup> Ping, et al., "Carbon-based archiving," 6.

<sup>84</sup> "Understanding the NGS workflow," *Illumina*, accessed 24 April 2021, <https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html>.

<sup>85</sup> Potomac Institute, "The Future of DNA Data Storage," 30.

<sup>86</sup> Bornholt, et al., "Toward a DNA-Based Archival Storage System," 103.

need to extract small files quickly with no error rate from larger bitstreams. Essentially, DNA-based data storage will have to offer a robust random access for data stored.

Platform	Error rate	Runtime	Instrument cost	Cost per Gb
Illumina MiSeq	0.10%	4-56 hours	\$99,000	\$110-1,000
Illumina HiSeq 2000	0.26%	3-10 days	\$654,000	\$41
Illumina HiSeq 2500	0.10%	7 hours-6 days	\$690,000	\$30-320
Illumina NextSeq	0.20%	11-29 hours	\$250,000	\$33-43
Oxford Nanopore MinION	8.0%	Less than 48 hours	\$1,000	\$70

*Table 1. Summary of sequencing platforms as of 2018.<sup>87</sup>*

Random access to data is one of the most pressing necessities for any new storage format technology. For computing, lack of random access inhibits the ability to increase data size because the sequencing and decoding of whole datasets must be done in order to access any amount of data. As was briefly mentioned in the discussion of oligonucleotides and encoding, each individual strand must carry a copy of its own address as strands will not be organized spatially on the storage substrate.<sup>88</sup> In encoding, this poses a problem for reaching a happy medium between oligo size, maximum payload, and address size. In addition to storing the address, each oligo has primers on each end of the strand – essentially key identifiers that enable efficient sequencing. Currently, most random access methodologies in DNA-based data storage rely heavily on polymerase chain reaction (PCR). PCR is an enzymatic process that amplifies, or replicates, selected strands from a pool of molecules, identifying them from the primers. After

<sup>87</sup> Ping, et al., “Carbon-based archiving,” 7.

<sup>88</sup> Bornholt, et al., “Toward a DNA-Based Archival Storage System,” 100.

amplification, a sample will contain a large number of copies of the desired strand, which can be identified and sequenced.<sup>89</sup> Bornholt et al. suggest that a physical library could contain individual pools with 100 Terabytes each of data in order to enable more efficient information extraction. Organick et al. also identify how the generation of effective PCR primers and devising a standardized way of doing so will determine much of random access success.<sup>90</sup> Random access methods, much like other aspects of DNA-based data storage, will need to look towards biological processes to be competitive.

### *Decoding*

Decoding is the step by which data extracted from DNA by sequencing – in the form of nucleotide sequences – is mapped back into binary data. Not as simple as using a key to decode an encrypted set of information, decoding is also the process by which most error detection and correction takes place. As seen in the discussion of synthesis and sequencing, error production is very much a possibility and should be expected depending on what methods are employed for both. Therefore PCR, not just helpful in the random access of information, is necessary to create redundancy, which can help detect errors.<sup>91</sup> For the most part, redundancy is what allows the extraction of aDNA; mammoth bones do degrade over time, but because millions of samples are available during analysis, a very accurate genome can be sequenced through error detection and comparison. Column-based synthesis can be 100% accurate, but microarray-based synthesis has a higher error rate (and is cheaper). Sequencing often introduces errors regardless of technology. For these reasons, a knowledge of what technology has been used is necessary to knowing whether errors have been introduced during sequencing only or during synthesis as well.<sup>92</sup>

---

<sup>89</sup> Bornholt, et al., “Toward a DNA-Based Archival Storage System,” 100.

<sup>90</sup> Organick, et al., “Random access in large-scale DNA data storage,” 243.

<sup>91</sup> Antonini, et al., “DNA-based Media Storage,” 9.

<sup>92</sup> Bornholt, et al., “Toward a DNA-Based Archival Storage System,” 102.



The fact that sequencing and synthesis (as well as any random access) can introduce errors to the DNA means that any encoding and decoding schemes will need to employ robust error detection and correction. This effect is called “code resilience,” or the ability to “recover the information under the observed error regime.”<sup>93</sup> With any encoded data, we should have the ability to recover or decode the information without knowing the encoding scheme, the ways errors were introduced, and any past sequencing history, although, of course, that information could be integral to data recovery of corrupted sequences. In DNA, substitutions are the most common type of error, with deletions next, and inserts last. Additionally, primers tend to have less or no errors. In the experiment conducted by Organick et al., successfully recovering 200MB of information required between 4 and 14 reads per sequence.<sup>94</sup> These experiments demonstrate that, like with aDNA recovery, usually a complete set of information can be recovered with more reads; however, decoding, error detection, and error correction will need to be robust in order to lower the cost of these multiple reads. There are several schemes for providing error correction, but as with encoding, these can be divided into computational resilience and biological factors. PCR serves as an error corrector in the way that it amplifies sets of data, and different encoding schemes that maximize entropy coding can provide that resilience.<sup>95</sup>

The discussion of decoding concludes this section’s overview of how the technology of DNA-based data storage works. Still very much an assemblage of many different burgeoning technologies, DNA-based data storage has a long way to go before it becomes a standardized format and gains wide commercial usage. What this extended discussion of the various methods that have been used as well as those in development shows is that DNA-based data storage

---

<sup>93</sup> Organick, et al., “Random access in large-scale DNA data storage,” 245.

<sup>94</sup> Organick, et al., “Random access in large-scale DNA data storage,” 245.

<sup>95</sup> Reinhard Heckel, “Low cost DNA data storage with noisy synthesis and advanced error correction,” (Zoom slide presentation, 3<sup>rd</sup> JPEG DNA Workshop, 9 April 2021).

requires a lot of advancement unilaterally. More specifically, those methodologies that will become integral to a DNA-based data storage workflow will rely primarily on advancements in identifying and harnessing biological processes to increase efficiency. By attacking the issues of efficiency, error rates, and random access from both the coding structures and biological processes, many see DNA-based data storage becoming a viable, commercial format within the next ten years. With anticipated cost reduction, speed increases, and expanded features by then, archivists have that much time to wield power as primary designated use cases to imbue these technologies with archival standards, ethics, and, perhaps most importantly, capabilities.

### **Why DNA-Based Data Storage?**

In much of the literature describing DNA-based data storage and advocating for its application as a data storage format, authors list four main features and conditions that make DNA an attractive option for long term, offline storage. These features are: the *data density* available in DNA, the *longevity* made possible with DNA, the *stability* of the macromolecule, and the advertised *sustainability* of DNA when compared with other data storage formats. For most advocating for DNA-based data storage, the current context of exponentially growing amounts of data worldwide lend the development and promise of DNA a great urgency. Moreover, what links DNA to data storage in the way many authors both of science articles and of marketing materials is the slippage between genetic information and binary information. If DNA can hold the ‘code’ of genes, of life, it can certainly hold the code to our most precious data. In this section of the first chapter, I examine these four features in some detail, describe the associated studies and claims made to support the effectiveness of these features, and finally, offer a more in-depth comparison with LTO tape, which is perhaps the closest potential competitor to DNA.

The largest catalyst driving the sense of urgency with which many speak about DNA-based data storage is the exponential growth of data worldwide. In his talk to the 2<sup>nd</sup> JPEG DNA working group, Marc Antonini began his talk describing how many – specifically the International Data Corporation – predict that the world’s data will grow to 175 zettabytes by 2025.<sup>96</sup> David Rosenthal points out that this number from the IDC corresponds to data generated, not necessarily data stored.<sup>97</sup> However, “a significant fraction of this data is called ‘cold’ or infrequently accessed.”<sup>98</sup> Antonini then contrasted how cold data growth will see a 60% increase where data storage density in current technologies will only see a 20% improvement. It is within this gap – between growth and current long term storage capability – that DNA will, according to those believers, find its place. Of course, concurrent projects – notably Microsoft Research’s Project Silica, which stores data onto silica quartz glass with the same lasers used in Lasik surgery – offer decent competition.<sup>99</sup> However, where current glass can store around 75.6 Gigabytes of data in a 11.25cm<sup>3</sup> space, DNA can store 215 petabytes of data per gram, with a theoretical limit of 455 exabytes per gram.<sup>100</sup> Advances in technology that increase encoding efficiency and synthesis capability (notably, being able to synthesize longer oligos) will make approaching that theoretical limit possible. Compare that density to LTO-9 tape, the most recent LTO generation, which can store 18 terabytes of native data, or 45 terabytes compressed.<sup>101</sup> In

---

<sup>96</sup> Marc Antonini, “Image coding for long term storage on synthetic DNA,” (Slide presentation, JPEG, 30 September 2020), <https://www.youtube.com/watch?v=U5EKe-RDYN4>.

<sup>97</sup> David Rosenthal, “The Medium-Term Prospects for Long-Term Storage Systems,” *DSHR’s Blog*, 13 December 2016, accessed 2 May 2021, <https://blog.dshr.org/2016/12/the-medium-term-prospects-for-long-term.html>.

<sup>98</sup> Antonini, “Image coding.”

<sup>99</sup> Jennifer Langston, “Project Silica proof of concept stores Warner Bros. ‘Superman’ movie on quartz glass,” *Microsoft News*, 4 November 2019, accessed 27 April 2021, <https://news.microsoft.com/innovation-stories/ignite-project-silica-superman/>.

<sup>100</sup> Kohll, et al., “Stabilizing synthetic DNA,” 3613.

<sup>101</sup> “LTO-9,” *Ultrium LTO*, accessed 27 April 2021, [lto.org/lto-9/](http://lto.org/lto-9/).

all, density is perhaps DNA's greatest asset; if it can be harnessed to its full potential, this density makes DNA formidable in the world of data storage.

The proposed longevity of DNA is perhaps its second greatest appeal thrust forward by marketers. Longevity of a storage format can be deceiving, however, especially in the realm of digital preservation. The following chapter will consider more specifically the needs of digital preservationists in the face of claims of storage superiority. However, in brief, codecs for files, or the programs that encode and decode the bitstream of data, the operating systems they run on, and a whole other host of dependencies – software and hardware – do not last forever. In much of the literature surrounding DNA-based data storage, longevity is most often confirmed by the recovery of ancient DNA (aDNA) from fossils. DNA is “time tested by nature,” and can last thousands of years.<sup>102</sup> This comparison with aDNA recovery is not necessarily a false one. DNA in fossil, in cold and relatively dry circumstances and in large, salvageable quantities, can last that long, which is an indication of its longevity. However, the storage of digital data onto synthetic DNA on a substrate has yet to be tested the same way it has in bone. Moreover, aDNA recovery is much more complex, involving purification of DNA, the reconstruction of many different fragmented and degraded samples, and more.<sup>103</sup> This is to say that we should be wary when the only discussion of longevity is grounded in aDNA extraction.

Aside from the discussion of aDNA, there are several studies performing thermal damage in order to simulate the passage of time and calculate a possible longevity of synthetic DNA on various substrates. As the aforementioned study on encapsulating in salts suggests, these

---

<sup>102</sup> Luis Ceze, Jeff Nivala, and Karin Strauss, “Molecular digital data storage using DNA,” *Nature Reviews Genetics* 20 (8 May 2019): 456, <https://doi.org/10.1038/s41576-019-0125-3>.

<sup>103</sup> Maria A. Nieves-Colón, Andrew T. Ozga, William J. Pestle, Andrea Cucina, Vera Tiesler, Travis W. Stanton, Anne C. Stone, “Comparison of two ancient DNA extraction protocols for skeletal remains from topicals,” *American Journal of Physical Anthropology* 166, no. 4 (August 2018): 824, <https://doi.org/10.1002/ajpa.23472>.

decisions – on what material to store the DNA and under what conditions – will be a better indicator of DNA’s longevity rather than aDNA extraction. Perhaps more important to the immediate need of archival storage than whether DNA will last a hundred or a thousand years (film negatives can last at least a hundred) will be how the encoding, the error correction from errors produced during synthesis and sequencing, and the ease of retrieval affect what use that longevity will be. That being said, DNA’s stability and ability to persist for a long time without degradation or errors during storage do make DNA extremely conducive to long term storage. The double helix structure and the base stacking interactions in the sequence of nucleotides in that structure make DNA a stable macromolecule.<sup>104</sup> DNA has a half-life of 500 years, which increases in cold and dry conditions.<sup>105</sup> There is no doubt that DNA as a macromolecule will outlast hard drives and tape; it is another question whether the data stored on it will.

Lastly, one of the largest claims of superiority to other storage media is that of sustainability. Sustainability can mean many things, but in the discussion of DNA-based data storage, sustainability primarily refers to moving away from increasingly rare memory-grade silicon. In the move away from these rare materials – the Potomac Institute predicts storing the world’s data in 2040 will require 1000-kilograms of single-crystal wafer grade silicon, of which there will only be 108 – DNA does offer some promise.<sup>106</sup> DNA’s competition with glass, which is constructed from similar silicon,<sup>107</sup> may disappear as silicon does. Still, studies have yet to be

---

<sup>104</sup> Ping, et al., “Carbon-based archiving,” 1.

<sup>105</sup> Potomac Institute, “The Future of DNA Data Storage,” 11.

<sup>106</sup> Potomac Institute, “The Future of DNA Data Storage,” 7.

<sup>107</sup> Patrick Anderson, Richard Black, Ausra Cerkaskaite, Andromachi Chatzieleftheriou, James Clegg, Chris Dainty, Raluca Diaconu, Austin Donnelly, Rokas Drevinskas, Alexander Gaunt, Andreas Georgiou, Ariel Gomez Diaz, Peter G. Kazansky, David Lara, Sergey Legtchenko, Sebastian Nowozin, Aaron Ogus, Douglas Phillips, Ant Rowstron, Masaaki Sakakura, Ioan Stefanovici, Benn Thomsen, Lei Wang, “Glass: A New Media for a New Era?” *10<sup>th</sup> USENIX Workshop on Hot Topics in Storage and File Systems* HotStorage 18 (July 2018): 1, <https://www.microsoft.com/en-us/research/uploads/prod/2018/07/hotstorage18-paper-anderson.pdf>.

done on the materials and energy cost of DNA-based data storage. DNA may not use as much energy in storage due to its inert stability, but the amount of energy and chemicals required to synthesize and sequence DNA must be accounted for. This energy consumption and sustainability rating may lie more acutely in DNA-based data storage's chemical footprint.<sup>108</sup> For a recent, comprehensive study on the reliance on finite materials and data storage energy usage, Danielle Calle's MIAP thesis provides an excellent overview; in the future, DNA-based data storage coalesces into a standardized format, it will need to be integrated into that work.<sup>109</sup>

### **The State of the Field**

In November 2020, Microsoft, Twist Bioscience, Illumina, and Western Digital announced that they were in the process of forming the DNA Data Storage Alliance (DDSA).<sup>110</sup> This alliance, notably brings together companies with backgrounds in medical research and synthesis (Twist), commercial DNA sequencing (Illumina), and the information technology and computing sectors (Western Digital and Microsoft). The Alliance now includes at least thirteen other members, forming a base across the sectors of molecular research, DNA synthesis, DNA sequencing, data storage manufacturers, other computing companies, and various academic research institutions. The DDSA represents a step forward in establishing a standard as well as in marketing DNA-based data storage to a wide variety of use cases. More importantly, having competitors work together in developing a new technology could suggest interest in developing a technology that

---

<sup>108</sup> Twist Bioscience, "Reducing Our Chemical Footprint for a Sustainable Future in Biotechnology," Company News & Updates, 22 April 2021, accessed 23 April 2021, <https://www.twistbioscience.com/blog/company-news-updates/reducing-our-chemical-footprint-sustainable-future-biotechnology>.

<sup>109</sup> Danielle Calle, "Climate Change as a Prompt for the Digital Archive," MA Thesis (New York University, 2020).

<sup>110</sup> Vitak, Sarah, "Technology alliance boosts efforts to store data in DNA," *Nature.com*, 03 March 2021, <https://www.nature.com/articles/d41586-021-00534-w>.

can communicate across platforms. That is to say, DNA could be encoded, synthesized, stored, and sequenced all by different parties using different platforms, ensuring the format's sustainability into the future.

The DDSA's mission statement is to "Create and promote a new, zettabyte scale, enduring, always recoverable, format-stable, widely adopted, and cost effective archival storage ecosystem for the explosive growth of digital data based on DNA." More specifically, in a presentation given to the 2<sup>nd</sup> JPEG DNA Workshop, Daniel Chadash, Director of Digital Product Management at Twist Bioscience, stated that he and others had explicitly looked to the success of other formats – such as LTO and USB – that had relied heavily on industry collaboration to set a standard – in the development and establishment of the DDSA.<sup>111</sup> The Alliance comes at a point where the technology has been tested and proven in the lab and in some pilot projects but has yet to gain significant interest due to a number of constraints outlined in this section: financial cost, uneven application and results, poor random access, and slow latency or access times. However, with the DDSA's future white papers – promised are "Intro to DNA Data Storage," "DNA Data Storage Technical Roadmap," and "Requirements and Use Cases"<sup>112</sup> – the Alliance hopes to cement interest and grow a user base. Not yet a legal entity – a promoter group, in Chadash's words – the DDSA's reach may be limited until then, but its goals stated here mark it as the first coalition advocating for the use of DNA.

DNA-based data storage has established its merits, especially in its stability and durability. The next greatest hurdle in its journey to becoming a commercial storage format and potentially changing computing forever – by virtue of introducing biotechnology – will lie in the

---

<sup>111</sup> Daniel Chadash, "An Overview of the DNA Data Storage Alliance," (Zoom slide presentation, 2<sup>nd</sup> JPEG DNA Workshop, 27 January 2021), <https://www.youtube.com/watch?v=IFNE47YfVWg>.

<sup>112</sup> Chadash, "An Overview of the DNA Data Storage Alliance."

ability of the Alliance or other consortia to address the needs of their user base. As the next chapter outlines how DNA-based data storage researchers and companies have identified those use cases in archives, and have addressed those needs to date, we should keep in mind the technological limitations of DNA-based data storage. Chadash is not wrong when he discusses how explaining that DNA-based data storage could even function as a replacement for offline long term storage is a great learning curve for those familiar with current data storage.<sup>113</sup> LTO data tape and hard drives are building on long histories of development, trust, and integration into existing data storage industries. DNA, as a newcomer, must not only be able to provide all the services established competitors can, but it must also gain the trust of users by conforming to the different standards we set. As chapter one leaves us with a good grasp of how the technology works and what its current limitations are, chapter two will pick up with how archivists and cultural heritage institutions can use the way DNA-based data storage companies and the Alliance itself are catering to us in order to ensure that this storage format will be able to operate within our existing frameworks of digital preservation.

## **Chapter Two: DNA in the Archive**

In much of the language describing DNA-based data storage, the term ‘archival’ appears periodically. Primarily in scientific articles, its use indicates an understanding of archival storage from the perspective of those developing DNA storage technology. In the simplest of terms, these descriptions define archival storage as a space where objects are not accessed often but must be kept for a very long time. The perception of time – of what constitutes long term and what the implications for long term preservation are – varies wildly across these many

---

<sup>113</sup> Chadash, “An Overview of the DNA Data Storage Alliance.”



disciplines which converge at the site of DNA-based data storage. For that reason, the use of the word ‘archival’ is worth analyzing as a source of value and as a determinant for what qualities DNA-based data storage will have. It may be obvious to state, but scientific narratives – as well as archival scientific ones – rely heavily on the precision of language and the trust that users and readers will be clued into those precise meanings. One of the guiding texts for digital preservation, the Open Archival Information System (OAIS) reference model, which will steer much of the ensuing conversation here, also meta-textually provides a model for this kind of communication. In brief, OAIS defines a ‘Designated Community’<sup>114</sup> as “an identified group of potential Consumers who should be able to understand a particular set of information.”<sup>115</sup> As the first chapter of this thesis works to make the language of DNA-based data storage accessible to archivists, this second chapter works to mediate the language of the archive by ensuring that all parties are able to understand this set of information.

The limitations of DNA-based data storage make long term preservation an obvious choice as a primary use case. As discussed, DNA has a much longer latency or access period than any other storage format – the difference between minutes and hours or even days – and the random access methods are still precarious. DNA’s potential lifespan and durability make it attractive as a long-lasting format. For these reasons, many scientific articles point to long term preservation as the primary use case for DNA. The very high density in turn makes it an obvious choice for storing audiovisual media, which tend to have much higher file sizes than other media types. For these reasons, write the JPEG DNA working group, “DNA-based storage seems to firstly target large scale, long term preservation archives.”<sup>116</sup> With other papers echoing these

---

<sup>114</sup> See the Designated Community statement earlier in this thesis.

<sup>115</sup> *Reference Model for the Open Archival Information System (OAIS)*, 1-11.

<sup>116</sup> Antonini, et al., “DNA-based Media Storage,” 18.

sentiments, it is important, from the perspective of these targeted long term preservation archives, to understand how scientists developing DNA-based data storage use the term archive, in order to identify how to communicate our needs explicitly. Bornholt et al. claim to demonstrate a “DNA-based archival storage system,” which includes the writing and reading of DNA.<sup>117</sup> In their paper, storing data in “archival form” means data should last for a long time, occupy little space, and be accessed very few times or even never.<sup>118</sup> The Twist Bioscience white paper uses the word ‘permanent’ in association with “cold” archival storage.<sup>119</sup> Ping et al.’s paper “Carbon-based archiving,” using the gerund form of archive, calls DNA the “ultimate solution,” for archiving.<sup>120</sup> Charles Choi, in an article surveying the field of DNA-based data storage, pits DNA against LTO as many do. He also identifies how “the most immediate applications will likely be to archive ‘very valuable data that needs to last for a long time,’” in an interview with Karin Strauss, a researcher at Microsoft.<sup>121</sup> While only a sampling of perceptions, none of which is egregious or misguided, these descriptions of the archival application of DNA-based storage accomplish two things: they position DNA as a competitor with current technology used in archives (such as LTO), and they claim without differentiation between archives that DNA’s best use case is as a storage technology for archives.

What is particularly evocative is the continued association of long term preservation with the perceived infinite preservation of an object in space. From DNA’s first proof-of-concept in *Microvenus* to Ping et al.’s example of DNA in the Lunar Library, those working with this

---

<sup>117</sup> Bornholt, et al., “Toward a DNA-Based Archival Storage System,” 98.

<sup>118</sup> Bornholt, et al., “Toward a DNA-Based Archival Storage System,” 102.

<sup>119</sup> Twist Bioscience, “DNA-Based Digital Storage White Paper,” last modified 29 October 2018, accessed 27 April 2021, <https://www.twistbioscience.com/resources/white-paper/dna-based-digital-storage>.

<sup>120</sup> Ping, et al., “Carbon-based Archiving,” 7.

<sup>121</sup> Choi, Charles Q, “Nature’s Databank,” *Prism*, February 2020, <http://www.asee-prism.org/natures-databank>.

technology have envisioned it persisting in space, there to be accessed in an indeterminate future, possibly not even by humans. The Lunar Library is a project from the Arch Mission Foundation, a nonprofit organization that seeks to “back up planet Earth.”<sup>122</sup> The Library’s first concrete effort was in the delivery of 25 nickel discs stacked together and sealed, etched with 30 million pages of material, sent to the moon via Israel’s SpaceIL mission, which eventually crash-landed there in 2019.<sup>123</sup> The second phase of the Lunar Library, a collaboration between Microsoft, Twist Bioscience, University of Washington, and the Arch Mission Foundation, lies in “Memories in DNA,” an effort to dump that information onto the moon stored this time in DNA.<sup>124</sup> In many ways, the Arch Mission Foundation is a frustrating counterexample to the actual needs of Earth-bound archives and libraries. It is not exactly surprising that Arch Mission Foundation is bound up with tech billionaire Elon Musk’s SpaceX efforts to explore and eventually colonize Mars.<sup>125</sup> Here is not really the space to critique Musk’s (or Bezos’ or Branson’s) extraterrestrial efforts on the grounds of class, climate, colonization, global politics – many others have and will continue to do so. However, this approach to librarianship and to archives is dangerous to the actual archival application of DNA-based data storage and to its development as a technology. That all of these issues – of capital, of settler-colonialism, of apocalypse – converge in the realm of librarianship should cue archivists to intervene.

---

<sup>122</sup> Nova Spivack, “There is Now a 30 Million Page Backup of Planet Earth, on the Moon! – New NASA Images Show Location of the Lunar Library,” *Medium*, 16 May 2019, accessed 30 April 2021, <https://medium.com/@novaspivack/there-is-now-a-30-million-page-backup-of-planet-earth-on-the-moon-d0504458aca1>.

<sup>123</sup> Spivack, “There is Now a 30 Million Page Backup of Planet Earth, on the Moon!”

<sup>124</sup> Ping, et al., “Carbon-based Archiving,” 7.

<sup>125</sup> Nova Spivack, “Arch Mission Foundation Announces Our Payload On SpaceX Falcon Heavy,” *Medium*, 6 February 2018, accessed 30 April 2021, <https://medium.com/arch-mission-foundation/arch-mission-foundation-announces-our-payload-on-spacex-falcon-heavy-c4c9908d5dd1>.

Incredibly reminiscent of the Voyager records Joe Davis' *Microvenus* critiqued, the Arch Mission Foundation et al.'s Lunar Libraries (and libraries elsewhere) have the dual existential missions of "backing up" humanity as well as putting forth a version of humanity to the future, and possibly to extraterrestrial lifeforms. The Lunar Library and Arch Mission Foundation is grounded in some understanding of librarianship. Nova Spivack, Arch's founder, wrote of the SpaceIL crash's payload of nickel discs:

A library is not really a library unless it can be accessed by its intended audience.

Because the Lunar Library is at least partially intact, then *whether or not it is ever retrieved, it could be retrieved – it is not irretrievable* – and therefore it really is a library.<sup>126</sup>

A library is indeed not a library unless people use it. However, the assumption that simply because the information is not lost forever – destroyed in the Israeli space crash on the moon – that it must still be a library is problematic. We can indulge in the fantasy of a post-apocalyptic archive of the remains of humanity,<sup>127</sup> but really only as such. That Arch claims "the benefit of this is also not only for hypothetical potential beings in a million years," but also "now," impedes the actual work of librarianship.<sup>128</sup> This view of a library that does not ever need to be retrieved, or is difficult to retrieve by its designated community, conflicts with the work of archivists and digital preservationists whose daily work is to make information accessible. More importantly, with efforts such as the Lunar Libraries widely cited in scientific articles and referenced as

---

<sup>126</sup> Spivack, "There is Now a 30 Million Page Backup of Planet Earth, on the Moon!"

<sup>127</sup> Alexis Pauline Gumbs' *M Archive* accomplishes what a post-apocalyptic archive without a failed Israeli space launch; her poetry excavates from the post-apocalyptic dig of the leftovers of humanity the interweaving of class, race, information, technology, and ancestry.

<sup>128</sup> Nova Spivack, quoted in Clara Moskowitz, "DNA-Coded 'Lunar Library' Aims to Preserve Civilization for Millennia," *Scientific American*, 28 September 2018, <https://www.scientificamerican.com/article/dna-coded-lunar-library-aims-to-preserve-civilization-for-millennia>.

archival pilot projects – as opposed to fantastical art projects? billionaire pipe dreams? political extensions of settler-colonial states? – further obscure the actual work that needs to be done in order to make DNA-based data storage a viable format for preservation.

This chapter unravels the statement “it could be retrieved” through an analysis of integrating DNA-based data storage into an archival workflow. The possibility of retrieval is simply not enough when it comes to the preservation of digital objects. Many preservationists reiterate time and time again that it is not film that is the most vulnerable audiovisual media – it is digital objects, whose access rests on so many dependencies that the work to maintain their accessibility is endless. The temporality of a digital preservationist is very different from that of an unmanned and inaccessible library on the moon. While the latter is fantastical and perhaps exciting – whether you believe in aliens or not – it will not sustain a format whose proprietors wish it to become commercially viable. To that end, the first section of this chapter describes frameworks for digital preservation that define very clearly what long term archival storage is, what temporality digital preservationists operate within, and what advances need to occur for DNA-based data storage to become viable. The second half of this chapter shifts the thesis into the realm of the theoretical and asks questions about what it means to store information onto a substrate, and about what that substrate – DNA, here – does to the data it carries. Within the archival framework, and in the spirit of work within the audiovisual archiving community that has sought to contend with the effects formats have on information, I try to understand how DNA can show us new ways of caring for our cultural heritage and of caring for our formats and what they do.

### **“Care and Feeding”**

In the realm of digital preservation work are several frameworks and organizational schemes that offer models, and other self-rating systems that standardize the labor of librarianship. These include OAIS, as well as the National Digital Stewardship Alliance (NDSA) Levels of Preservation or the Trusted Digital Repository (TDR) Checklist (ISO 1363). Whether or not institutions that perform digital preservation follow these standards and models or qualify to be a TDR, the language emerging from these documents has provided a common vocabulary for preservation. Besides lending specificity to this conversation about DNA-based data storage, they situate it within an archival praxis. Not completely irrelevant to the discussion of space, OAIS was developed to provide a model by which international space agencies could share standardized data in support of space research.<sup>129</sup> I focus on OAIS, not because of this connection to space, but because it offers a conceptual model for the ingestion, storage, and access of information into an archive, which are three aspects of digital preservation those developing DNA-based data storage must contend with.

OAIS, also referred to as the Magenta Book because of the cover color, describes a minimum-viable conceptual model for how archives interact with their environment, namely the entities of producer, management, and consumer (see fig. 4). Information is packaged with its metadata, described as a Submission Information Package (SIP) when the archive receives it from the producer, where it becomes an Archival Information Package (AIP). From there, archivists make derivatives in order to deliver that information to consumers as a Dissemination Information Package (DIP). To bridge the varied use of ‘archival’ described in the introduction

---

<sup>129</sup> Brian Lavoie, “Meeting the challenges of digital preservation: The OAIS reference model,” *OCLC Newsletter* 243, no. 26-30 (January/February 200), <https://www.oclc.org/research/publications/2000/lavoie-oais.html>.

to this chapter, the OAIS definition provides a standardized and specific outline for “archival storage”:

The Archival Storage Functional Entity (labeled ‘Archival Storage’ in the figures in this section) provides the services and functions for the storage, maintenance and retrieval of AIPs. Archival Storage functions include receiving AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which Archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders.<sup>130</sup>

What this definition tells us, first and foremost, is that archival storage is never a singular place. It is a process – a process by which archivists can continue to carry out maintenance on collections. The term ‘archival storage’ necessarily describes many different activities, some of which involve the physical space where items are stored; more critical are all of the other activities that go into managing the longevity of that package of information. Any contribution to the development of archival storage, in technology or method, must support these goals of providing function for storage, maintenance, and retrieval.

---

<sup>130</sup> *Reference Model for the Open Archival Information System (OAIS)*, 4-2.

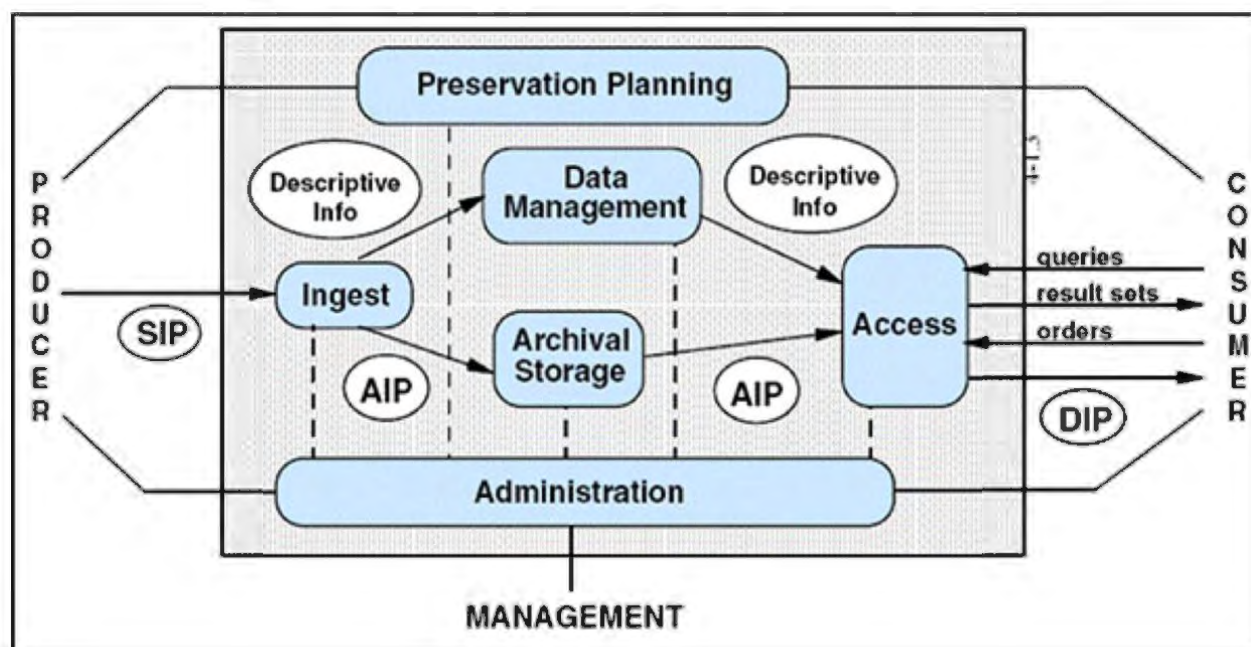


Figure 4. The Open Archival Information System conceptual model.

Often in the descriptions of DNA as a storage format, the word ‘archival’ is used in tandem with ‘long term,’ or is used in order to mean ‘long term.’ ‘Archival’ also tends to connote infrequently or never accessed data. However, as the following OAIS definition of *long term* shows, temporality is not just the amount of time a storage format can endure before its information is retrieved.

“Long Term: A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in OAIS. This period extends into the indefinite future.”<sup>131</sup>

In many ways, the OAIS definition of ‘long term’ implies many different temporalities operating within archival storage (and all of the labor implied by archival storage). The identification of an

<sup>131</sup> Reference Model for the Open Archival Information System (OAIS), 1-12.



object requiring long term care does not mean the object should be put away forever; instead, this definition of long term implies many, many short terms. Long term means actively engaging with and working within the present in order to be able to perform maintenance into the “indefinite future.” We should not take claims of long term preservation or storage as a be-all, end-all to the archive; conversely, an object requiring long term preservation implicates the archive in constantly and consistently attending to the needs of that object in the present. Objects requiring long term preservation must exist in the present, if they are to exist in the future.

The mediation of long and short has been difficult, increasingly so within digital preservation, where objects demand much more care in shorter periods of time than formats like paper or film. That is not to say film and paper do not require attention, but that the lifespan of a digital object is contingent upon the maintenance of so many dependent systems – proprietary, obsolete, or ill-attended – or upon the migration from one format to the other, that they end up being much shorter. In a study conducted by archivists from a variety of different institutions, Blumenthal et al. followed up on a survey from the NDSA showing that only a small minority of digital stewards were satisfied with the organization of current digital preservation methods. In their efforts to identify the source of this dissatisfaction through a phenomenological methodology, these researchers interviewed practitioners – here, called digital stewards. These practitioners have common understanding that digital stewardship is an “inherently long-term and open-ended pursuit.”<sup>132</sup> The researchers’ results identified seven themes practitioners noted as contributing to their dissatisfaction, the first of which was “good digital stewardship is active

---

<sup>132</sup> Blumenthal, Karl, Peggy Griesinger, Julia Y. Kim, Shira Peltzman, Vicky Steeves, “What’s Wrong with Digital Stewardship: Evaluating the Organization of Digital Preservation Programs from Practitioners’ Perspectives,” *Journal of Contemporary Archival Studies* 7, no. 13 (2020): 2, <https://elischolar.library.yale.edu/jcas/vol7/iss1/13>.

and life-long.”<sup>133</sup> More specifically, researchers identify a recurring metaphor – “care and feeding” – from practitioners that represents the maintenance of digital assets.<sup>134</sup> Not a simple deduction, the refrain of “care and feeding” of digital assets point to these dueling temporalities: that digital stewards operate in the realm of short-term actions that create present access to objects in the effort of long term preservation.

The repeated rhetoric of “care and feeding” is reminiscent of other phrases practitioners use to describe this misconception in temporalities that often occurs in the field of archiving and preservation. The researchers of the digital stewardship study quote an interviewee who says “It’s not ‘set and forget’...there is no ‘this thing is preserved’ in digital preservation.”<sup>135</sup> In a conversation on this subject, a mid-career digital preservationist referred to this phenomenon: “even if it lasts a hundred years, it’ll look stupid in ten.”<sup>136</sup> More specifically, he noted that digital preservationists only really need to worry about data surviving 5-10 years before the next preservationist, under different circumstances, can make preservation decisions. This sentiment is echoed by the practitioners in the digital stewardship study, in which practitioners discuss how the expectations of digital preservation – that objects will be preserved – often comes into conflict with financial and staffing realities. Many of these practitioners cite the lack of institutional buy-in and allowance for authority for stewards that impedes institutional support as opposed to, often, short-term funding models and project-based cycles.<sup>137</sup> What this shows us with respect to the application of DNA-based data storage is that the perception of an object’s temporality in the archive is both long and short, with respect to staff and labor.

---

<sup>133</sup> Blumenthal, et al., “What’s Wrong with Digital Stewardship,” 8.

<sup>134</sup> Blumenthal, et al., “What’s Wrong with Digital Stewardship,” 8.

<sup>135</sup> Blumenthal, et al., “What’s Wrong with Digital Stewardship,” 9.

<sup>136</sup> Interview with an archivist, 19 November 2020.

<sup>137</sup> Blumenthal, et al., “What’s Wrong with Digital Stewardship,” 11.

More specifically, an object's temporality is also mediated by its digital dependencies. The "care and feeding" practitioners use to describe their work comes in the form of numerous preservation activities such as migration, emulation, transferring, fixity checks, among others. These preservation activities are not necessarily due because of failing storage, although that certainly could be the case. What much of the papers describing the longevity of DNA-based data storage fail to acknowledge is the longevity of the digital files themselves. While DNA may last tens if not hundreds or even thousands of years, what good is that if the digital files contained within DNA are irretrievable due to format obsolescence? A way to understand how these levels of obsolescence that interact with each other is in dissecting the different dependencies a digital object has. For example, the recently patented Digital Object Obsolescence Database from Digital Bedrock tracks these vulnerabilities "related to formats, codecs, software, operating systems, hardware, and ownership."<sup>138</sup> Other common failures include media, hardware, and software failure as well as operator error, and perhaps most ignored by companies selling new technology, economic and organizational failure.<sup>139</sup> A digital object, while completely accessible today, may not be in a couple months, years, or decades because of loss of support from these dependencies. When considering the many threats to the ongoing accessibility of an object that have nothing to do with the longevity of a storage format, it becomes clear that longevity beyond the inevitable failure of a digital object due to the loss of these dependencies can be irrelevant. While DNA – and all the bits on DNA – may persist well

---

<sup>138</sup> "Format Obsolescence," *Digital Bedrock*, accessed 2 May 2021, <https://www.digitalbedrock.com/format-obsolescence>.

<sup>139</sup> David S. H. Rosenthal, Thomas Robertson, Tom Lipkis, Vicky Reich, and Seth Morabito, "Requirements for Digital Preservation Systems: A Bottom-Up Approach," *D-Lip Magazine* 11, no. 11 (November 2005), <http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html>.

into the future, those bits may be irretrievable because of the conditions of the dependencies at that future point.

### **Archival Requirements**

As the development of DNA-based data storage began in earnest, especially since the Church et al. 2012 paper, and the rise of companies such as those a part of the DNA Data Storage Alliance, they have been courting the cultural heritage institution industry. Since that attention, and as more archiving professionals become involved in DNA-based data storage, these companies have been learning about the requirements necessary to the technology in order for it to become viable. Daniel Chadash of Twist Bioscience demonstrated this interest during his presentation on the DDSA in which he acknowledged how the DDSA is taking the initiative to interview archiving professionals and build out those features. The following few paragraphs build on a presentation given by Linda Tadic of Digital Bedrock on what those necessary features are: random access, interoperability, cost, speed, sustainability, error resilience, and fixity. Very much like the qualities of DNA-based data storage discussed in Chapter One, the discussion of these minimum requirements also references the archival reasoning behind these requirements. While archivists should participate in the development of these necessary developments, archivists have also long been wary of new formats that claim “immortal media.”<sup>140</sup> Rosenthal, a digital preservationist, lists numerous issues with immortal media in general: first that these formats are incredibly expensive compared to others; second, as an offline format, DNA is at best a second copy; and third, a format that cannot be edited and has a very long access latency is not appropriate for preserving data that will become obsolete. The following requirements aim to

---

<sup>140</sup> David Rosenthal, “More on long-lived media,” 19 June 2014, <https://blog.dshr.org/2014/06/more-on-long-lived-media.html>.

counter these criticisms of immortal media, but if it is to become a commercially scalable medium that archivists want to use, DNA will have to market itself beyond its immortality.

The first necessary requirement for DNA-based data storage as an archival storage format is interoperability. Interoperability means that DNA produced by one company should be able to be read by any read software and hardware. In Chadash's presentation to the JPEG DNA working group, he lists the development of LTO as a model to look to.<sup>141</sup> Tadic also mentions Linear Tape File System (LTFS) as an open standard DNA-based data storage developers could look to.<sup>142</sup> Up until the development of LTFS in 2010, an LTO tape sold by HP could not be read by a deck sold by IBM, which allowed for multivendor architectures and the migration of data from one vendor to another.<sup>143</sup> Without this level of interoperability, the benefits of DNA marketed such as longevity are useless, and DNA can quickly become obsolete as a data storage format. The DDSA represents a promising effort in building interoperability between vendors from an early stage, setting a standard that will be governed by a diverse body of users and vendors. Interoperability of formats is absolutely crucial to archives, which must strive to be independent of the rise and fall of companies, open source communities, and lack of technological buy in.

What LTFS also offered for LTO was the separation of metadata from respective digital objects, which enabled an emulation of random access.<sup>144</sup> As mentioned in Chapter One, the development of random access technology will be absolutely imperative to the success of DNA-based data storage. What DNA developers can learn from LTFS is in data structures and

---

<sup>141</sup> Chadash, "An Overview of the DNA Data Storage Alliance."

<sup>142</sup> Tadic, "Audiovisual digital preservation."

<sup>143</sup> "The History of Magnetic Tape and Computing: A 65-Year-Old Marriage Continues to Evolve," *Iron Mountain*, accessed 2 May 2021, <https://www.ironmountain.com/resources/general-articles/t/the-history-of-magnetic-tape-and-computing-a-65-year-old-marriage-continues-to-evolve>.

<sup>144</sup> "The History of Magnetic Tape and Computing," *Iron Mountain*.

encoding, but what they should look to as format specific is the biological power of DNA computing in enzymatic processes. If DNA is to store immense amounts of data, it must offer single file retrieval and indexing. Tadic writes that “we need to be able to retrieve/restore just that one file, not a 2 PB string of data just to get to a 1 MB file.”<sup>145</sup> This demand also raises an important aspect of DNA-based data storage that has not been addressed by scientists: that DNA is a write-once technology, meaning data cannot be overwritten or edited once it is encoded. Given the great expense of storing data on DNA currently, the inability to modify files once they have been stored will also be a difficult hurdle to overcome. This inability to modify files can also introduce issues as archivists are used to performing fixity checks on data regularly, and will continue to need to do so even if that data is on DNA. Archivists are a skeptical people and will not take the word of a molecular biologist that the data is safe. As DNA is offline and biological, we must ask how checksums – hashes unique to a file – will be encoded and recorded. Additionally, will the validation of these checksums require an enzymatic process? As mentioned earlier, archivists cannot “set it and forget it,” but must continue to “care and feed,” even if digital objects are on DNA.

As many technologists and advocates for DNA-based data storage have discussed, cost will probably be the greatest barrier towards implementing DNA at scale commercially. Because of the lack of migration, restoring, or refreshing available in DNA-based data storage – that DNA is write once – many would expect the costs to be lower than the current market for LTO. Additionally, because of the expertise involved in synthesizing and sequencing DNA – and therefore the current inability of archivists to perform their own preservation work in-house – DNA will only become viable commercially to archives once it is competitive with or cheaper

---

<sup>145</sup> Tadic, “Audiovisual digital preservation.”

than current physical storage media, like data tape. Deemed by many to be “not a good business,” archival storage media can only function at the very bottom of a tiered data storage hierarchy and is a market where “customers are reluctant to spend, because there is no return on the investment.”<sup>146</sup> DNA must be significantly cheaper than LTO. On the positive, many struggle with LTO because of the migration and perceived lack of backwards compatibility – an issue those buying into DNA-based data storage will not have. Factoring into the cost of using DNA will also be the speed at which files are written and read – Tadic requires “at minimum one Gigabyte per second.”<sup>147</sup> Additionally, the storage environment must be reasonable; one of the greatest costs for current data storage are the cool environments necessary to maintain servers, hard drives, and other storage formats for a long period of time.

Lastly, sustainability is a key minimum requirement in order for DNA-based data storage to become competitive to archives. Sustainability can mean many things for an archivist choosing to buy into a new technology, but all of these factors interact with each other. From being a green technology, to being financially sustainable, to being a format that keeps over time from community buy in and userbases, sustainability will affect DNA’s attractiveness to archives. Currently, those selling DNA-based data storage argue that DNA will be a good ecological alternative to current formats due to its composition and lack of rare earth metals. However, we must also consider the energy and chemical use of writing and reading data as well as the material in which DNA is encapsulated. Archivists must remain skeptical of the marketing and descriptive language used in discussions of DNA, and in order to do so, we must understand how DNA works and how it could be integrated into current storage systems. Where in a tiered

---

<sup>146</sup> David Rosenthal, “Archival Media: Not a Good Business,” *DSHR’s Blog*, 1 March 2018, accessed 2 May 2021, <https://blog.dshr.org/2018/03/archival-media-not-good-business.html>.

<sup>147</sup> Tadic, “Audiovisual digital preservation.”

storage plan could DNA fit? Large archives with audiovisual digital objects, which can represent very large files, might want to incorporate DNA as a last-resort geographically separated solution. With large costs, a lack of standardization, and the middling stages of technological development, DNA is not a storage format ready to be implemented at scale today. That being said, DNA-based data storage does offer a promising and attractive solution to long-term preservation; but digital stewards, archivists, and administrators must recognize where DNA fits in their own archival storage systems and must not neglect the need to “care and feed” their digital objects.

## **Containers**

DNA-based data storage raises myriad questions with respect to time and labor. The application of archival media in the archive – that is, media that markets itself as “set it and forget it” – can be a point of contention for digital stewards whose labor is markedly opposed to permanent storage solutions. The issue of temporality and the mediation between human lifetime and the goal of “indefinite future” lifetimes for digital objects, not to mention those of institutional memory, retirement, and staff turnover, complicate the vision of DNA-based data storage. What DNA-based data storage also pushes archivists to consider, as these conflicting views of media storage formats collide, is what containment of a cultural heritage object means. Here, a brief look at the history of genomics and the development of DNA as a metaphor for life intermingles with views of storage as a dynamic or static quality in the archive. Reading Zoë Sofia’s theory of container technologies alongside a variety of scholars’ discussion on the perceived operability of DNA – as code, as life, as metaphor – I briefly dissect how the rendering of archival storage as static and forever conflicts with the labor of containers.



OAIS and the digital stewards interviewed by Blumenthal et al. view archival storage as necessarily dynamic, evolving, and changing through the labor of archivists, the maintenance of object integrity, and the delivery of access to a user base. Archival storage, therefore, is constantly producing work. In her discussion of containment, Sofia grounds her article within a feminist framework. She does so, not by pursuing an analysis of anti-maternal or reproductive labor bias, but by instead analyzing how containers obscure their labor. She argues that the “indebtedness to the spatial/maternal environment and the labors of those who sustain this facilitating space” is often ignored, or “considered ‘menial’ because they do not produce some dynamic and heroically discovered object to be wondered at, but reproduce an unobtrusively and incrementally ordered space which can be taken for granted as a background for other activities.”<sup>148</sup> Reading this essay, which focuses on all manners of containers, from jugs and ovens, to media apparatuses, the connection to archival storage is resounding. The perception of archival storage as static, unobtrusive, unchanging, or simply waiting to be plundered mirrors the devaluation of stabilizing processes. Stabilizing processes, like that of continually making digital objects orderly and accessible in the face of continued disorder, change, and obsolescence, are masked by the very containers on which they reside.

In library sciences, different conceptual models – such as OAIS – provide frameworks within which archivists can describe, organize, and maintain objects. One such model is the Functional Requirements of Bibliographic Records (FRBR; pronounced ‘ferber’), which specifies when to use words like “work,” “object,” “edition,” or “item.” One example of FRBR is to define a Work as a “distinct intellectual or artistic creation,” which is realized an Expression, or “the intellectual or artistic realization of a Work,” which is embodied in a

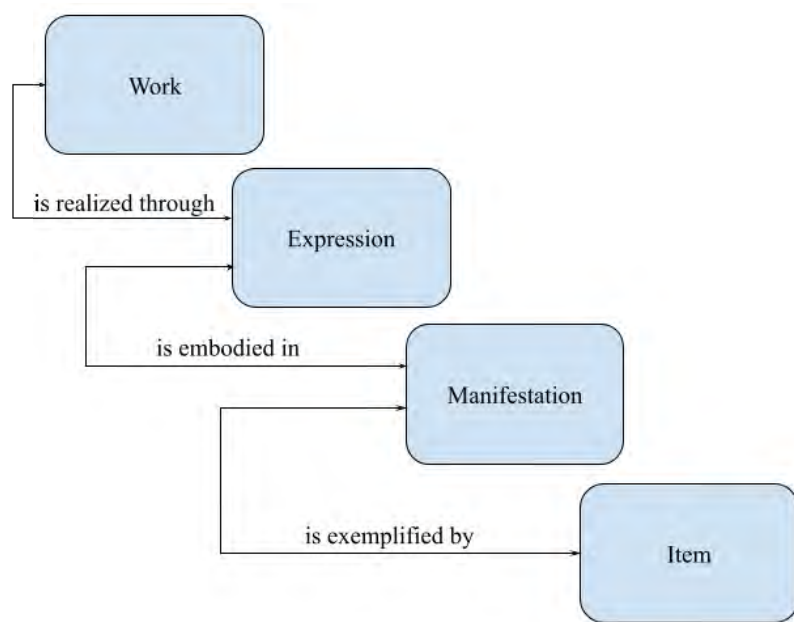
---

<sup>148</sup> Zoë Sofia, “Container Technologies,” *Hypatia A Journal of Feminist Philosophy* 15, no. 2 (Spring 2000): 188, <https://muse.jhu.edu/article/14107>.

Manifestation, “the physical embodiment of an Expression of a Work,” which is subsequently exemplified by an Item – “a single exemplar of a Manifestation” (see fig. 5).<sup>149</sup> To FRBRize a collection is to describe each aspect of a resource through these different entities. For example, *Microvenus* is a conceptual work realized through either a binary string or a sequence of nucleotides, embodied in a bacterium, exemplified by the specific *E. coli* bacterium Joe Davis worked with in 1988. FRBRization helps librarians understand the different levels at which preservation work must take place or at which a resource can be described. I bring up FRBR in order to demonstrate the different levels at which librarians interact with storage in order to pinpoint how Works, Expressions, Manifestations, and Items can be separated from each other, re-order, and reconfigured. Moreover, separating an intellectual resource into these components emphasizes all of the different points at which a resource faces a threat of obsolescence: at the DNA, at the bit, at the codec, or even at the intellectual level.

---

<sup>149</sup> IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records*, last modified February 2009, accessed 3 May 2021, 13, [https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)



*Figure 5. FRBR Group 1 Entities.*

That these entities can be separated from each other, migrated, transferred, emulated, and mapped onto different manifestations and expressions of the work forms the foundation of much of the labor of digital preservation. That they can be disassociated from each other leaves them both vulnerable and malleable in ways that are beneficial to the archivist's paradigm. Still, we should challenge that the separation and reconfiguration is without trace, without effect. When a process, such as these preservation actions, is meant to cover up its own work, a remnant of that process is the trace of labor performed. As she continues this discussion of the description of containers as static, in opposition to dynamic, Sofia writes that "it could be debated whether holding or containing is simply to be considered as a passively inhering property of a shaped space, or whether containing is rather to be thought of as a form of action in itself."<sup>150</sup> If containing is doing, as Sofia presupposes, that obfuscation of dynamism for static inactivity is

<sup>150</sup> Sofia, "Container Technologies," 190.

“encouraged by the unobtrusiveness of containers, traces of whose productive roles are not necessarily evident in the final product.”<sup>151</sup> However, in digital preservation, where the dynamic labor is constantly that of preventing against change in integrity, the traces are evident. They are evident in our storage decisions, in the labor done to maintain or upheave those decisions. Specifically, they *should* be visible in the preservation documentation – potentially in a Preservation Metadata Maintenance Activity (PREMIS) standardized log that accompanies an object. With DNA, the choice to store onto DNA at first will not be mandated by necessity. It will be a choice of container – suggested as “not just about what holds or houses us, but what we put our stuff into, and thereby identify with; what of ourselves we can and cannot contain.”<sup>152</sup> Sofia claims a value judgment upon our containers, a value judgment that collides directly with understandings how DNA functions in our bodies, and now, in our storage.

Taking container technologies into account is not new for archives. For audiovisual archivists in particular, maintaining collections has meant contending with the ways different storage formats affect that media. For example, at the Association of Moving Image Archiving conference in 2020, CK Ming facilitated a roundtable for audiovisual archivists to discuss inherent bias in preserving skin tones. Citing how “Broadcast range values are based on film color cards, but as leader ladies have shown us both film and video color and lighting values don’t account for darker skin tones,” the roundtable sought to invite a small group of audiovisual archivists to discuss the practice of accurately capturing skin tone values.<sup>153</sup> The inability for many photographic media to capture darker skin tones accurately has been discussed by many

---

<sup>151</sup> Sofia, “Container Technologies,” 198.

<sup>152</sup> Sofia, “Container Technologies,” 185.

<sup>153</sup> CK Ming, “Roundtable: Out of Range: Inherent Bias in Preserving Skin Tones,” (Webinar, Association of Moving Image Archivists, 17 November 2020), <http://www.amiaconference.net/preliminary-program-6>.

media scholars, including Lorna Roth, whose study, “Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity,” explore the ways in which media technologies carry biases imparted to them by their creators and stewards. She writes that “if the social machine manufactures representations, it also manufactures *itself* from representations, the latter operative at once as means, matter, and condition of sociality.”<sup>154</sup> Roth identifies how the bias developed within the chemistry of photographic film has played not just in the formation of beauty standards but also in the co-production of social cognition and technology. Other scholars (Tara McPherson, “US Operating Systems at Mid-Century: The Intertwining of Race and Unix,” 2011; Lisa Nakamura, “Indigenous Circuits,” 2011) have written about similarly visually ingrained biases into the very fabric of these technologies. What, then, if anything, does DNA do to the media soon to be stored on it?

Photochemical film and analog video, computer software and hardware, and – as the following chapter aims to demonstrate – even the technology of DNA are embedded with social, political, and economic structures that reflect on their usage as well as the items that reside within these systems. Especially in the realm of “immortal media,” where options abound in the form of glass, film, and DNA, if the choice of format reflects on an identification *with* the container, archivists should inquire their relationship with that identity. In an object’s life in storage, all of these choices become part of its meta-narrative, its story of preservation. Each preservation action, then, becomes tithed to the intellectual resource itself. Even though, as FRBR shows us how intellectual resources can be cleaved to these decisions and then separated

---

<sup>154</sup> Lorna Roth, “Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity,” *Canadian Journal of Communication* 34, no. 1 (2009): 127, <https://doi.org/10.22230/cjc.2009v34n1a2196>.

from them, with the proper documentation, archivists wield a power to affect perception of materials through these choices.

### Chapter Three: A Case Study for Audiovisual Archives

On September 25, 2020, the National Film and Sound Archive of Australia (NFSA) announced via a press release that their institution would be the first to preserve a film onto synthetic DNA. Along with their partner, the Olympic Foundation for Culture and Heritage (OFCH), the NFSA chose the footage of Aboriginal-Australian athlete Cathy Freeman winning the 400-meter race at the 2000 Sydney Olympics. Funded partially by financial and in-kind donations from the International Olympic Committee (IOC) and Australia Olympic Committee (AOC), this pilot project ushered DNA-based data storage into the field of archiving at audiovisual memory institutions.<sup>155</sup> Jan Müller, under whose leadership the DNA project was conducted, became the CEO of the NFSA in 2017 with the goal of pushing the archive into the digital future.<sup>156</sup> “We are not an archive anymore, we are an IT company,” announced Müller, as he unveiled a five-year strategy for the archive during his first year.<sup>157</sup> During his tenure, Müller also published the NFSA’s new Strategic Vision Plan and Digitisation Strategy, both of which detail the archive’s belief that “the future of audiovisual archives is digital.”<sup>158</sup> The Digitisation Strategy details that, while they currently store data on commonly used LTO, “the NFSA needs to look at options to store digital collection material” like, for example, cloud services.<sup>159</sup> Later in the document, the NFSA writes that it will “review the use of existing and new technology throughout the life of

---

<sup>155</sup> Linda Morris, “In our DNA: Cathy Freeman’s Olympic win projected onto Opera House,” *The Sydney Morning Herald*, 25 September 2020, <https://www.smh.com.au/culture/art-and-design/in-our-dna-cathy-freeman-s-olympic-win-projected-onto-opera-house-20200922-p55xzx.html>.

<sup>156</sup> Müller resigned from his position as CEO of the NFSA amidst the COVID-19 pandemic on 10 December 2020 to return to the Netherlands, where his family resides: <https://www.if.com.au/nfsa-ceo-jan-muller-resigns>.

<sup>157</sup> Helen Musa, “Muller puts the ‘N’ back into the NFSA,” *City News*, 31 May 2018, <https://citynews.com.au/2018/muller-puts-the-n-back-into-the-nfsa>.

<sup>158</sup> National Film and Sound Archive of Australia, “NFSA: Strategic Vision for a Digital Archive,” 2018, accessed 2 May 2021, <https://www.nfsa.gov.au/collection/curated/nfsa-strategic-vision-2018>.

<sup>159</sup> National Film and Sound Archive of Australia, “Digitisation Strategy 2018-2025,” 2018, accessed 2 May 2021, 10, <https://www.nfsa.gov.au/collection/curated/nfsa-digitisation-strategy-2018-2025>.

this Strategy,” indicating the archive’s intention to look outside the immediately available options for data and digitizing solutions.<sup>160</sup> It is in the context of the NFSA’s commitment to being a “digital leader” that they pursued the collaboration with OFCH and Twist Bioscience, the California Silicon Valley biotechnology start-up that performed the work of synthesizing the DNA strand onto which Freeman’s winning race now resides.

Selected, presumably, because of the collaboration with the OFCH and the funding dynamics with the IOC and AOC, the footage of Cathy Freeman’s race also holds great importance in the cultural memory of Australia. Freeman was the first Aboriginal athlete to win the gold medal in an individual event. Likening its own data storage first to Freeman’s athletic feat, the NFSA wrote of the footage: “an iconic moment which has metaphorically become part of Australia’s DNA is now stored on actual DNA.”<sup>161</sup> The confluence of metaphor and material in this statement – “Australia’s DNA” and “actual DNA” – seems to be the justification for this curatorial choice. The metaphoric power of DNA is often harnessed to connect cultural or ancestral heritage to conceits of nation and race using scientific language. DNA’s materiality is now physically bound up with the imagery of Cathy Freeman. At the same time, this metaphor of Australia’s DNA cannot be separated from the ongoing histories of fraught management of Indigenous materials in the national archive. In this chapter, I seek to unravel the ways in which the storage of cultural heritage onto DNA in national memory institutions intersects with critical discussions in the archival field of the stewardship of Indigenous collections. In order to do so, I look to DNA’s participation in racializing projects like the migration and genetic diversity offshoots of the Human Genome Project that have exploited Indigenous genetic material. If, as

---

<sup>160</sup> National Film and Sound Archive of Australia, “Digitisation Strategy 2018-2025,” 11.

<sup>161</sup> Miguel Gonzalez, “Technological World First for NFSA,” National Film and Sound Archive, 25 September 2020, accessed 8 January 2020, <https://www.nfsa.gov.au/latest/nfsa-dna>.



the NFSA wishes to indicate, DNA will become a primary form of moving image data storage in the near future, archival institutions must contend with the biases that a 50-year history of genomic study has imbued into the medium itself. Archives must acknowledge DNA as a technology of colonization and of race, not only of digital data storage.

### **Cathy Freeman and Australia's DNA**

In the NFSA's Charter of Curatorial Values, written by Paolo Cherchi Usai in 2006, Cherchi Usai writes that, while the curator may not have the "overwhelming power" of history as a "selective, powerful and often unforgiving curator of the cultural heritage," their responsibility is "to decide what should be preserved first."<sup>162</sup> In selecting this footage of Cathy Freeman for the pilot project, the NFSA has made a political decision to champion this moment as emblematic of Australia, a part of Australia's DNA. Many Black Indigenous athletes – including Freeman herself, in years leading up to the Olympics – suffered public retribution for their statements about their race and heritage during athletic events. In contrast to controversial proclamations like these, Freeman carrying both the Aboriginal and Australian flags during her victory lap was described as "bringing the nation together" both in contemporary press<sup>163</sup> and twenty years later in the NFSA's own description of the event.<sup>164</sup> In framing Freeman's accomplishment as a nation-building moment that unified the country at a time of immense political and social discord

---

<sup>162</sup> Cherchi Usai, Paolo, "A Charter of Curatorial Values," *Journal of the National Film and Sound Archive*, Australia 1, no. 1 (Spring 2006), 8.

<sup>163</sup> *Herald Sun*, 5 October 2000, 4, cited in Gary Osmond and Matthew Klugman, "A Forgotten Picture: Race, Photographs and Cathy Freeman at the Northcote Koori Mural," *Journal of Australian Studies* 43, no. 2 (2019): 204, <https://doi.org/10.1080/14443058/2019.1581247>.

<sup>164</sup> Miguel Gonzalez, "Technological World First for NFSA," National Film and Sound Archive, published 25 September 2020, accessed 8 January 2020, <https://www.nfsa.gov.au/latest/nfsa-dna>.

between Indigenous- and settler-Australians, the NFSA has reinforced a narrative that belies the complexity of the Sydney Olympics and the politic of reconciliation it championed.

In the decade leading up to her gold medal win in 2000, Freeman not only trained as an athlete but also as an Aboriginal activist. She wielded her growing fame over the years to raise consciousness for the struggles and histories of her ancestors in the national and international arena. In the years since her Olympic win, scholars have discussed these mirrored histories in light of each other as well as the ways in which Freeman's activism was contrasted with her contemporaries'. In her discussion of why national and international structures deem some political actions appropriate and others not, Christine O'Bonsawin writes that "there exist decisive manifestations that continue to privilege the interests of colonial nation-states that are responsible for the historic sufferings and ongoing oppression of Indigenous peoples."<sup>165</sup> Actions condoned or even facilitated by settler states represented by Black and Indigenous athletes may be similar or the same as actions condemned by the same state. The reception of political activism relies on how much that nation-state can benefit from it. Freeman's victory lap in 2000 was applauded by Australia despite her other, similar actions that were either forgotten by the public or condemned by the state. Analysis of how Freeman's Olympic win came to be construed as "uniting the nation" must contend with Australia's relevant political history and the context of Freeman's legacy as an activist.

Throughout her athletic career, Freeman was consistently portrayed as naïve, unaware of her own political activism.<sup>166</sup> A closer look at her athletic history reveals the opposite: Freeman,

---

<sup>165</sup> Christine O'Bonsawin, "From Black Power to Indigenous Activism: The Olympic Movement and the Marginalization of Oppressed Peoples (1968-2012)," *Journal of Sport History* 42, no. 2 (Summer 2015): 201, <https://www.jstor.org/stable/10.5406/jsporthistory.42.2.0200>.

<sup>166</sup> Osmond and Klugman, "A Forgotten Picture," 210.

who self-describes as a “proud Kuku Yalanji woman,”<sup>167</sup> consistently brought her Aboriginal heritage into the arena visually,<sup>168</sup> made public statements about her family history and the “Stolen Generation,”<sup>169</sup> and sought out media opportunities to make visible the painful and violent history of Australia’s treatment of Aboriginals and Torres Strait Islanders.<sup>170</sup> Close reading of her activism shows Freeman working, almost by trial-and-error, to usher the Aboriginal flag and people into the sports arena, and therefore into the public eye. In athletic meets early in her career, Freeman wore the Aboriginal flag as ornamentation on her wrist- and headbands.<sup>171</sup> In a foreshadowing of her victory lap in 2000, after she won the women’s 400-meter event at the 1994 Commonwealth Games in Victoria, Canada, Freeman first took a lap with only the Aboriginal flag and then took a second lap with both Aboriginal and Australian flag. Almost immediately publicly excoriated by Australian government officials for putting her indigeneity before nation, Freeman made an adjustment when, a few days later, she won the 200-meter.<sup>172</sup> She picked up both flags at the same time for a single victory lap, and the resulting images were called “Cathy’s uniting double,” setting a precedent for becoming the “embodiment of reconciliation” in 2000.<sup>173</sup>

These images, of Freeman holding both Aboriginal and Australian flags, were lauded by the wider Australian public. If her celebration of her Aboriginal identity outshined her patriotic duty as an athlete, she was derided or disregarded. That same year, Freeman devised another

---

<sup>167</sup> CathyFreeman.com.au, “Cathy Freeman,” *Estrella Sports Management*, accessed 19 April 2021, <https://estrellasports.com.au/portfolio/items/cathy-freeman/>.

<sup>168</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 205.

<sup>169</sup> David Mayeda, “Characterizing Gender and Race in the 2000 Summer Olympics: NBC’s Coverage of Maurice Greene, Michael Johnson, Marion Jones, and Cathy Freeman,” *Social Thought & Research* 24, no. 1/2, *The Politics of Gender* (2001): 169, <https://www.jstor.org/stable/23250078>.

<sup>170</sup> Osmond and Klugman, “A Forgotten Picture,” 208.

<sup>171</sup> Osmond and Klugman, “A Forgotten Picture,” 211.

<sup>172</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 205.

<sup>173</sup> Osmond and Klugman, “A Forgotten Picture,” 204.

stunt, creating an image that should have become iconic, but was instead ignored and then forgotten. In order to make their argument about acceptable imagery in the eye of the settler state, Gary Osmond and Matthew Klugman turn to a photograph Freeman staged in front of a mural depicting a famous 1906 photograph of Aboriginal elders in chains that was organized by the local Aborigines Advancement League.<sup>174</sup> A striking photograph, Freeman runs in front of the mural of Australia's colonial past in a present where inequity between white settlers and Aboriginals was rampant. In this image, Freeman was able to make that violent history a part of the present instead of moving past it, as Australian reconciliation sought to do. Here, there is no denial of the past in favor of a bright future. Osmond and Klugman describe how this picture, arguable as powerful as the flag images, never endured in public memory. They attribute this phenomenon to the ways it went against the public conception of Freeman as an innocent and apolitical character, evidenced by the way the newspaper re-printing the image without the original caption in order to lesser its political impact as well as Freeman's authorial motivations.<sup>175</sup> Osmond and Klugman analyze the gendered and racialized language used to describe Freeman's personality, as does David Mayeda, who analyzes how contemporary sports commentary "pacify her" in order to make her a figurehead for the Australia.<sup>176</sup> The stunt Freeman created with this image did not fit into that narrative and therefore never became a part of it.

These constructions of Freeman as a non-participant in her own activism and thus an acceptable representation of Aboriginal activism by the settler state run in contrast to the reception of similar actions by contemporary athletes. Freeman, because of her gender, perhaps

---

<sup>174</sup> Osmond and Klugman, "A Forgotten Picture," 206.

<sup>175</sup> Osmond and Klugman, "A Forgotten Picture," 209.

<sup>176</sup> Mayeda, "Characterizing Gender and Race in the 2000 Summer Olympics," 171.

because of her sport, was framed as “shy” and “naïve.” She is often compared to footballer Nicky Winmar, whose 1993 response to racist remarks on the pitch was to lift his shirt and point at his skin, declaring “I’m black and I’m proud to be black.”<sup>177</sup> Not coincidentally, Freeman hired the same photographer who took this picture, Wayne Ludbey, to shoot her running in front of the mural. Perhaps more striking because of the similarities, boxer Damien Hooper, of Kamilaroi and Manandanyi ancestry, wore an Aboriginal flag on his shirt and faced accusations from the IOC and AOC of breaching Rule 50, the same rule Freeman ostensibly broke, which “forbids any kind of political, religious, or racial demonstration inside an Olympic venue.”<sup>178</sup> Essentially the same political action Freeman took twelve years earlier caused the AOC to chastise and reprimand Hooper.<sup>179</sup> O’Bonsawin writes of the incident: “in the Olympic domain, Indigenous athletes, such as Hooper, have to accept the identity of a colonizing settler citizenry, thereby further validating the political authority of an illegally imposed governing structure.”<sup>180</sup> That governing bodies decide what activism is appropriate based on political gain means that athletic activism cannot rely on precedents set by previous athletes to judge what will be approved of and what will not. The hypocrisy of one person’s use of the Aboriginal flag being cause for reprimand while the other’s is a cause for celebration can be parsed through Australia’s desire to incorporate reconciliation into the Sydney Olympics.

As Cathy Freeman’s career took off in the 1990s, building in momentum towards the new millennium, the nation of Australia was facing its history of colonialism and racism in courts, on

---

<sup>177</sup> Michael Gordon, “The day Nicky Winmar drew the line,” *The Age*, 16 April 2013, accessed 2 May 2021, <https://www.theage.com.au/sport/afl/the-day-nicky-winmar-drew-the-line-20130416-2hydd.html>.

<sup>178</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 201.

<sup>179</sup> Karen Barlow, “Boxer Hooper apologises for Aboriginal shirt,” *ABC News*, 30 July 2012, accessed 2 May 2021, <https://www.abc.net.au/news/2012-07-31/boxers-aboriginal-t-shirt-draws-ire-of-officials/4165260>.

<sup>180</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 201.

the streets, and in the general public. The Indigenous rights movement during this period had pressured the government past the point of ignorance. A year after Freeman became the first Aboriginal-Australian to win a Commonwealth Games gold medal in 1990, the Australian parliament established the Council for Aboriginal Reconciliation in response to the Royal Commission into Aboriginal Deaths in Custody.<sup>181</sup> Over the next ten years, the Council was tasked with promoting “a process of reconciliation between Aboriginal and Torres Strait Islander peoples and the wider Australian community.”<sup>182</sup> *Mabo v Queensland (No 2)* decision in 1992 rejected the *terra nullius* doctrine and led to the 1993 Native Title Act which began to recognize certain Aboriginal and Torres Strait Islander rights to land.<sup>183</sup> In 1997, the Human Rights and Equal Opportunity Commission published its report, *Bringing Them Home*, which detailed the history of Australia’s forcibly removing children from their families between 1910 and 1970.<sup>184</sup> The history of the Stolen Generations was highly contested and denied by the government at the time. A year later, the first National Sorry Day – later renamed the National Day of Healing for All Australians – was held, for “Australians [to] express regret for the historical mistreatment of Aboriginal people.” At the same time, the Native Title Amendment Act was enacted in the absence of Indigenous voices.<sup>185</sup> In April 2000, a United Nations Committee on the Elimination of All Forms of Racial Discrimination (CERD) report criticized the treatment of Indigenous

---

<sup>181</sup> William Jonas, “Chapter 2: Reconciliation and human rights,” in *Social Justice Report 2000*, Australian Human Rights Commission, 1 January 2000, <https://humanrights.gov.au/our-work/social-justice-report-2000-chapter-2-reconciliation-and-human-rights>.

<sup>182</sup> Council for Aboriginal Reconciliation, “Vision Statement,” *Australasian Legal Information Institute*, accessed 18 April 2021, [http://www5.austlii.edu.au/au/orgs/car/council/spl98\\_20/council.htm](http://www5.austlii.edu.au/au/orgs/car/council/spl98_20/council.htm).

<sup>183</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 207.

<sup>184</sup> Human Rights and Equal Opportunity Commission, *Bringing Them Home: National Inquiry into the Separation of Aboriginal and Torres Strait Islander Children from Their Families*, Commonwealth of Australia, 1997, [https://humanrights.gov.au/sites/default/files/content/pdf/social\\_justice/bringing\\_them\\_home\\_report.pdf](https://humanrights.gov.au/sites/default/files/content/pdf/social_justice/bringing_them_home_report.pdf).

<sup>185</sup> SBS News, “Australia marks 20-year anniversary of Sorry Day,” *SBS News*, 26 May 2018, <https://www.sbs.com.au/news/australia-marks-20-year-anniversary-of-sorry-day>.

Australians on the eve of the Olympics.<sup>186</sup> A month later, hundreds of thousands of Indigenous and non-Indigenous Australians marched at the People's Walk for Reconciliation in order to protest the lack of a formal apology from the government for the Stolen Generations.<sup>187</sup>

The People's Walk for Reconciliation took place the day after the Council for Aboriginal Reconciliation had released its final recommendations in the "Australian Declaration Towards Reconciliation" and "Roadmap for Reconciliation," which were expected to be received by the government with a formal response. Instead, Prime Minister John Howard disagreed with the Declaration and made a public statement regarding the "areas of difference relating to customary law, the general application of the laws of Australia to all citizens, self-determination and a national apology as distinct from an expression of sorrow and sincere regret."<sup>188</sup> Howard, as the opposition leader a decade earlier, had championed a "One Australia" policy that railed against multiculturalism, immigration, and treaties with Aboriginal Australians. As prime minister, he had cut \$A400 million from the Aboriginal affairs budget.<sup>189</sup> Reconciliation encompasses many different legal, political, and social justice frameworks. In her analyses of the various definitions of reconciliation, Kim TallBear notes that "settler-colonial definitions" such as "the restoration of friendly relations" wrongly imply that there once were friendly relations.<sup>190</sup> Others, such as "the action of making one view or belief compatible with another" – which is one that she claims

---

<sup>186</sup> O'Bonsawin, "From Black Power to Indigenous Activism," 208.

<sup>187</sup> SBS News, "Australia marks 20-year anniversary of Sorry Day."

<sup>188</sup> John Howard, quoted in William Jonas, "Chapter 6: Reconciliation – National progress one year on," in *Social Justice Report 2001*, 1 January 2001, <https://humanrights.gov.au/our-work/aboriginal-and-torres-strait-islander-social-justice/publications/social-justice-report-9>.

<sup>189</sup> John Pilger, "Cruelty and xenophobia stir and shame the lucky country," *The Guardian*, 18 January 2017, accessed 2 May 2021, <https://www.theguardian.com/commentisfree/2007/jan/19/australia.guantanamo>.

<sup>190</sup> Kim TallBear, "Diversity v. Decolonization in the Academy: A Conversation with Kim TallBear," (Zoom Lecture, McGill University Department of English, 10 February 2021), Video, 53:30, <https://www.youtube.com/watch?v=qrL5Kc4wJgQ>.

persists in the general public in Canada – do not contend with the disparate belief systems between the settler state and Indigenous peoples. Moreover, she claims, in line with work by Adam Gaudry and Danielle Lorenz, reconciliation is the work done to find common ground between different epistemological frameworks.<sup>191</sup> This labor, which policies like “One Australia” work against, cannot be accomplished by universalizing language that seeks to flatten differences in favor of one culture and one nation.

In the midst of this pervasive cultural discourse and pressure to bring true equity, justice, and reconciliation to the Aboriginal and Torres Strait Islander peoples, Australia found itself, in 2000, on the international stage hosting the Olympics. The Council for Aboriginal Reconciliation was well aware that their final year coincided with this event, anticipating it: “During this time, Australia will be subject to intense global interest. Attention will be given to the progress of relationships between Indigenous peoples and the wider community.”<sup>192</sup> Due to the consistent media depictions of Freeman as shy and politically innocent – despite her releasing statements that April deriding the government’s claims that the “Stolen Generation” was nonexistent<sup>193</sup> - the Sydney Olympics and contemporary news reports propagandized Freeman as a “symbol that stood for both Aboriginal and non-Aboriginal Australians” (emphasis by author).<sup>194</sup> Freeman is an athlete who consistently engages in the painful, colonial past of Australia and its treatment of her people, but her agency in that activism was subsumed into the state-approved running “for all

---

<sup>191</sup> Adam Gaudry and Danielle Lorenz, “Indigenization as inclusion, reconciliation, and decolonization: navigating the different visions of indigenizing the Canadian Academy,” *AlterNative: An International Journal of Indigenous People* 14, no. 3 (2018), <https://doi.org/10.1177/1177180118785382>.

<sup>192</sup> Council for Aboriginal Reconciliation, “About the Council for Aboriginal Reconciliation,” *Australasian Legal Information Institute*, accessed 19 April 2021, <http://www5.austlii.edu.au/au/orgs/car/council/about.htm>.

<sup>193</sup> Mayeda, “Characterizing Gender and Race in the 2000 Summer Olympics,” 169.

<sup>194</sup> Mayeda, “Characterizing Gender and Race in the 2000 Summer Olympics,” 170.



Australia.”<sup>195</sup> The authors who discuss the afterlife of Freeman’s Olympic actions in the context of the larger Indigenous rights movement and her career argue that the portrayal of her victory lap as “the embodiment of reconciliation” belies the precarity Australia faced at the Olympics. Instead, argues O’Bonsawin, the “tokenistic strategy” of pushing Freeman into the Olympic limelight was “used to censor truths of ongoing political oppression and racial tyranny directed toward Indigenous peoples of Australia.”<sup>196</sup> The nationalistic media of the time – the same outlets which had suppressed Freeman’s more radical activism – framed reconciliation within the patriotism Freeman’s win represented for Australia, encompassing her Aboriginal identity instead of engaging with it.

Why engage in an extended discussion of Cathy Freeman, her activism, and its afterlife in the public imaginary twenty years after her “golden win”? Just as archivists are the stewards for the physical longevity of items under our care, we must also consider the ways our preservation strategies and decisions affect those materials. The slogan of “running for all Australia” subsumes Freeman’s activism on behalf of her family, ancestors, land, and culture into Australia’s John Howard-approved project for reconciliation, a much less radical racial project than Freeman’s and the Council for Aboriginal Reconciliation’s. “Running for all Australia” – eerily reminiscent of “One Australia” – is what “metaphorically [became] part of Australia’s DNA,” not Cathy Freeman’s long history of aggressive athletic activism. Here exists a double meaning: Australia’s DNA is not only the metaphoric cultural memory of history, colonialism, violence, reconciliation, but also the co-option of genetic material of Indigenous Australians. The next section explores how these two meanings of DNA – of memory and material – intersect to

---

<sup>195</sup> Mayeda, “Characterizing Gender and Race in the 2000 Summer Olympics,” 171.

<sup>196</sup> O’Bonsawin, “From Black Power to Indigenous Activism,” 211.

create a narrative of subsuming Indigenous DNA into the history and project of Australia-the-nation's DNA.

### **Indigenous DNA, Reconciliation, and Repatriation**

In the over 50 years since DNA first wormed its way into the public imaginary, the so-called “code of life” has become a common metaphor for essential or determining parts of a larger whole, whether that be a person, a people, or a nation.<sup>197</sup> As technological developments have now made it possible to use DNA as a data storage format, the metaphorical use of DNA – i.e., “the DNA of Australia” – is employed as justification for storing specific items onto the new format. DNA is not a storage format that indiscriminately carries its contents; because of its metaphorical inheritance, its structure has been imbued with the weighty sentiment that, like a sidecar metadata file, argues that the content is worth being contained by DNA. This implication – that the essence of material qualifies its treatment – can become pernicious when it echoes the histories of archival abuse and the exclusion or mistreatment of Indigenous materials. Not only that, the use of the DNA metaphor in the national context perpetuates how settler states selectively include and exclude Indigenous peoples and jurisdiction at their convenience. Examining how scientific genomic projects value Indigenous genetic material for research confronts the metaphor “\_\_\_ is the DNA of \_\_\_” with the material commodification of Indigenous DNA. In this section I use that confrontation to unravel how the racial and colonial projects imbued in genomics reappear when cultural heritage is stored onto DNA.

This chapter has primarily concerned the Australian context; however, in their decision to store the footage of Cathy Freeman, the NFSA and OFCH (an international organization) as well

---

<sup>197</sup> Weigmann, “The code, the text and the language of God,” 116.

as their United States-based technology partner Twist Bioscience have brought DNA data storage into the realm of critical race and Indigenous studies. Here is an opportunity for archives – especially national archives of settler states – to engage in questions of how these colonizing social structures are embedded in data storage formats at their inception. DNA, synthetic or biological, retains a double helical structure that has accumulated fifty years of genetic research and bioethics that inform our understanding of what DNA is and what it can do. In this section, I look primarily to Kim TallBear and Jenny Reardon’s research on the use of Indigenous genetic material as a raw resource by the Genographic Project. While the differences between synthetic DNA for storage and biological DNA from humans (and nonhumans) are many, they are discussed in tandem. In order to explain the long lasting qualities of DNA at the 2020 Association of Moving Image Archivists conference, Jan Müller repeatedly referred to Ötzi, a 5000-year old mummy found in the Swiss Alps whose DNA was extracted and analyzed.<sup>198</sup> References to Ötzi will resonate in a discussion of the treatment of the 9000-year-old Kennewick Man, skeletal remains found in Washington state. This section compares the treatment of biological DNA with the treatment of Indigenous artifacts and people in order to understand how the storage of cultural heritage materials on synthetic DNA can perpetuate cycles of archival abuse if not addressed properly.

The Genographic Project, a privately funded research endeavor founded in 2005, and its predecessor, Stanford University’s Human Genome Diversity Project, do not have specific links to the Human Genome Project (HGP), they rely heavily on the work accomplished by the HGP. Upon the completion of the first survey of the human genome, then-President Bill Clinton

---

<sup>198</sup> Jan Müller and Yasmin Meichtry, “Content as Data Stream: The First Heritage Video Stored on DNA – A Case Study About the Future of Digital Storage,” (Online Presentation at Association of Moving Image Archivists Conference, 17 November 2020).

announced that this information would inevitably lead to the prevention, treatment, and cure for all diseases.<sup>199</sup> Ramya M. Rajagopalan, Alondra Nelson, and Joan H. Fujimura, in their survey of the field of Science and Technology Studies identify the HGP as marking “an era during and after which DNA and genetics came to be seen as many as the ultimate source of disease,” as opposed to, say, a new source of research.<sup>200</sup> This view, write Rajagopalan et al., “is problematic for a number of reasons, not least of which are the conceptual connections that continue to drive popular perceptions of the association between disease incidence, genetics, and certain racialized groups.”<sup>201</sup> While the Genographic Project’s purpose was not to research disease, the Project was built on a framework that grounds difference, specifically negative difference in the manifestation of disease, in genetics. The Genographic Project and Human Genome Diversity Project cannot be divorced from their own intellectual genealogies in genomic studies.

The Genographic Project sought to trace human migratory paths back to Africa. As Spencer Wells, the lead researcher for the project, says, “we are all African under the skin,” highlighting the goals of proving that all humans come from the same place.<sup>202</sup> In order to find that common ancestor, the Genographic Project solicited blood samples from “highly unadmixed” people in “isolated indigenous populations.”<sup>203</sup> While claims of anti-racism are rife in the Genographic Project’s literature. It claims repeatedly that “racism is...scientifically

---

<sup>199</sup> White House Office of the Press Secretary, “President Clinton Announces the Completion of the First Survey of the Entire Human Genome: Hails Public and Private Efforts Leading to This Historic Achievement,” *Human Genome Project Information Archive 1990-2003*, 26 June 2000, accessed 2 May 2021, [https://web.ornl.gov/sci/techresources/Human\\_Genome/project/clinton1.shtml](https://web.ornl.gov/sci/techresources/Human_Genome/project/clinton1.shtml).

<sup>200</sup> Ramy M. Rajagopalan, Alondra Nelson, and Joan H. Fujimura, “Race and Science in the Twenty-First Century,” in *The Handbook of Science and Technology Studies*, ed. Ulrike Felt (Cambridge: MIT Press, 2016), 357.

<sup>201</sup> Rajagopalan, Nelson, and Fujimura, “Race and Science in the Twenty-First Century,” 357.

<sup>202</sup> Spencer Wells, quoted in Kim TallBear, *Native American DNA: Tribal Belonging and the False Promise of Genetic Science* (Minneapolis: University of Minnesota Press, 2013), 146.

<sup>203</sup> TallBear, *Native American DNA*, 146.

incorrect.”<sup>204</sup> Yet, the project relies heavily on scientific articulations of indigeneity and racial purity that are liable to “usurp claims to identity and perhaps legal rights” when considered in isolation from Indigenous knowledge, jurisdiction, and articulations of identity.<sup>205</sup> Moreover, at the core of the project is Wells’ belief that, while Indigenous peoples have their traditions, oral histories, and creation myths, white Europeans like himself are at a loss. They only have science:

We use science to tell us about [origin stories] because we don’t have the sense of direct continuity. Our ancestors didn’t pass down the stories. We’ve lost them, and we have to go out and find them. We use science, which is a European way of looking at the world to do that.<sup>206</sup>

Repurposing the language of exploration, Wells affirms that white Europeans like himself must search for their origin story, which can only be done through the extraction of the resource that is Indigenous genetic material. Only nominally anti-racist, the Genographic Project sought to create a white origin myth by the dissection, co-option, and exploitation of indigeneity.

While the Genographic Project depends on distinct haplotypes (genes inherited from a single parent) to trace ancestry, it also relies on social constructions that create and define difference in human populations. In this way, Wells perpetuates the “familiar position of a European making a moral claim on the natural resources of indigenous peoples” in that indigenous DNA is his means to claiming something that Indigenous peoples already possess in their history and tradition.<sup>207</sup> The cultural value Wells sees in indigeneity mirrors his belief that

---

<sup>204</sup> TallBear, *Native American DNA*, 146.

<sup>205</sup> Kim TallBear, “Narratives of Race and Indigeneity in the Genographic Project,” *The Journal of Law, Medicine & Ethics* 35, no. 3 (14 August 2007): 416, <https://doi.org/10.1111/j.1748-720X.2007.00164.x>.

<sup>206</sup> Spencer Wells, quoted in Jenny Reardon and Kim TallBear, “‘Your DNA Is Our History’: Genomics, Anthropology, and the Construction of Whiteness as Property,” *Current Anthropology* 53, no. S5, *The Biological Anthropology of Living Human Populations: World Histories, National Styles, and International Networks* (April 2012): S233, <https://www.jstor.org/stable/10.1086/662629>.

<sup>207</sup> Reardon and TallBear, “Your DNA Is Our History,” S234.

white Europeans deserve the same. Valuation of Indigenous heritage becomes justification for mining their genetic resources. The triumphant metaphor of “Australia’s DNA” also subsumes Indigenous accomplishment and existence into a nation-building narrative, while biological anthropologists “continue to value studying and understanding whiteness...and believe that a study of ‘redness’ is a constitutive part of this project.”<sup>208</sup> Just as the AOC selectively picked representations of indigeneity that benefitted the terms of reconciliation most favorable for Australia the nation, biological anthropologists too have a history of wielding their power to decide when it is appropriate for us all to be “African under the skin” and when to respect the rights of “isolated indigenous populations.”

There exist numerous examples of the misuse of Indigenous genetic material by large institutions that violate consent on many levels. The Navajo nation in 2002 instated a moratorium on genetic research following numerous studies attempting to connect disease to genetic “errors.”<sup>209</sup> The Indigenous Peoples Council on Biocolonialism (IPCB), an organization working from the early-90s to the mid-2000s composed of scientists and Indigenous leaders, deemed genetic research “a form of colonization” and “genetic vampirism.”<sup>210</sup> The case at Arizona State University, which violated the consent of Havasupai tribal members when they used samples collected for other purposes besides diabetes research.<sup>211</sup> The long history of medicalized violations against tribal nations accompanies the anthropological research of these tribes’ materials and bodies. The United States Native American Graves Protection and

---

<sup>208</sup> Reardon and TallBear, “Your DNA Is Our History,” S235.

<sup>209</sup> Oscar Schwartz, “A Geneticist’s Dilemma,” *The Washington Post Magazine*, 23 November 2020, accessed 2 May 2021, <https://www.washingtonpost.com/magazine/2020/11/23/many-scientists-believe-that-dna-holds-cure-disease-that-poses-problem-some-native-americans>.

<sup>210</sup> Indigenous Peoples Council on Biocolonialism, “Human Genetics Issues,” *IPCB.org*, accessed 19 April 2021, [http://www.ipcb.org/issues/human\\_genetics/index.html](http://www.ipcb.org/issues/human_genetics/index.html).

<sup>211</sup> Robyn L. Sterling, “Genetic Research among the Havasupai: A Cautionary Tale,” *Virtual Mentor* 13, no. 2 (February 2011): 113, <https://doi.org/10.1001/virtualmentor.2011.13.2.hlwa1-1102>.

Repatriation Act (NAGPRA) of 1990 attempts to remediate these relationships between American cultural heritage and research institutions and federally recognized tribal nations through repatriation. However, the case of the Kennewick Man demonstrates how DNA can become a tool wielded selectively by scientists. By their scientific authority, biological anthropologists adhere to Indigenous jurisdiction haphazardly and at their research's convenience.

In 1996, two students uncovered the 9000-year-old remains of a man now known as the Kennewick Man after the town in Washington state where it was found. The Umatilla tribal members and Colville tribes referred to the body as “The Ancient One.”<sup>212</sup> Invoking NAGPRA, several Columbia River Basin Indian tribes and bands demanded to repatriate and bury the bones before they could be researched. In a 2014 *Smithsonian Magazine* review of the book written by the scientists researching the Kennewick Man, Douglas Preston writes of the ensuing legal fight: “If it weren’t for a harrowing round of panicky last-minute maneuvering worthy of a legal thriller, the remains might have been buried and lost to science forever.”<sup>213</sup> Quite succinctly, Preston demonstrates how the mostly white scientific community lauded the manipulation of legal loopholes and their own scientific know-how to evade the unscientific claims made by tribal nations to protect their religious freedom and tribal jurisdiction. The phrase “lost to science forever” echoes the myth of the “vanishing Indian” perpetuated by the Genographic Project in their urgent language describing efforts to capture Indigenous DNA before populations become

---

<sup>212</sup> Kim TallBear, “Genomic articulations of indigeneity,” *Social Studies of Science* 43, no. 4, Special Issue: Indigenous Body Parts and Postcolonial Technoscience (August 2013): 520, <https://www.jstor.org/stable/43284191>.

<sup>213</sup> Douglas Preston, “The Kennewick Man Finally Freed to Share His Secrets,” *Smithsonian Magazine*, September 2014, accessed 2 May 2021, <https://www.smithsonianmag.com/history/kennewick-man-finally-freed-share-his-secrets-180952462>.

too admixed and disappear.<sup>214</sup> Ultimately, the scientists won their lawsuit on the basis that the bones could not possibly be related to a living tribe, and therefore could not establish cultural kinship to invoke NAGPRA. This argument was supplemented by the forensic anthropologists' claim that the bones did not appear Native American enough. In other words, the Ancient One was not indigenous enough to the area and not Indigenous enough, despite pre-dating colonization by several millennia.

The court made their decisions based on settler articulations of ancestry and cultural heritage rooted in anthropological bioscience opposed to tribal jurisdiction and Indigenous articulations of identity. Instead, scientists fit the Kennewick Man into the Bering Strait migration narrative, attaching him not to the Indigenous peoples who reside on the land of his final resting place but instead to "mysterious people [who] have long since disappeared."<sup>215</sup> TallBear and Reardon use Cheryl Harris' framework of whiteness as property to describe how "claims by Native Americans to a property interest in their own biological materials and history often are viewed...as an obstruction of 'the original or current distribution of power, property, and resources [that are] the result of 'right' and 'merit.''"<sup>216</sup> The value of Indigenous genetic materials only exists when it contributes to the settler scientific norms; when that genetic material cannot conform, it cannot be Indigenous. While DNA tests at the time were not possible, technological developments allowed for DNA analysis in 2015. The Colville tribes claiming the remains reluctantly consented to DNA testing, and results linked them and the

---

<sup>214</sup> TallBear, "Genomic articulations of indigeneity," 513.

<sup>215</sup> Preston, "The Kennewick Man Finally Freed to Share His Secrets."

<sup>216</sup> Reardon and TallBear, "Your DNA Is Our History," S240.



Ancient One.<sup>217</sup> The Kennewick Man was repatriated and buried in 2017, long after some bones were lost and others tested and researched against the wishes of his ancestors.<sup>218</sup>

In *Native American DNA*'s extended discussion of the Kennewick Man, TallBear examines how NAGPRA juggles settler notions of race, culture, and science with tribal ones. The requirement of establishing "cultural affiliation" can become difficult considering the shifting definitions of race alongside the history of colonization in the Americas that have profoundly changed definitions of "belonging" and "difference" since 1492.<sup>219</sup> This reading also fails to take into account the diverse contemporary tribal membership requirements – which can include a signed affidavit from a relative, DNA testing, symbolic blood, and blood rules – to which the Kennewick Man could not necessarily conform.<sup>220</sup> Writing before his repatriation, TallBear argues that knowing the "Kennewick Man walked North America 8,500 prior to European colonization" is the indisputable fact.<sup>221</sup> The genetic links between the Kennewick Man and Native Americans and non-Native Americans cannot dispute that fact. However, as she notes in an article published after the Kennewick Man's DNA was sequenced, "science shares with the law – and with white people in general – the power to define and ultimately police race."<sup>222</sup> Despite the numerous treaties in place meant to counteract violations of Indigenous

---

<sup>217</sup> Carl Zimmer, "New DNA Results Show Kennewick Man Was Native American," *The New York Times*, 18 June 2015, accessed 2 May 2021, <https://www.nytimes.com/2015/06/19/science/new-dna-results-show-kennewick-man-was-native-american.html>.

<sup>218</sup> "Tribes lay remains of Kennewick man to rest," *The Spokesman Review*, 20 February 2017, accessed 2 May 2021, <https://www.spokesman.com/stories/2017/feb/20/tribes-lay-remains-of-kennewick-man-to-rest>.

<sup>219</sup> TallBear, *Native American DNA*, 157.

<sup>220</sup> TallBear, "Genomic articulations of indigeneity," 523-4.

<sup>221</sup> TallBear, *Native American DNA*, 157.

<sup>222</sup> Kim TallBear, "Who Owns the Ancient One?" *Buzzfeed News*, 23 July 2015, accessed 2 May 2021, <https://www.buzzfeednews.com/article/kimtallbear/how-the-man-stole-ancient-man-from-his-native-descendents>.

jurisdiction like the research done on the Kennewick Man, the settler state will continue to mandate how people can be Indigenous and when their rights will be respected.

NAGPRA, the scientific community, and the settler state require tribal nations to operate within white narratives and methodologies in order to prove their rights. Ultimately, the DNA sequencing of the Kennewick Man is irrelevant to the incontrovertible fact that he existed in North America before colonization. As Debra Harry and Frank Dukepoo of the IPCB wrote of white scientists' desire to prove the Bering Strait theory, "It's possible these new 'scientific findings' concerning our origins can be used to challenge aboriginal rights to territory, resources and self-determination."<sup>223</sup> The case of the Kennewick Man demonstrates that the IPCB's fear of delegitimization by so-called 'scientific findings' forces tribal nations to conform to settler standards of race and ethnicity in order to maintain sovereignty. These standards in turn destabilize the very traditional knowledge bases and oral histories which people like Spencer Wells are so keen to possess. DNA is a tool that has been used consistently to reinforce settler understandings of indigeneity in order to reinforce conceptualizations of race that prioritize whiteness. As a metaphor, these same systems of abuse accomplish the same task: at the settler state's convenience, the metaphor of DNA seamlessly absorb Indigenous narratives that benefit the state while dismissing, condemning, or forgetting those that do not. These two cases also demonstrate how institutions with repositories of Indigenous materials can perpetuate abuse by violating Indigenous jurisdiction. Moreover, they weaponize white European constructions that continue to promote indigeneity only as a means of constructing and reinforcing whiteness.

---

<sup>223</sup> Debra Harry and Frank Dukepoo, quoted in TallBear, *Native American DNA*, 154.

### **Indigenous Collections and the *Protocols***

This chapter has sought to contend with the ways metaphors – “the DNA of Australia” – and materials, represented by the exploitation of Indigenous genetic resources, can reinforce structures of colonization and can undermine Indigenous jurisdiction. I have argued that the rhetorical language used to describe the storage of Cathy Freeman’s “golden win” onto DNA reinforces a narrative of state-approved reconciliation, ignores Freeman’s own legacy of activism, and subsumes her indigeneity into “all Australia.” I have also argued that the language of DNA, synthetic or otherwise, is intertwined with the exploitation of Indigenous DNA by scientific projects tracing human migration. The co-option and framing of Indigenous DNA as a raw material resource that can create a white origin myth cannot be disentangled from a metaphor which essentializes indigeneity as part of a nation-building story. Here, these two narratives of DNA come together and reveal an opportunity for archives and libraries to partake in reconciliation. I offer that tools like the protocols for managing Indigenous collections should be used not only as directives for handling specific materials but also as a foundation that can help to grapple with the ethical questions DNA data storage can pose. Cultural heritage institutions are entering the field of genomics. In order to pursue reconciliation meaningfully, archives must engage with these material, technological structures for the equitable longevity, sustainability, and accessibility of collections.

While each analysis of national archives must take into account respective national context, the conditions of settler state archives – in particular Australia, New Zealand, the United States, and Canada – are not dissimilar. In her “Genomic Articulations of Indigeneity,” Kim TallBear uses the sociocultural anthropology methodology of articulation in order to analyze how “previously disparate elements are conjoined into new cultural and social formations in acts

of borrowing, interpretation, and reconfiguration.”<sup>224</sup> In particular, TallBear applies articulation to understand how the genomic formation of indigeneity interacts with the indigenous one. There, she analyzes how power structures affect the manifestation of different tools for articulation within these groups. In synthesizing various Indigenous definitions of indigeneity, TallBear sees a crucial difference between those and the genomic articulations in understandings of the environment and human divide: “Indigenous notions of *peoplehood* as emerging *in relation with* particular lands and waters and their nonhuman actors differ from the concept of a genetic *population*, defined as moving *upon or through* landscapes.”<sup>225</sup> From this distinction between peoplehood and population, TallBear argues that Indigenous peoples globally are united in self-determination and opposition to colonialism. The questions asked by projects like the Genographic Project presuppose a colonial relation to land that undermines Indigenous jurisdiction and self-determination. Moreover, the assimilation of indigeneity into a nationalizing project disavows Indigenous articulations as necessarily anticolonial. For TallBear, these genomic articulations exist within a larger, colonial structure that imposes concepts of land and population onto Indigenous articulations.

The disparate epistemological frameworks of these articulations of indigeneity reappear within the context of reconciliation. Reconciliation itself will never be decolonization, an ethic that, as Eve Tuck and K. Wayne Yang describe in their “Decolonization is not a metaphor,” when employed as a metaphor “makes possible a set of evasions that problematically attempt to reconcile settler guilt and complicity, and rescue settler futurity.”<sup>226</sup> The metaphor of DNA can be wielded as an essentializing force that performs a very similar function. In acknowledging

---

<sup>224</sup> TallBear, “Genomic articulations of indigeneity,” 512.

<sup>225</sup> TallBear, “Genomic articulations of indigeneity,” 514-5.

<sup>226</sup> Eve Tuck and K. Wayne Yang, “Decolonization is not a metaphor,” *Decolonization: Indigeneity, Education & Society* 1, no. 1: 1, <https://jps.library.utoronto.ca/index.php/des/article/view/18630>.

that “the desire to reconcile is just as relentless as the desire to disappear the Native” and that reconciliation will never fulfill a promise of decolonization, scholars like Adam Gaudry and Danielle Lorenz argue for a critical approach to diversity and inclusion standards within the Canadian Academy. In this context, Gaudry and Lorenz navigate how reconciliation indigenization “requires power sharing, a transformation of decision-making processes, and a reintegration of Indigenous peoples... into policymaking that affects them, and their Canadian peers.”<sup>227</sup> In short, a reconciliation indigenization (their term), as opposed to inclusion, alters how the academy proper operates. Kim TallBear additionally stresses that reconciliation should fall to the colonizers as a self-education and behavior modification.<sup>228</sup> Reconciliation is not decolonization, which “brings about the repatriation of Indigenous land and life,” and must repatriate.<sup>229</sup> Starting from a place that stresses the action of reconciliation and decolonization allows archives to move past metaphors of decolonization and toward definitive restructuring that can change how institution operate in a restorative way.

I outline and address these different concepts of reconciliation and decolonization in an order to be specific in how I discuss them within the archival context and to be direct about the effectiveness and efforts of different directiveness. Tuck and Yang urge us to be careful with how we use “decolonization;” any metaphorical use of the term diverts from the concrete actions of repatriation and restoration that define decolonization. Following this strict definition, decolonization in the archive, especially a national archive, whose nation-building mission is necessarily in opposition to the efforts of decolonization, is difficult, but not impossible to move towards. Best practices documents like the *Protocols for Native American Archival Materials*,

---

<sup>227</sup> Gaudry and Lorenz, “Indigenization as inclusion, reconciliation, and decolonization,” 223.

<sup>228</sup> TallBear, “Diversity v. Decolonization in the Academy.”

<sup>229</sup> Tuck and Yang, “Decolonization is not a metaphor,” 1.

which draws heavily from the *Aboriginal and Torres Strait Islander Library, Information and Resource Network Protocols for Libraries, Archives, and Information Services* (ATSILIRN *Protocols*), are acts of reconciliation and encourage decolonization through the repatriation of materials. The *Protocols*, which were written in 2006 and ratified as an external standard by the Society of American Archivists in 2018, establish “best professional practices for culturally responsive care and use of American Indian archival material held by non-tribal organizations.”<sup>230</sup> In the American context, the *Protocols* rose specifically out of the experiences of Native Americans attempting to access documents required for tribal federal recognition.<sup>231</sup> Additionally, for the most part, the vast collections of Native American materials in non-Native institutions comes primarily from 19<sup>th</sup> century anthropological collecting stemming from the belief in the “vanishing Indian,” which mirrors the fear of admixture expressed by the Genographic Project.<sup>232</sup> The *Protocols* seek to contend with these collecting policies and colonizing practices and offer concrete solutions archiving professionals can take in order to better the status of Indigenous collections.

Because of the necessary engagement with legal language and Indigenous jurisdiction, the *Protocols* outline policies non-Native organizations should put in place in order to be culturally responsive in their care for Indigenous archival materials. They thoroughly document how archives have a responsibility to reconciling their own colonizing past in order to initiate building mutual respect. Established, importantly, through the legal language of rights, the *Protocols* place the responsibility on cultural heritage institutions: “Libraries and archives must

---

<sup>230</sup> Protocols for Native American Archival Materials, “Home,” *Northern Arizona University*, accessed 19 April 2021, <https://www2.nau.edu/libnap-p>.

<sup>231</sup> Jennifer O’Neal, “‘The Right to Know’: Decolonization Native American Archives,” *Journal of Western Archives* 6, no. 1 Native American Archives Special Issue (2015): 7, <https://doi.org/10.26077/fc99-b022>.

<sup>232</sup> O’Neal, “The Right to Know,” 5.

recognize that Native American communities have primary rights for all culturally sensitive materials that are culturally affiliated with them. These rights apply to issues of collection, preservation, access, and use of or restrictions to these materials.”<sup>233</sup> As TallBear says of reconciliation, the onus and responsibility of establishing and respecting rights is on the colonizer, in this case, the archive. The *Protocols* focus specifically on the task of libraries and archives to provide responsible context for Indigenous materials and provides concise and direct guidelines for action. Moreover, the *Protocols* frame the importance of descriptive language as an issue of access: “the use of outdated, inaccurate, derogatory, or Eurocentric language impedes access.”<sup>234</sup> Access, a crucial part of an archive’s structure, relies on accurate description and metadata in order to make it possible for designated communities and userbases to find materials.

The *Protocols* stress that descriptive language can cause injury to user groups if it is offensive. Additionally, they stress how “injurious perspectives and information [that] may be inherent to the content of some of the original materials” can impede access without appropriate context from the archiving institution. The *Protocols* outline many different decolonizing and reconciliatory actions archives and libraries can take in order to adhere to these best practices. They prioritize handle different epistemological systems – between Native American communities and non-Native archives – in order to encourage the proper management of Indigenous materials that may or may not be prohibited, classified, secret, or otherwise restricted. Closer reading also demonstrates that the *Protocols* that the reconciliation between these “perspectives” goes beyond the management of specific materials and calls on archives and

---

<sup>233</sup> Protocols for Native American Archival Materials, “Protocols for Native American Archival Materials,” *Northern Arizona University*, accessed 19 April 2021, <https://www2.nau.edu/libnap-p/protocols.html>

<sup>234</sup> Protocols for Native American Archival Materials, “Protocols for Native American Archival Materials.”

libraries to question concepts of property, to integrate moral rights (*droit moral*) into library policy, and, specifically within the *ATSILIRN Protocols*, to “ensure sustainable choices of formats, descriptive methods and access and preservation strategies for Aboriginal and Torres Strait Islander peoples’ knowledge, creativity and experience.” The incorporation of these less collection-specific calls to action emphasizes how *Protocols* seek to become foundational standards that can be incorporated into archives’ mission statements and collecting policies. By embracing these directives for *all* materials – not just ones explicitly claimed or named (which again, engages in colonial understandings of property) as Indigenous – archives can practice, at the least, reconciliation from the outset.

The *Protocols*, both in the American and Australian context, primarily refer to historical materials. They are indebted to Vine Deloria’s concept of the “Right To Know,” a 1978 call to action for specific steps required for the federal government to fulfill treaty educational provisions and allow Indigenous peoples access to their own history. As Jennifer O’Neal writes in her historiography of Native American Archives, “enveloped within this call to action is Deloria’s recognition that information and knowledge are critical to the sovereignty and self-determination of Native nations.”<sup>235</sup> This stress on intellectual access as critical to self-determination reappears in the *Protocols*, themselves. These *Protocols* were written specifically with the intention to “support archival theory and practices through the theoretical concepts of the post-custodial model for participatory and community archives,” and purposefully argue that non-western perspectives are crucial and “desperately needed in archival education, practice, and the profession at large.”<sup>236</sup> In this way, the *Protocols* engage not only with the management of the past, but the active profession of the present and the future. Using the *Protocols* as a means to

---

<sup>235</sup> O’Neal, “The Right to Know,” 3.

<sup>236</sup> O’Neal, “The Right to Know,” 14-5.



manage Indigenous collections of the present also invests in “cultural survival” and actively resists the “myth of the vanishing Indian” perpetuated by those 19<sup>th</sup> century collections.<sup>237</sup> Additionally, non-Native archives must apply *Protocols* to more recently produced collections in order to empower an Indigenous futurity and reaffirm the Indigenous jurisdiction. And, as the *Protocols* outline, the maintenance and preservation of Indigenous records – perhaps all records held by colonizing institutions – must include the necessary work of repatriation, and the inclusion, understanding, and expression of Indigenous values, perspectives, and contexts. Without this labor, archives cannot adhere to the best practices outlined by the *Protocols*.

When archives attempt to adhere to those best practices, thinking beyond descriptive metadata or limited access is crucial. The ATSILIRN *Protocols* stipulate that archives must “ensure sustainable choices of formats, descriptive methods, access, and preservation strategies for Aboriginal and Torres Strait Islander peoples’ knowledge, creativity, and experience.”<sup>238</sup> In this thesis, sustainability has implied economic sustainability, ecological sustainability, community buy in, ease of use, amongst a myriad of meanings. Here, the ATSILIRN *Protocols* implore archives to question how a choice of format can be sustainable for Aboriginal and Torres Strait Islander peoples. We must ask if DNA is sustainable – not only as a green technology – but as it exists in our public imaginary, and how it imparts meaning in the context of Indigenous collections and the portrayal of Indigenous peoples. We must ask what work archivists need to do in order to make DNA a viable format beyond those technical specifications, especially as DNA carries with it this history of violating Indigenous jurisdiction and itself is a mode of racialization and colonization. DNA is a tool. In the archive, it must not be a tool by which

---

<sup>237</sup> O’Neal, “The Right to Know,” 9.

<sup>238</sup> Aboriginal and Torres Strait Islander Library, Information and Resource Network, “Protocols for Libraries, Archives and Information Services,” *ATSILIRN*, 2012, accessed 3 May 2021, <https://atsilirn.aiatsis.gov.au/protocols.php>.

curators and archivists confine Indigenous collections within racializing Western knowledge systems and systems of value.

## Conclusion

DNA-based data storage has been in development in earnest since at least 2012. In recent years, many companies already within the medical technology sector – namely synthesis and sequencing companies like Twist Bioscience and Illumina – have expanded their operations to include DNA-based data storage. With the formation of the DNA Data Storage Alliance, a consortium of companies and institutions participating in the development of and commercialization of DNA-based data storage, the technology has investment and momentum behind it. Moreover, many of these companies, consortia, and working groups have specified that archives are a primary use case. Identifying archives as a use case has meant that DNA-based data storage companies have courted a variety of cultural heritage institutions for pilot projects, sponsorship, and collaboration. This interest resulted in the NFSA joining with the OFCH and Twist Bioscience for their pilot project to explore DNA-based data storage. Still, the high costs, slow speeds, and technical limitations of DNA at this point in time do not make it an attractive solution for archival storage right now. The outlined necessary technical requirements for archival storage not only call for the further development of current features of DNA but also stipulate new features – such as editing and fixity – that have yet to be considered by the field. In the next ten years, when many predict the technology will be commercially viable and available, archivists must intervene in these working groups and consortia in order to advocate for our needs.

As the DDSA gains credibility and those proffering DNA-based data storage solicit archives, archivists who are invested in the ways our daily work affects the life of digital objects in the public imaginary through access must contend with the merging of DNA and those objects. Archivists are uniquely positioned to ask these onto-epistemological questions about the

co-production of technology and social order. The access to and responsibility to these digital objects lends archivists unique perspectives on how the technology affects these objects at the level of the bit, grain, or nucleotide. In the case study of the video footage of Cathy Freeman stored onto DNA by the NFSA, the corporate structure of the NFSA and the use of this new technology for press ultimately undermine claims of preservation. More insidiously, the archive's adoption of the metaphor of life and essentialism so often used with respect to the co-option of Indigenous DNA by biological anthropologists exemplifies how technology and social production underwrite each other at the site of the archive. Archives have had to engage with other technologies of oppression in the efforts of preservation. Media technologies in particular are fraught with prejudices ingrained in their very structures.

This thesis has offered that the *Protocols* written by Indigenous and non-Indigenous archivists can point to methods of preservation, of “care and feeding,” that account for these structures. This early in the technology's development, archivists must ask what our handling of DNA will show us about our current practices? How can DNA reveal our relations to the objects onto which we store cultural heritage? What DNA has shown is that archival storage, labor, and “care and feeding,” are all critical sites where issues of race, sovereignty, and nation are negotiated through the archive's institutional power to wield DNA as a technology. This thesis has approached the archive for its material conditions: what the use of archival storage means, and how curatorial and archival choices manifest at the level of the bitstream, and now the molecule. The introduction of synthetic DNA – and by proxy, the metaphor of life, and all the prejudice of life – to these conditions has the power to reconfigure the archive. DNA makes readily apparent that the archive is not a metaphor. Its technological materialities must be taken

seriously in order to contend with the racial and colonial projects that have reappeared as genomics and cultural heritage institutions collide at the site of data storage.

## Works Cited

- “DNA-based Data Storage and Computation Provider CATALOG, Founded by MIT Scientists, Raises \$10 Million and Names IBM Executive as CTO.” Catalog DNA. 8 September 2020. Accessed 24 April 2021. <https://www.catalogdna.com/blog/dna-based-data-storage-and-computation-provider-catalog-founded-by-mit-scientists-raises-10-million-and-names-ibm-executive-as-cto>.
- “Format Obsolescence.” *Digital Bedrock*. Accessed 2 May 2021. <https://www.digitalbedrock.com/format-obsolescence>.
- “The History of Magnetic Tape and Computing: A 65-Year-Old Marriage Continues to Evolve.” *Iron Mountain*. Accessed 2 May 2021. <https://www.ironmountain.com/resources/general-articles/t/the-history-of-magnetic-tape-and-computing-a-65-year-old-marriage-continues-to-evolve>.
- “LTO-9.” *Ultrium LTO*. Accessed 27 April 2021. <https://lto.org/lto-9>.
- Reference Model for an Open Archival Information System*. Washington, DC: The Consultative Committee for Space Data Systems, June 2012. <https://public.ccsds.org/pubs/650x0m2.pdf>.
- “President Clinton: Announcing the Completion of the First Survey of the Entire Human Genome.” The White House at Work Archives. 26 June 2000. Accessed 24 April 2021. <https://clintonwhitehouse3.archives.gov/WH/Work/062600.html>.
- “Tribes lay remains of Kennewick Man to rest.” *The Spokesman Review*, 20 February 2017. <https://www.spokesman.com/stories/2017/feb/20/tribes-lay-remains-of-kennewick-man-to-rest>.

“Understanding the NGS workflow.” *Illumina*. Accessed 24 April 2021.

<https://www.illumina.com/science/technology/next-generation-sequencing/beginners/ngs-workflow.html>.

Aboriginal and Torres Strait Islander Library, Information and Resource Network. “Protocols for Libraries, Archives and Information Services.” *ATSILIRN*. 2012. Accessed 3 May 2021.

<https://atsilirn.aiatsis.gov.au/protocols.php>

Anavy, Leon, Inbal Vaknin, Orna Atar, Roe Amit, and Zohar Yakhini. “Data storage in DNA with fewer synthesis cycles using composite DNA letter.” *Nature Biotechnology* 37 (9 September 2019): 1229-1236. <https://doi.org/10.1038/s41587-019-0240-x>.

Anderson, Patrick, Richard Black, Ausra Cerkauskaitė, Andromachi Chatzieleftheriou, James Clegg, Chris Dainty, Raluca Diaconu, Austin Donnelly, Rokas Drevinskas, Alexander Gaunt, Andreas Georgiou, Ariel Gomez Diaz, Peter G. Kazansky, David Lara, Sergey Legtchenko, Sebastian Nowozin, Aaron Ogus, Douglas Phillips, Ant Rowstron, Masaaki Sakakura, Ioan Stefanovici, Benn Thomsen, Lei Wang. “Glass: A New Media for a New Era?” *10<sup>th</sup> USENIX Workshop on Hot Topics in Storage and File Systems* Hot Storage 18 (July 2018): 1-6. <https://www.microsoft.com/en-us/research/uploads/prod/2018/07/hotstorage18-paper-anderson.pdf>.

Antonini, Marc, Luis Cruz, Eduardo da Silva, Melpomeni Dimopoulou, Touradj Ebrahimi, Siegfried Foessel, Fernando Pereira, António Pinheiro, and Mohama Raad. “DNA-based Media Storage: State-of-the-Art, Challenges, Use Cases and Requirements version 3.0.” Edited by Touradj Ebrahimi and Fernando Pereira. *JPEG DNA Exploration*. Lausanne: ISO/IEC JTC 1/SC29/WG1N90023, 22 January 2021.

<http://ds.jpeg.org/documents/jpegdna/wg1n90023-REQ-DNA-based Media Storage State-of-the-Art Challenges Use Cases and Requirements.pdf>.

Antonini, Marc. "Image coding for long-term storage on synthetic DNA." Slide presentation, JPEG, 30 September 2020. Video. 22:54-01:08:00.

<https://www.youtube.com/watch?v=U5EKe-RDYN4>.

Barlow, Karen. "Boxer Hooper apologises for Aboriginal shirt." *ABC News*, 30 July 2012.

Accessed 2 May 2021. <https://www.abc.net.au/news/2012-07-31/boxers-aboriginal-t-shirt-draws-ire-of-officials/4165260>.

Bhattarai-Kline, Santi, Sierra K. Lear, and Seth L. Shipman. "One-step data storage in cellular DNA." *Nature Chemical Biology* 17 (26 January 2021): 232-233.

<https://doi.org/10.1038/s41589-021-00737-2>.

Blumenthal, Karl, Peggy Griesinger, Julia Y. Kim, Shira Peltzman, Vicky Steeves. "What's Wrong with Digital Stewardship: Evaluating the Organization of Digital Preservation Programs from Practitioners' Perspectives." *Journal of Contemporary Archival Studies* 7, no. 13 (2020). <https://elischolar.library.yale.edu/jcas/vol7/iss1/13>.

Boettner, Benjamin. "New CRISPR technology takes cells to the movies." *The Harvard Gazette*, July 12, 2017. <https://news.harvard.edu/gazette/story/2017/07/taking-cells-out-to-the-movies-with-new-crispr-technology>.

Bornholt, James, Randolph Lopez, Douglas M. Carmean, Luis Ceze, Georg Seelig, and Karin Strauss. "Toward a DNA-Based Archival Storage System." *IEEE Micro* 37, no. 3 (2017): 98-104. 10.1109/MM.2017.70.

Calle, Danielle. "Climate Change as a Prompt for the Digital Archive." MA Thesis. New York University, 2020.



Ceze, Luis, Jeff Nivala, and Karin Strauss. "Molecular digital data storage using DNA." *Nature Review Genetics* 20 (2019). 456-466. <https://doi.org/10.1038/s41576-019-0125-3>.

Chadash, Daniel. "An Overview of the DNA Data Storage Alliance." Zoom slide presentation, 2<sup>nd</sup> JPEG DNA Workshop (27 January 2021). Video. 03:45-15:33.  
<https://www.youtube.com/watch?v=IFNE47YfVWg>.

Cherchi Usai, Paolo. "A Charter of Curatorial Values." *Journal of the National Film and Sound Archive, Australia* 1, no. 1 (Spring 2006): 1-11.

Choi, Charles Q. "Nature's Databank." *Prism*, February 2020. <http://www.asee-prism.org/natures-databank>.

Church, George, Yuan Gao, and Sriram Kosuri. "Next-Generation Digital Information Storage in DNA." *Science* 337, no. 6102 (28 September 2012): 1628.  
<https://www.jstor.org/stable/41703606>.

Church, George and Seth Shipman. "CRISPR-Cas: Molecular Recording." Harvard University. 12 July 2017. Video, 3:10. <https://www.youtube.com/watch?v=o8ONjPHLr6s>.

CathyFreeman.com.au. "Cathy Freeman." *Estrella Sports Management*. Accessed 19 April 2021.  
<https://estrellasports.com.au/portfolio/items/cathy-freeman/>.

Council for Aboriginal Reconciliation. "About the Council for Aboriginal Reconciliation." *Australasian Legal Information Institute*. Accessed 19 April 2021.  
<http://www5.austlii.edu.au/au/orgs/car/council/about.htm>.

Council for Aboriginal Reconciliation. "Vision Statement." *Australasian Legal Information Institute*. Accessed 18 April 2021.  
[http://www5.austlii.edu.au/au/orgs/car/council/spl98\\_20/council.htm](http://www5.austlii.edu.au/au/orgs/car/council/spl98_20/council.htm).

Davis, Joe. "Microvenus." *Art Journal* 55, no. 1 (Spring 1996): 70-74.

[www.jstor.org/stable/777811](http://www.jstor.org/stable/777811).

Ebrahimi, Touradj, Fernando Pereira, Linda Tadic, Reinhard Heckel, Sergey Yekhanin. "3<sup>rd</sup>

JPEG DNA Workshop." Zoom meeting, Third JPEG DNA Workshop, 9 April 2021.

[https://jpeg.org/items/20210316\\_jpeg\\_dna\\_3rd\\_workshop\\_announcement.html](https://jpeg.org/items/20210316_jpeg_dna_3rd_workshop_announcement.html).

Eisenstein, Michael. "Enzymatic DNA synthesis enters new phase." *Nature Biotechnology* 38

(October 2020): 1113-1115. <https://doi.org/10.1038/s41587-020-0695-9>.

Ehrenberg, Rachel. "DNA could prove practical for long-term data storage." *Science News* 183,

no. 4 (23 February 2013): 5-6. <https://www.jstor.org/stable/235998785>.

Ernst, Wolfgang. "Aura and Temporality: The Insistence of the Archive." *Quaderns portàtils* 29

(2013): 1-25. <https://www.macba.cat/en/learn-explore/publications/aura-and-temporality-insistence-archive>.

Gaudry, Adam and Danielle Lorenz. "Indigenization as inclusion, reconciliation, and

decolonization: navigating the different visions for indigenizing the Canadian Academy."

*AlterNative: An International Journal of Indigenous People* 14, no. 3 (2018): 218-227.

<https://doi.org/10.1177/1177180118785382>.

Greenfieldboyce, Nell. "As Made-To-Order DNA Gets Cheaper, Keeping It Out of the Wrong

Hands Gets Harder." *National Public Radio*, 24 September 2019. Accessed 23 April

2021. [https://www.npr.org/sections/health-shots/2019/09/24/762834987/as-made-to-](https://www.npr.org/sections/health-shots/2019/09/24/762834987/as-made-to-order-dna-gets-cheaper-keeping-it-out-of-the-wrong-hands-gets-harder)

[order-dna-gets-cheaper-keeping-it-out-of-the-wrong-hands-gets-harder](https://www.npr.org/sections/health-shots/2019/09/24/762834987/as-made-to-order-dna-gets-cheaper-keeping-it-out-of-the-wrong-hands-gets-harder).

Goldman, Nick, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M. LeProust, Botond

Sipos, and Ewan Birney. "Towards practical, high-capacity, low-maintenance

information storage in synthesized DNA.” *Nature* 494 (23 January 2013): 77-80.

doi:10.1038/nature11875.

Gonzalez, Miguel. “Technological World First for NFSA.” *National Film and Sound Archive of Australia*. Last modified 25 September 2020. Accessed 19 April 2021.

<https://www.nfsa.gov.au/latest/nfsa-dna>.

Gordon, Michael. “The day Nicky Winmar drew the line.” *The Age*, 16 April 2013. Accessed 2 May 2021. <https://www.theage.com.au/sport/afl/the-day-nicky-winmar-drew-the-line-20130416-2hydd.html>,

Heckel, Reinhard. “Low cost DNA data storage with noisy synthesis and advanced error correction.” Zoom slide presentation at 3<sup>rd</sup> JPEG DNA Workshop (9 April 2021).

Hellmold, Steffen. “Storage Innovation Is in our DNA.” Western Digital. Zoom slide presentation at Small Data Industries (20 November 2020).

Hughes, Randall A. and Andrew D. Ellington. “Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology.” *Cold Spring Harbor Perspectives in Biology* 9, no. 1 (January 2017): 1-17. <https://doi.org/10.1101/cshperspect.a023812>.

Human Rights and Equal Opportunity Commission. *Bringing Them Home: National Inquiry into the Separation of Aboriginal and Torres Strait Islander Children from Their Families*.

Commonwealth of Australia. 1997.

[https://humanrights.gov.au/sites/default/files/content/pdf/social\\_justice/bringing\\_them\\_home\\_report.pdf](https://humanrights.gov.au/sites/default/files/content/pdf/social_justice/bringing_them_home_report.pdf).

IFLA Study Group on the Functional Requirements for Bibliographic Records. *Functional Requirements for Bibliographic Records* (February 2009). Accessed 3 May 2021.

[https://www.ifla.org/files/assets/cataloguing/frbr/frbr\\_2008.pdf](https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf)

- Indigenous Peoples Council on Biocolonialism. "Human Genetics Issues." *IPCB.org*. Accessed 19 April 2021. [http://www.ipcb.org/issues/human\\_genetics/index.html](http://www.ipcb.org/issues/human_genetics/index.html).
- Jasanoff, Sheila. *States of Knowledge: The Co-Production of Science and Social Order*. New York: Routledge, 2006.
- Jonas, William. "Chapter 2: Reconciliation and human rights." In *Social Justice Report 2000*. Australian Human Rights Commission. 1 January 2000. <https://humanrights.gov.au/our-work/social-justice-report-2000-chapter-2-reconciliation-and-human-rights>.
- Jonas, William. "Chapter 6: Reconciliation – National progress one year on." In *Social Justice Report 2001*. 1 January 2001. <https://humanrights.gov.au/our-work/aboriginal-and-torres-strait-islander-social-justice/publications/social-justice-report-9>.
- Keast, Jackie. "NFSA CEO Jan Müller resigns." *IF*, 11 December 2020. <https://www.if.com.au/nfsa-ceo-jan-muller-resigns>.
- Koch, Julian, Silvan Gantenbein, Kunal Masania, Wendelin J. Stark, Yaniv Erlich, and Robert N. Grass. "A DNA-of-things storage architecture to create materials with embedded memory." *Nature Biotechnology* 38 (09 December 2019): 39-43. <https://doi.org/10.1038/s41587-019-0356-z>.
- Kohll, A. Xavier, Philipp L. Antkowiak, Weida D. Chen, Bichlien H. Nguyen, Wendelin J. Stark, Luis Ceze, Karin Strauss, and Robert N. Grass. "Stabilizing synthetic DNA for long-term data storage with earth alkaline salts." *Chemical Communications* 56 (24 February 2020): 3613-3616. <https://doi.org/10.1039/D0CC00222D>.
- Langston, Jennifer. "Project Silica proof of concept stores Warner Bros. 'Superman' movie on quartz glass." *Microsoft News*, 4 November 2019. Accessed 27 April 2021. <https://news.microsoft.com/innovation-stories/ignite-project-silica-superman/>.

Lavoie, Brian. "Meeting the challenges of digital preservation: The OAIS reference model."

*OCLC Newsletter* 243, no. 26-30 (January/February 2000).

<https://www.oclc.org/research/publications/2000/lavoie-oais.html>.

Ling, Wing-Kuen. *Nonlinear Digital Filters*. Academic Press, 2007.

<https://doi.org/10.1016/B978-012372536-3/50001-6>.

Mayeda, David. "Characterizing Gender and Race in the 2000 Summer Olympics: NBC's

Coverage of Maurice Greene, Michael Johnson, Marion Jones, and Cathy Freeman."

*Social Thought & Research* 24, no. 1/2, The Politics of Gender (2001): 145-186.

<https://www.jstor.org/stable/23250078>.

Morris, Linda. "In our DNA: Cathy Freeman's Olympic win projected onto Opera House." *The*

*Sydney Morning Herald*, 25 September 2020. [https://www.smh.com.au/culture/art-and-](https://www.smh.com.au/culture/art-and-design/in-our-dna-cathy-freeman-s-olympic-win-projected-onto-opera-house-20200922-p55xzx.html)

[design/in-our-dna-cathy-freeman-s-olympic-win-projected-onto-opera-house-20200922-p55xzx.html](https://www.smh.com.au/culture/art-and-design/in-our-dna-cathy-freeman-s-olympic-win-projected-onto-opera-house-20200922-p55xzx.html).

Moskowitz, Clara. "DNA-Coded 'Lunar Library' Aims to Preserve Civilization for Millennia."

*Scientific American*, 28 September 2018. <https://www.scientificamerican.com/article/dna-coded-lunar-library-aims-to-preserve-civilization-for-millennia>.

Müller, Jan. "NFSA preserves historic Australian Olympic win on DNA." National Film and

Sound Archive of Australia (NFSA). 25 September 2020. Video, 3:20.

<https://www.youtube.com/watch?v=oEFGcXSefTI>.

Müller, Jan and Yasmin Meichtry. "Content as Data Stream: The First Heritage Video Stored on

DNA – A Case Study About the Future of Digital Storage." Online presentation at

Association of Moving Image Archivists Conference, 17 November 2020.

Musa, Helen. "Muller puts the 'N' back into the NFSA." *City News*, 31 May 2018.

<https://citynews.com.au/2018/muller-puts-the-n-back-into-the-nfsa>.

Nakamura, Lisa and Peter Chow-White (editors). *Race After the Internet*. New York: Routledge, 2011.

National Film and Sound Archive of Australia. "NFSA: Strategic Vision for a Digital Archive." 2018. Accessed 2 May 2021. <https://www.nfsa.gov.au/collection/curated/nfsa-strategic-vision-2018>.

National Film and Sound Archive of Australia. "Digitisation Strategy 2018-2025." 2018. Accessed 2 May 2021. <https://www.nfsa.gov.au/collection/curated/nfsa-digitisation-strategy-2018-2025>.

National Human Genome Research Institute. "The Cost of Sequencing a Human Genome." Accessed 18 April 2021. <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.

National Human Genome Research Institute. "Deoxyribonucleic Acid (DNA) Fact Sheet." Accessed 17 April 2021. <https://www.genome.gov/about-genomics/fact-sheets/Deoxyribonucleic-Acid-Fact-Sheet>.

Neiman, Mikhail Samoilovich. "Some fundamental issues of microminiaturization." *Radiotekhnika* 19, no. 1 (1964). [https://2a008ed5-a-62cb3a1a-sites.googlegroups.com/site/msneiman1905/Neiman-1964\\_Micromini.pdf](https://2a008ed5-a-62cb3a1a-sites.googlegroups.com/site/msneiman1905/Neiman-1964_Micromini.pdf).

Neumann-Held, Eva M. and Christoph Rehmann-Sutter. *Genes in Development: Re-Reading the Molecular Paradigm*. Durham: Duke University Press, 2006.

Nieves-Colón, Maria A., Andrew T. Ozga, William J. Pestle, Andrea Cucina, Vera Tiesler, Travis W. Stanton, Anne C. Stone. "Comparison of two ancient DNA extraction

- protocols for skeletal remains from tropicals.” *American Journal of Physical Anthropology* 166, no. 4 (August 2018): 824-836. [https://doi.org/ 10.1002/ajpa.23472](https://doi.org/10.1002/ajpa.23472).
- O’Bonsawin, Christine. “From Black Power to Indigenous Activism: The Olympic Movement and the Marginalization of Oppressed Peoples (1968-2012).” *Journal of Sport History* 42, no. 2 (Summer 2015): 200-219.  
<https://www.jstor.org/stable/10.5406/jsporthistory.42.2.0200>.
- O’Neal, Jennifer R. “‘The Right to Know’: Decolonizing Native American Archives.” *Journal of Western Archives* 6, no. 1 Native American Archives Special Issue (2015): 1-17.  
<https://doi.org/10.26077/fc99-b022>.
- Organick, Lee, Bichlien H. Nguyen, Rachel McAmis, Weida D. Chen, A. Xavier Kohll, Siena Dumas Ang, Robert N. Grass, Luis Ceze, Karin Strauss. “An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage.” *Small Methods* (25 January 2021). <https://doi.org/10.1002/smt.202001094>.
- Organick, Lee, Siena Dumas Ang, Yuan-Jyue Chen, Randolph Lopez, Sergey Yekhanin, Konstantin Makarychev, Miklos Z. Racz, Govinda Kamath, Parikshit Gopalan, Bichlien Nguyen, Christopher N. Takahashi, Sharon Newman, Hsing-Yeh Parker, Cyrus Rashtchian, Kendall Stewart, Gagan Gupta, Robert Carlson, John Mulligan, Douglas Carmean, Georg Seelig, Luis Ceze, and Karin Strauss. “Random access in large-scale DNA data storage.” *Nature Biology* 36, no. 3 (29 February 2018): 242-248.  
<https://doi.org/10.1038/nbt.4079>.
- Osmond, Gary and Matthew Klugman. “A Forgotten Picture: Race, Photographs and Cathy Freeman at the Northcote Koori Mural.” *Journal of Australian Studies* 43, no. 2 (2019): 203-217. <https://doi.org/10.1080/14443058/2019.1581247>.

Pasadena Community College. *General Chemistry and Chemical Analysis*. Chemistry

LibreTexts. 10 June 2020. <https://batch.libretexts.org/print/Finished/chem-219108/Full.pdf>.

Paunescu, Daniela, Michela Puddu, Justus O. B. Soellner, Philipp R. Stoessel, and Robert N.

Grass. “Reversible DNA encapsulation in silica to produce ROS-resistant and heat-resistant synthetic DNA ‘fossils.’” *Nature Protocols* 8 (7 November 2013): 2440-2448. <https://doi.org/10.1038/nprot.2013.154>.

Perkel, Jeffrey M. “The race for enzymatic DNA synthesis heats up.” *Nature.com*, 25 February 2019. <https://www.nature.com/articles/d41586-019-00682-0>.

Phillips, Megan, Jefferson Bailey, Andrea Goethais, and Trevor Owens. “The NDSA Levels of Digital Preservation: An Explanation and Uses.” *National Digital Stewardship Alliance*, 2013.

[https://www.digitalpreservation.gov/documents/NDSA\\_Levels\\_Archiving\\_2013.pdf](https://www.digitalpreservation.gov/documents/NDSA_Levels_Archiving_2013.pdf).

Pilger, John. “Cruelty and xenophobia stir and shame the lucky country.” *The Guardian*, 18 January 2017. Accessed 2 May 2021.

<https://www.theguardian.com/commentisfree/2007/jan/19/australia.guantanamo>.

Ping, Zhi, Dongzhao Ma, Xiaoluo Huang, Shihong Chen, Longying Liu, Fei Guo, Sha Joe Zhu, and Yue Shen. “Carbon-based archiving: current progress and future prospects of DNA-based data storage.” *GigaScience* 8 (2019): 1-10. doi: 10.1093/gigascience/giz075.

Potomac Institute for Policy Studies. “The Future of DNA Data Storage.” Arlington, September 2018. [https://potomac institute.org/images/studies/Future\\_of\\_DNA\\_Data\\_Storage.pdf](https://potomac institute.org/images/studies/Future_of_DNA_Data_Storage.pdf).

Preston, Douglas. “The Kennewick Man Finally Freed to Share His Secrets.” *Smithsonian Magazine*, September 2014. Accessed 2 May 2021.



<https://www.smithsonianmag.com/history/kennewick-man-finally-freed-share-his-secrets-180952462>.

Protocols for Native American Archival Materials. “Home.” *Northern Arizona University*.

Accessed 19 April 2021. <https://www2.nau.edu/libnap-p>.

Protocols for Native American Archival Materials. “Protocols for Native American Archival

Materials.” *Northern Arizona University*. Last modified 9 April 2007. Accessed 19 April 2021. <https://www2.nau.edu/libnap-p/protocols.html>.

Rajagopalan, Ramy M., Alondra Nelson, and Joan H. Fujimura. “Race and Science in the

Twenty-First Century.” In *The Handbook of Science and Technology Studies*. Edited by Ulrike Felt. Cambridge: MIT Press, 2016.

Reardon, Jenny. *The Postgenomic Condition: Ethics, Justice & Knowledge after the Genome*.

Chicago: University of Chicago Press, 2017.

Reardon, Jenny and Kim TallBear. “‘Your DNA Is Our History’: Genomics, Anthropology, and

the Construction of Whiteness as Property.” *Current Anthropology* 53, no. S5, The Biological Anthropology of Living Human Populations: World Histories, National Styles, and International Networks (April 2012): S233-S245.

<https://www.jstor.org/stable/10.1086/662629>.

Roosth, Sophia. *Synthetic: How Life Got Made*. Chicago: University of Chicago Press, 2017.

Rosenthal, David. “Archival Media: Not a Good Business.” *DSHR’s Blog*, 1 March 2018.

Accessed 3 May 2021. <https://blog.dshr.org/2018/03/archival-media-not-good-business.html>.

Rosenthal, David. “The Future of Storage.” *DSHR’s Blog*, 12 May 2016. Accessed 2 May 2021.

<https://blog.dshr.org/2016/05/the-future-of-storage.html>.

Rosenthal, David. "Ithaka's Perspective on Digital Preservation." *DSHR's Blog*, 1 November 2018. Accessed 30 April 2021. <https://blog.dshr.org/2018/11/ithakas-perspective-on-digital.html>.

Rosenthal, David. "The Medium-Term Prospects for Long-Term Storage Systems." *DSHR's Blog*, 13 December 2016. Accessed 2 May 2021. <https://blog.dshr.org/2016/12/the-medium-term-prospects-for-long-term.html>.

Rosenthal, David. "More on long-lived media." *DSHR's Blog*, 19 June 2014. Accessed 2 May 2021. <https://blog.dshr.org/2014/06/more-on-long-lived-media.html>.

Rosenthal, David, Thomas Robertson, Tom Lipkis, Vicky Reich, Seth Morabito. "Requirements for Digital Preservation Systems: A Bottom-Up Approach." *D-Lib Magazine* 11, no. 11 (November 2005). <http://dx.doi.org/10.1045/november2005-rosenthal>.

Roth, Lorna. "Looking at Shirley, the Ultimate Norm: Colour Balance, Image Technologies, and Cognitive Equity." *Canadian Journal of Communication* 34, no. 1 (2009): 111-136. <https://doi.org/10.22230/cjc.2009v34n1a2196>.

SBS News. "Australia marks 20-year anniversary of Sorry Day." *SBS News*, 26 May 2018. <https://www.sbs.com.au/news/australia-marks-20-year-anniversary-of-sorry-day>.

Scudellari, Megan. "Inner Workings: DNA for data storage and computing." *Proceedings of the National Academy of Sciences of the United States of America* 112, no. 52 (29 December, 2015): 15771-15772. <https://www.jstor.org/stable/10.2307/26466228>.

Schwartz, Oscar. "A Geneticist's Dilemma." *The Washington Post Magazine*, 23 November 2020. Accessed 2 May 2021. <https://www.washingtonpost.com/magazine/2020/11/23/many-scientists-believe-that-dna-holds-cure-disease-that-poses-problem-some-native-americans>.

- Shendure, Jay, Shankar Balasubramanian, George M. Church, Walter Gilbert, Jane Rogers, Jeffery A. Schloss, and Robert H. Waterston. *Nature* 550, no. 7676 (October 2019): 345-353. <https://doi.org/10.1038/nature24286>.
- Sofia, Zoë. "Container Technologies." *Hypatia A Journal of Feminist Philosophy* 15, no. 2 (Spring 2000): 181-201. <https://muse.jhu.edu/article/14107>.
- Spivack, Nova. "Arch Mission Foundation Announces Our Payload On SpaceX Falcon Heavy." *Medium*, 6 February 2018. Accessed 30 April 2021. <https://medium.com/arch-mission-foundation/arch-mission-foundation-announces-our-payload-on-spacex-falcon-heavy-c4c9908d5dd1>.
- Spivack, Nova. "There is Now a 30 Million Page Backup of Planet Earth, on the Moon! – New NASA Images Show Location of the Lunar Library." *Medium*, 16 May 2019. Accessed 30 April 2021. <https://medium.com/@novaspivack/there-is-now-a-30-million-page-backup-of-planet-earth-on-the-moon-d0504458aca1>.
- Steinhauer, Jillian. "Lynn Hershman Leeson: The Artist Is Prescient." *New York Times*. 8 July 2021. Accessed 8 July 2021. <https://www.nytimes.com/2021/07/08/arts/design/hershman-leeson-review-art-museum.html>.
- Sterling, Robyn L. "Genetic Research among the Havasupai: A Cautionary Tale." *Virtual Mentor* 13, no. 2 (February 2011): 113-117. <https://doi.org/10.1001/virtualmentor.2011.13.2.hlaw1-1102>.
- Sterne, Jonathan. *MP3: The Meaning of a Format*. Durham: Duke University Press, 2012.
- Tadic, Linda. "Audiovisual digital preservation and DNA storage requirements." Zoom slide presentation at Third JPEG DNA Workshop (9 April 2021).

- Takakura, Kazuki, Atsushi Kawamura, Yuichi Torisu, Shigeo Koido, Naohisa Yahagi, and Masayuki Saruta. "The Clinical Potential of Oligonucleotide Therapeutics against Pancreatic Cancer." *International Journal of Molecular Sciences* 20, no. 3331 (2019). doi:10.3390/ijms20133331.
- TallBear, Kim. "Diversity v. Decolonization in the Academy: A Conversation with Kim TallBear." Zoom Lecture, McGill University Department of English, 10 February 2021. Video, 53:30. <https://www.youtube.com/watch?v=qrL5Kc4wJgQ>.
- TallBear, Kim. "Genomic articulations of indigeneity." *Social Studies of Science* 43, no. 4, Special Issue: Indigenous Body Parts and Postcolonial Technoscience (August 2013): 509-533. <https://www.jstor.org/stable/43284191>.
- TallBear, Kim. "Narratives of Race and Indigeneity in the Genographic Project." *The Journal of Law, Medicine & Ethics* 35, no. 3 (14 August 2007): 412-424. <https://doi.org/10.1111/j.1748-720X.2007.00164.x>.
- TallBear, Kim. *Native American DNA: Tribal Belonging and the False Promise of Genetic Science*. Minneapolis: University of Minnesota Press, 2013.
- TallBear, Kim. "Who Owns the Ancient One?" *Buzzfeed News*, 23 July 2015. <https://www.buzzfeednews.com/article/kimtallbear/how-the-man-stole-ancient-man-from-his-native-descendents>.
- Tuck, Eve and K. Wayne Yang. "Decolonization is not a metaphor." *Decolonization: Indigeneity, Education & Society* 1, no. 1: 1-40. <https://jps.library.utoronto.ca/index.php/des/article/view/18630>.

Twist Bioscience. “DNA-Based Digital Storage White Paper.” Last modified 29 October 2018.

Accessed 27 April 2021. <https://www.twistbioscience.com/resources/white-paper/dna-based-digital-storage>.

Twist Bioscience. “Reducing Our Chemical Footprint for a Sustainable Future in Biotechnology.

Company News & Updates. 22 April 2021. Accessed 23 April 2021.

<https://www.twistbioscience.com/blog/company-news-updates/reducing-our-chemical-footprint-sustainable-future-biotechnology>.

Vitak, Sarah. “Technology alliance boosts efforts to store data in DNA.” *Nature.com*, 03 March 2021. <https://www.nature.com/articles/d41586-021-00534-w>.

Walker, Francis. “Huntington’s disease.” *The Lancet* 369, no. 9557 (20 January 2007): 218-228. [https://doi.org/10.1016/S0140-6736\(07\)60111-1](https://doi.org/10.1016/S0140-6736(07)60111-1).

Weigmann, Katrin. “The code, the text and the language of God.” *European Molecular Biology Organization Reports* 5, no. 2 (2004): 116-118.

White House Office of the Press Secretary. “President Clinton Announces the Completion of the First Survey of the Entire Human Genome: Hails Public and Private Efforts Leading to This Historic Achievement.” *Human Genome Project Information Archive 1990-2003*. 26 June 2000. Accessed 2 May 2021.

[https://web.ornl.gov/sci/techresources/Human\\_Genome/project/clinton1.shtml](https://web.ornl.gov/sci/techresources/Human_Genome/project/clinton1.shtml).

Yim, Sung Sun, Ross M. McBee, Alan M. Song, Yiming Huang, Ravi U. Sheth, and Harris H. Wang. “Robust direct digital-to-biological data storage in living cells.” *Nature Chemical Biology* 17 (11 January 2021): 246-253.

Zimmer, Carl. “New DNA Results Show Kennewick Man Was Native American.” *The New York Times*, 18 June 2015. Accessed 2 May 2021.

<https://www.nytimes.com/2015/06/19/science/new-dna-results-show-kennewick-man-was-native-american.html>.