

MASTERARBEIT | MASTER'S THESIS

Titel | Title

DNA Data Storage

Eine mögliche Zukunft der digitalen Archivierung

verfasst von | submitted by

Hanna Mitter BA

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on the
student record sheet:

UA 066 804

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Historische Hilfswissenschaften und
Archivwissenschaft

Betreut von | Supervisor:

Mag. Dr. Elisabeth Schöggli-Ernst MAS

Danksagung

Ich möchte mich ganz herzlich bei meiner Betreuerin Dr. Elisabeth Schöggli-Ernst für die umfassende Betreuung bedanken.

Ebenso möchte ich mich bei Dr. Adrian Keller und Dr. Thomas Peharz des NEO-Projektes dafür bedanken, dass sie meine Fragen über ihr Projekt „Next Generation Molecular Data Storage“ so ausführlich und umfassend beantwortet haben und damit einen wichtigen Teil zu meiner Masterarbeit beigetragen haben.

Inhalt

1. Einleitung	8
2. Abstract.....	11
3. Digitale Archivierung.....	14
3.1. Gegenwärtiger Stand der digitalen Archivierung	15
3.1.1. OAIS-Referenzmodell.....	16
3.1.1.1. Datenübernahme	17
3.1.1.2. Datenaufbewahrung.....	17
3.1.1.3. Datenmanagement.....	18
3.1.1.4. Systemverwaltung	19
3.1.1.5. Digital Preservation	19
3.1.1.6. Zugriff	20
3.1.2. Dateiformate	20
3.1.2.1. PDF/A (Portable Document Format/ Archive)	21
3.1.2.2. JPEG2000 (Joint Experts Photograph Group)	21
3.1.2.3. ODT, ODP (ODF) (Offene Textdatei, Open Document Präsentation, Open Document Format)	22
3.1.2.4. TXT (Text)	22
3.1.3. Migration und Emulation	23
3.2. Vorteile der digitalen Archivierung.....	24
3.3. Beispiele digitaler Archivierung in Österreich	25
3.4. Problemstellungen der digitalen Archivierung	26
3.4.1. Kosten	27
3.4.2. CO2-Verbrauch des digitalen Speichers	28
3.5. Fazit.....	28
4. Optical Storages.....	30
4.1. Archium.....	31
4.2. MOM-Memory of Mankind.....	33
4.3. WORF (Write Once, Read Forever)	35
4.4. Microsoft Silica.....	35
4.5. DNA Data Storage	36
4.6. Fazit.....	37
4.7. Abbildungen der vorgestellten Optical Storages.....	39
5. DNA Data Storage.....	43
5.1. Entwicklung der DNA Datenspeicherung.....	43
5.2. Vorteile der Methode	44
5.3. Technische Hintergründe	45

5.3.1. Codierung.....	46
5.3.2. Synthetisierung (Schreiben von DNA).....	51
5.3.3. Speicherung der Daten.....	53
5.3.4. Sequenzierung (Lesen von DNA)	55
5.3.5. Decodierung.....	57
5.3.6. Error Correction	58
5.4. Nachteile.....	59
6. DNA Data Storages als Möglichkeit der digitalen Archivierung	62
6.1. Theoretische Umsetzungen	62
6.1.1. Archival Storage System nach Church.....	62
6.1.2. Archival Storage System nach Bornholt.....	63
6.1.3. Archival Storage System nach Goldman.....	66
6.1.4. DNA MemoChip	67
6.2. Projekte zur praktischen Umsetzung des DNA Data Storage	68
6.2.1. NEO-Projekt – Next Generation Molecular Data Storage.....	68
6.2.2. Disco – DNA Data Storage	73
6.2.3. Microsoft Research von DNA Data Storage	74
6.2.4. OLIGOARCHIVE – Synthetische DNS als digitale Speicherlösung der nächsten Generation	75
6.3. Arbeitskonzept zur Speicherung von GIS-Daten auf DNA	78
7. Fazit	88
8. Literaturverzeichnis	92
9. Abbildungsverzeichnis.....	102
10. Abkürzungsverzeichnis.....	103
11. Anhang.....	105

1. Einleitung

In der vorliegenden Masterarbeit soll eine mögliche, zukünftige Methode der digitalen Datenspeicherung in Archiven analysiert werden. Es handelt sich hierbei um die Speicherung von digitalen Daten auf synthetisch hergestellter DNA.

Die vorliegende Arbeit widmet sich der genaueren Arbeitsweise der DNA-Datenspeicherung und deren bisherige Entwicklung innerhalb der Forschung. Allerdings soll auch hinterfragt werden, inwieweit DNA-Datenspeicherung eine Zukunft in österreichischen Archiven, beziehungsweise in Archiven generell haben könnte und wie so eine praktische Umsetzung, im Rahmen des heutigen Forschungsstandes, gestaltet werden würde.

Die Masterarbeit befasst sich auch mit einem naturwissenschaftlichen und informatischen Thema, welches ein inhaltliches Vorwissen für das Verständnis erfordert. Einzelne Begriffe und Abläufe werden daher im Laufe der Arbeit so gut wie möglich erklärt und aufbereitet. Für genauere Ausführungen und Vertiefungen sind an den jeweiligen Stellen weiterführende Literaturvorschläge angegeben.

Mit der laufenden Zunahme von digitalen Daten, sei es jetzt über das Internet, den regelmäßigen E-Mail-Verkehr, die Verwendung des Elektronischen Aktes in der österreichischen Verwaltung oder die Sozialen Medien, müssen sich auch Archive darüber Gedanken machen, wie sie in Zukunft diese neue Menge an digitalen Daten in ihrem System speichern können. Die Verwendung von herkömmlichen technischen Speichern weist einige Nachteile auf, unter anderem die laufenden Kosten der Datenmigration, welche regelmäßig durchgeführt werden muss, um einen Verlust von Daten zu vermeiden, aber auch die Zunahme von Speicherplatz oder die Kosten für den Strom für die Speicherung der Daten.¹

Diese und andere Aspekte sprechen dafür, sich nach einer umweltfreundlicheren und kostengünstigeren Methode der digitalen Speicherung umzuschauen. Hier kommen die sogenannten Optical Storages oder auch Write Once, Read Forever (WORF)-Speichermethoden ins Spiel. Auch DNA-Datenspeicherung fällt unter diese Begriffe. Diese Ansätze bringen unterschiedliche Vorteile im Vergleich zu jetzigen digitalen

¹ Reinhard ALTENHÖNER, Markus BRANTL, Klaus CEYNOWA, Digitale Langzeitarchivierung in Deutschland. Projekte und Perspektiven. *Zeitschrift für Bibliothekswesen und Bibliographien* 58/ 3-4 (2011) 184-196, 184.

Speichermethoden mit sich. Ein großer Pluspunkt dieser Methoden ist die Tatsache, dass es keiner Datenmigration bedarf, um die digitalen Daten zu erhalten. Zusätzlich können mit diesen Methoden große Mengen an Daten auf vergleichsweise kleinem Platz gespeichert werden. Hier sticht besonders die DNA Datenspeicherung hervor. Die Forschungsarbeiten und Paper, welche sich mit diesem Thema beschäftigen, geben dazu unterschiedliche Angaben, wie viel Daten tatsächlich auf DNA gespeichert werden können, aber ein ungefährender Richtwert sind 109 Gigabyte auf einem Kubikmillimeter DNA². DNA hat den weiteren Vorteil, dass sie eine hohe Stabilität aufweist. Und diese verliert sie auch unter schwierigeren, klimatischen Bedingungen nicht.³ Aufgrund dieser Gegebenheiten können Optical Storages und im Besonderen die DNA Datenspeicherung als kostengünstige und effiziente Speichermethode gesehen werden. Dennoch gibt es bei der DNA Datenspeicherung auch Nachteile. Die Herstellung der synthetischen DNA ist unter heutigen Voraussetzungen immer noch sehr teuer im Vergleich zu herkömmlichen digitalen Speicherträgern. Die finanziellen Vorteile ergeben sich daher erst über einen längeren Zeitraum und nicht direkt bei der Anschaffung dieses neuen Speichers.⁴ Besonders bei der Herstellung von DNA kommt es zu unterschiedlichen Schwierigkeiten. Einerseits ist die Synthetisierung von DNA fehleranfällig, wodurch immer ein Mehr an DNA hergestellt wird, um etwaigen Datenverlust zu verhindern. Andererseits ist es noch nicht möglich, längere Stränge von DNA ohne Probleme herzustellen, wodurch gerade große digitale Daten nicht in einem Stück auf DNA gespeichert werden können⁵. Und auch im Bereich der Lagerung und dem Auslesen ergeben sich noch Probleme, weil es teilweise noch nicht möglich ist, direkt auf nur ein Dokument in einem Pool aus DNA zuzugreifen, wodurch bei der Sequenzierung und der Decodierung Mehrkosten und Mehraufwand entstehen.⁶

Neben den theoretischen Ausführungen, wie DNA Datenspeicherung funktioniert und aufgebaut ist, soll auch ein Überblick gegeben werden, welche Forschungsprojekte sich mit DNA Datenspeicherung beschäftigen. Darunter fallen theoretische Ansätze, welche deren Theorie nur mit einer kleinen Datenmenge testen und an der Verbesserung und Weiterentwicklung der Arbeitsweise von DNA Datenspeicherung

² James BORNHOLT, Randolph LOPEZ, Douglas M. CARMEAN et al., A DNA-Based Archival Storage System. *ASPLOS* '16/ 2-6 (2016) 1f.

³ Reinhard HECKEL, An archive written in DNA. *Nature Biotechnology* 36/ 3 (2018) 236-237, 236.

⁴ Potomac Institute für Policy Studies, The Future of DNA Data Storage, 16.

⁵ HECKEL, An archive written in DNA, 236.

⁶ HECKEL, An archive written in DNA, 236.

arbeiten. Der zweite Teil des Kapitels beschäftigt sich mit praktischen Ansätzen und Projekten, welche entweder schon abgeschlossen sind oder an denen noch aktiv geforscht wird. Bei diesen Projekten handelt es sich hauptsächlich um Kooperationen zwischen Universitäten und Firmen, welche gemeinsam Ansätze der DNA Datenspeicherung testen.

Der Fokus der Arbeit liegt auf dem Aspekt der Umsetzung dieses Ansatzes im Archivalltag, daher wird abschließend ein Arbeitskonzept ausgearbeitet, welches darlegen soll, wie die Implementierung eines DNA Datenspeichers in einem Archiv ablaufen könnte. Es soll die nötige Vorarbeit für ein solches Projekt beinhalten, die zu setzenden Schritte aufzeigen und einen Überblick darüber geben, welche Anschaffungen ein Archiv tätigen muss oder welche zusätzlichen Anforderungen an Archivarinnen und Archivare gestellt werden müssen, damit ein digitales Archiv in DNA gespeichert werden kann. Es handelt sich hierbei um ein theoretisches Konzept, welches sich auf einen größeren Datensatz konzentriert, da bei großen Datenmengen die Vorteile der DNA Datenspeicherung besonders sichtbar sind. Bei den Daten handelt es sich um Geoinformationssystemdaten des Landes Steiermark, welche unter anderem Orthophotos und Laserdaten beinhalten. Bei einem Orthophoto handelt es sich um „eine verzerrungsfreie und maßstabsgetreue Abbildung der Erdoberfläche.“⁷ Das Konzept konzentriert sich auf den Back-Up Speicher der Daten, wodurch der Zugriff auf die Daten auf DNA nur im Notfall notwendig wird.

Insgesamt soll ein gesamtheitlicher Überblick über das Thema gegeben werden, in dem neben der Funktionsweise auch die Vorteile und die Nachteile der Methode besprochen werden sollen. Wie sich die Methode in Zukunft weiterentwickeln wird und ob es zu einer Anwendung in Archiven kommen wird, ist mit dem heutigen Stand der Forschung noch nicht vorauszusehen.

⁷Orthophoto, in: Bundesamt für Eich- und Vermessungswesen, <https://www.bev.gv.at/Services/Produkte/Luftbildprodukte/Orthophoto-Farbe.html> [19.08.2024].

2. Abstract

This master's thesis analyses a possible future method of digital data storage in archives, namely the storage of digital data on synthetically produced DNA.

This paper is dedicated to the more precise functioning of DNA data storage and its development within research to date. However, it also questions the extent to which DNA data storage could have a future in Austrian archives, or in archives in general, and how such a practical implementation would be organised in the context of the current state of research.

However, the topic of the Master's thesis also deals with a scientific and computer science topic, which requires prior knowledge of the content to understand it. Individual terms and processes are therefore explained and prepared as well as possible during the thesis. For more detailed explanations and in-depth information, further reading suggestions are provided at the relevant points.

With the constant increase in digital data, be it via the Internet, regular e-mail traffic, the use of electronic files in the Austrian administration or social media, archives must also consider how they want to and can store this new volume of digital data in their system in the future. The use of conventional technical storage systems has several disadvantages, including the ongoing costs of data migration, which must be carried out regularly to avoid losing data, but also the increase in storage space or the cost of electricity for storing the data. These and other aspects speak in favour of looking for a more environmentally friendly and cost-effective method of digital storage.

This is where so-called optical storage or Write Once, Read Forever (WORM) storage methods come into play. DNA data storage also falls under this term. These approaches have various advantages compared to current digital storage methods. A major point of these methods is the fact that no data migration is required to preserve the digital data. The storage medium preserves the data in its original state after the storage process. In addition, these methods can be used to store large amounts of data in a relatively small space. DNA data storage in particular stands out here. The research and papers that deal with this topic give different figures on how much data can be stored on DNA, but an approximate guideline is 109 gigabytes on one cubic

millimetre of DNA. DNA has the further advantage of being very stable. And it does not lose this stability even under difficult climatic conditions. Due to these circumstances, optical storage and DNA data storage can be seen as a cost-effective and efficient storage method. However, DNA data storage also has disadvantages. Under current conditions, the production of synthetic DNA is still very expensive compared to conventional digital storage media. The financial advantages therefore only become apparent over a longer period and not immediately when purchasing this new storage device. There are various difficulties, particularly in the production of DNA. On the one hand, the synthesis of DNA is prone to errors, which means that more DNA is always produced to prevent any loss of data. On the other hand, it is not yet possible to produce longer strands of DNA without problems, which means that large digital data cannot be stored in one piece on DNA. And there are still problems in storage and reading, because it is sometimes not yet possible to directly access just one document in a pool of DNA, which results in additional costs and effort for sequencing and decoding.

In addition to the theoretical explanations of how DNA data storage works and is structured, an overview will also be given of which research projects are dealing with DNA data storage. This includes theoretical approaches that test their theory with only a small amount of data and work on improving and further developing the way DNA data storage works. The second part of the chapter deals with practical approaches and projects that have either already been completed or are still actively researching. These projects are mainly collaborations between universities and companies that are jointly testing approaches to DNA data storage. The focus of the work is on the aspect of implementing this approach in everyday archive work, so a working concept will be developed at the end that will explain how the implementation of a DNA data storage facility in an archive could work. It should include the necessary preparatory work for such a project, show the steps to be taken and provide an overview of what purchases an archive must make or what additional requirements must be placed on archivists so that a digital archive can be stored in DNA.

This is a theoretical concept that focuses on a larger data set, as the advantages of DNA data storage are particularly evident with large amounts of data. The data is geoinformation system data from the state of Styria, which includes orthophotos and laser data. An orthophoto is "a distortion-free and true-to-scale image of the earth's

surface." The concept focuses on the back-up storage of the data, which means that access to the data on DNA is only necessary in an emergency. Overall, a comprehensive overview of the topic is to be given, in which the advantages and disadvantages of the method are to be discussed in addition to how it works. How the method will develop in the future and whether it will be used in archives cannot yet be predicted from today's perspective.

3. Digitale Archivierung

Im folgenden Abschnitt möchte ich einen Überblick darüber geben, wie der momentane Stand der digitalen Archivierung in Österreich ist und weiters darlegen, welche Vorteile diese Art der Archivierung bringt, aber auch noch vorhandene Nachteile aufzeigen.

Ich werde im Laufe der Arbeit von digitaler Archivierung sprechen und nicht von digitaler Langzeitarchivierung, wie dies in der Literatur hauptsächlich der Fall ist, da unter Archivierung an sich eine Langzeitarchivierung zu verstehen ist. Bei der digitalen Archivierung soll das Ziel erreicht werden, die zu archivierenden Daten auf lange Zeit zu erhalten, weswegen diese extra Unterscheidung zu einer nicht notwendigen Doppelung führt.

Warum braucht es also digitale Archivierung? Was gehört alles dazu und welche Schritte müssen für eine sinnvolle Arbeitsweise im digitalen Archiv gesetzt werden? Die Legitimation von Archiven beruht, im analogen wie im digitalen Bereich, auf der Benützung der Archivalien. Ohne eine Benützung wird die Rechtfertigung des Aufhebens schwierig. Benützung soll von allen, die ein Interesse darin haben, in Anspruch genommen werden können, denn das Archiv ist Bestandteil dessen, was das kollektive Gedächtnis der Gesellschaft bildet und erhält.⁸

Gerade die letzten Jahrzehnte sind durch die Entwicklung von digitaler Kommunikation, aber auch von vermehrter Verwendung digitaler Unterlagen geprägt. Auf diese Veränderung müssen auch Archive eingehen und sich den neuen Gegebenheiten anpassen.⁹ Das bedeutet auch, dass die Menge an Daten, welche genuin digital entstehen, immer und stetig wächst. Unter genuin digitalen Daten versteht man Objekte, deren Original digital ist, im Gegensatz zu Objekten, deren analoges Original digitalisiert wurden. Diese genuin digitalen Daten müssen nach einer gewissen Zeit ins Archiv übernommen werden. Darunter fallen digitale Dokumente, welche im Laufe des Verwaltungsprozesses entstehen, wie der Elektronische Akt (ELAK) der österreichischen Behörden, oder aber Webseiten,

⁸ Thomas SCHÄRLI, Kollektives Gedächtnis, Geschichtswissenschaft und digitale Archive, in: Digitales Gedächtnis. Archivierung und die Arbeit der Historiker der Zukunft, hg. von Patrick JUCKER-KUPPER - Christoph KOLLER - Gerold RITTER (Geschichte und Informatik 13/14, Zürich 2004) 57-82, 57f.

⁹ SCHÄRLI, Kollektives Gedächtnis, Geschichtswissenschaft und digitale Archive, 63ff.

welche mittlerweile von großer Bedeutung sind.¹⁰ Auch E-Mails müssen ins Archiv überführt werden. Denn diese überliefern einen wichtigen Bereich der Kommunikation in der Registratur, mittlerweile machen sie auch schon den Großteil der Kommunikation innerhalb von Firmen aus.¹¹ Außerdem fallen unter diese Kategorie auch neue Entwicklungen, wie Social Media oder YouTube. Auch für diese Plattformen muss eine Lösung der Archivierung gefunden werden.

Ein weiterer Punkt, warum digitale Archivierung immer bedeutender wird, liegt an den digitalen Daten selbst. Einerseits können sie aufgrund der technischen Umstände sehr schnell „unlesbar“ werden, in dem Sinne, dass entweder der Datenträger kaputt wird oder die Soft- oder Hardware die Daten nicht mehr öffnen und lesen kann. Andererseits nimmt die Menge der genuin digitalen Daten, wie schon erwähnt, sehr schnell zu und wird weiterhin stetig wachsen. Und ein weiterer Punkt ist die Tatsache, dass digitale Daten unterschiedlichste Formen annehmen können. Entweder Textdokumente, Film- und Tonaufnahmen oder Bildaufnahmen. Und zusätzlich können diese unterschiedlichen Formen auch innerhalb eines Dokumentes auftreten.¹²

3.1. Gegenwärtiger Stand der digitalen Archivierung

Besonders von Bedeutung bei der digitalen Archivierung ist die Verwendung von Standards und Referenzmodellen. Neben der Vereinfachung der Arbeitsschritte geben diese Standards Orientierung bei der Ausführung und können bei der Absicherung von Entscheidungen helfen.¹³ Vor allem bekannt ist das OAIS-Referenzmodell (Open Archival Information System), an dem sich viele andere Implementierungen orientieren. Dieses umfasst nicht nur Softwarelösungen, sondern auch die dahinterstehenden Prozesse und die ausführenden Menschen, beziehungsweise die anzusprechenden Zielgruppen.¹⁴ Weitere Standards sind die

¹⁰ Burkhard HEISE, Dagmar ULLRICH, Digitale Langzeitarchivierung. *PIK* 27/2 (2004) 103-109, 103.

¹¹ Daniel ABAWI, Oliver GREINER, E-Mails. Archivierung-Not und Tugend. *Wirtschaftsinformatik und Management (Internet)* 3/1 (2011) 56-63, 56.

¹² HEISE, ULLRICH, Digitale Langzeitarchivierung, 104.

¹³ ERNST, Standards und Normen im Bereich der Langzeitarchivierung, 62.

¹⁴ ERNST, Standards und Normen im Bereich der Langzeitarchivierung, 63f.

3.1.1. OAS-Referenzmodell



¹⁸ Christopher LEE, Open Archival Information System (OAIS) Reference Model. Encyclopaedia of Library and Information Sciences (2010). 4020-4030, 4025.

In Abbildung 1 werden die einzelnen Schritte des OAIS-Referenzmodells bildlich dargestellt. Die einzelnen Schritte, welche das Modell umfasst, werde ich auf den nachfolgenden Seiten beschreiben.

3.1.1.1. Datenübernahme

Der erste Aspekt des OAIS-Referenzmodells ist die Vorbereitung zur Übernahme der digitalen Daten aus der Registratur. Bevor diese Daten, auch Submission Information Packages (SIP) genannt, tatsächlich ins Archiv übernommen werden können, müssen diese auch auf ihre Lesbarkeit, ihre Vollständigkeit, ihre Authentizität und den richtigen Kontext geprüft werden. Ein weiterer Teil der Datenübernahme ist das Anreichern der Erschließungsdatenbank des digitalen Archives und die Organisation der durchzuführenden Updates und Migrationen der Daten. Nachdem all diese Schritte abgeschlossen sind, können die SIPs in Archival Information Packages (AIP) umgewandelt werden, welche dann in das digitale Archiv übernommen werden können.¹⁹

3.1.1.2. Datenaufbewahrung

Der Bereich der Datenaufbewahrung umfasst die Übernahme und die Einlagerung der AIPs in das digitale Archiv. Diese müssen einer regelmäßigen Überprüfung unterzogen werden, deren Organisation und Planung schon Teil des ersten Schrittes war. Neben der Überprüfung der Wiederauffindbarkeit der Daten werden hier auch die technischen Gegebenheiten der Daten überprüft und ob es hier, zur Verhinderung von Datenverlust, eventuell eine Datenmigration braucht. Hierbei wird im OAIS-Referenzmodell im Speziellen darauf hingewiesen, dass eine redundante

¹⁹Heike NEUROTH, Achim OSSWALD, Regine SCHEFFEL, Stefan STRATHMANN, Mathias JEHN (Hg.), Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Kapitel 4. Das Referenzmodell OAIS-Open Archival Information System, (2010) 13.

Speicherung der Daten mögliche Datenverluste verhindern kann.²⁰ Das Prinzip der Migration und Emulation werde ich in einem späteren Teil dieses Kapitels noch genauer ausführen und erklären.

3.1.1.3. Datenmanagement

Das OAIS-Modell beschreibt den Abschnitt des Datenmanagements folgendermaßen:

„Im Bereich Datenmanagement geht es um die Wartung und das Zugänglichhalten der Verzeichnungsinformationen und ihre kontinuierliche Ergänzung und Aufbereitung, dann aber auch das Verwalten verschiedener Archivdatenbanken und auch in diesem Bereich die Ausführung von verschiedenen Datenbank-Updates zur Sicherung von Lesbarkeit, Verständlichkeit und Nutzbarkeit.“²¹

Der Bereich des Datenmanagements umfasst somit mehrere Bereiche. Neben der Aufbereitung der Daten für die Benutzerinnen und Benutzer müssen auch die Übernahmen von neuen Informationspaketen und deren Nutzung ausgehandelt werden. Aber auch das Warten der jeweiligen technischen Lösungen bezüglich der Hardware und der Software des Archivsystems fallen in diesen Aufgabenbereich, ebenso wie die Regelungen der einzelnen Arbeitsvorgänge innerhalb des Systems. Ein Teil des Datenmanagements betrifft auch die Sicherung der Daten und somit auch den Erhalt der Lesbarkeit und die Nutzbarkeit der Daten.²²

Diese Arbeitspläne sind nie endgültig. Sie müssen einer laufenden Kontrolle unterzogen werden, um mit den Entwicklungen der technischen Welt mithalten zu können und nicht obsolet zu werden. Somit handelt es sich hierbei um eine ständige Aufgabe für Archive, welche niemals wirklich abgeschlossen ist.²³

²⁰ NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 13f.

²¹ NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 14.

²² NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 14.

²³ KLIMPEL, KEIPER (Hg), Was bleibt?, 38.

3.1.1.4. Systemverwaltung

Grundsätzlich umfasst dieser Bereich des OAIS-Referenzmodelles das Management des Modells. Dazu gehört einerseits die Verwendung der Software und der Hardware des Systems, aber auch die Beziehung zwischen Archivarinnen und Archivaren und den Benutzerinnen und Benutzern. Genauer gesagt bedeutet das, dass für jeden Bereich Arbeitsvorgänge geregelt werden und die einzelnen Zuständigkeiten definiert werden. Zusätzlich fällt auch die Qualitätssicherung in diesen Bereich, sowie die Regelung der Verwendung von Informationspaketen.²⁴

3.1.1.5. Digital Preservation

Digital Preservation ist ein großer und wichtiger Teil der digitalen Archivierung. Hierbei handelt es sich um Pläne und Strategien, mit denen digitale Daten erhalten und deren Lesbarkeit und Zugänglichkeit ermöglicht werden.

Daher müssen in diesem Abschnitt regelmäßige Updates auf neuere und passendere Datenformate geplant werden. Dabei müssen auch diverse technische Veränderungen der jeweiligen Datenträger im Blick behalten werden, ebenso wie die Nutzungsgewohnheiten der Benutzerinnen und Benutzer.²⁵

Neben dem Erhalt der digitalen Daten muss auch deren Entstehungskontext bewahrt und überliefert werden. Ohne diesen können digitale Daten, ebenso wie analoge Daten, nicht in den geschichtlichen und zeitlichen Kontext ihres Entstehungszeitraumes gesetzt werden, wodurch eine Interpretation der Quelle beinahe unmöglich wird. Außerdem muss auch die Authentizität und die Integrität der digitalen Quelle erhalten bleiben, wie dies mit analogen Daten auch der Fall ist. Eine

²⁴ NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 14.

²⁵ NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 14.

Unterschrift in einem genuin digitalen Dokument muss auch nach einer Datenmigration immer noch ihren Wert und ihre Authentizität behalten.²⁶

Für all diese Punkte braucht es Digital Preservation und einen Digital Preservation Plan innerhalb eines Archives, damit rechtzeitig auf die Veränderungen der digitalen Daten eingegangen werden kann, um diese weiterhin erhalten zu können. Mit diesem Plan können Veränderungen auch nachvollziehbar bleiben und diverse Erhaltungsschritte auch im Nachhinein und von Dritten verständlich sein.²⁷

3.1.1.6. Zugriff

Der letzte Punkt des OAIS-Referenzmodells beschäftigt sich mit dem Zugriff der Benutzerinnen und Benutzer auf die Daten, welche im digitalen Archiv aufbewahrt werden. Dieser Bereich arbeitet nicht mehr mit AIPs, sondern vergibt die Informationen an die Benutzerinnen und Benutzer in sogenannten Dissemination Information Packages (DIP). Neben dem Ausgeben dieser Informationspackages umfasst dieser Abschnitt auch die Hilfe bei der Benützung und beim Auffinden von diversen Informationen. Darunter fällt auch das Bearbeiten von Anfragen und die Koordination von Zugangsberechtigungen für Benutzerinnen und Benutzer.²⁸

3.1.2. Dateiformate

Auch bei der Wahl der Dateiformate ist es von Vorteil, wenn das Archiv auf standardisierte Formate zurückgreift. Da es zu einer regelmäßigen Datenmigration kommen muss, um die Informationen der Archivalien zu erhalten, sollten einheitliche

²⁶ BOUNTOURI, Archives in the Digitale Age. Standards, 39.

²⁷ BOUNTOURI, Archives in the Digitale Age. Standards, 39.

²⁸ NEUROTH, OSSWALD, SCHEFFEL STRATHMANN, JEHN (Hg.), Nestor Handbuch, 14f.

Formate für diese Migration gewählt werden. Allerdings gibt es durchaus Bereiche, für welche keine einheitlichen und standardisierten Formate vorliegen.²⁹

3.1.2.1. PDF/A (Portable Document Format/ Archive)

Dieser Standard für archivtaugliche PDF-Versionen entstand 2005 und wird als PDF/A-1 bezeichnet. Dieses Format wird noch in 1a und 1b unterschieden, wobei das Dateiformat PDF/A-1a zusätzlich eine Barrierefreiheit gewährleistet. 2011 wurde eine Erweiterung mit PDF/A-2 eingeführt, welche mit dem Einsatz von Ebenen arbeitet und abwärtskompatibel zu PDF/A-1 ist. Die letzten Ergänzungen sind PDF/A-3 und PDF/A-4, welche die Einbettung beliebiger Dateitypen ermöglichen, allerdings sind diese nicht für die Archivierung geeignet, da nur die Container bei der Überprüfung und Validierung im Schritt des Ingests überprüft werden und nicht die im PDF/A eingebetteten Dateien. Daher sind die beiden letzten Standards des PDF/As nicht für die Anwendung in Archiven geeignet.³⁰

3.1.2.2. JPEG2000 (Joint Experts Photograph Group)

Bei diesem Standard handelt es sich um ein Format zur verlustfreien Komprimierung von Bildern. Grundsätzlich sollte es Nachteile des Formates JPEG ausgleichen. JPEG2000 zählt zu den ISO-Standards (International Organization for Standardization). Durch die verlustfreie Komprimierung der Bilder kann eine hohe Speicherdichte erreicht werden. Auch wenn es ursprünglich nicht als Archivformat vorgesehen war, so weist das Format doch einige Vorteile für den Archivgebrauch auf.³¹

²⁹ ERNST, Standards und Normen im Bereich der Langzeitarchivierung, 68f.

³⁰ ERNST, Standards und Normen im Bereich der Langzeitarchivierung, 69.

³¹ KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/jpeg2000.html> [07.06.2024].

3.1.2.3.ODT, ODP (ODF) (Offene Textdatei, Open Document Präsentation, Open Document Format)

Bei diesem Format handelt es sich um ein Textformat, welches auf dem Office Format und somit auf XML (Extensible Markup Language) beruht. Das Format ist schon seit dem Jahr 2005 ein OAIS-Standard. Im Gegensatz zu DOCX-Dokumenten findet dieser Formattyp keine weite Verbreitung. Zusätzlich weist das Format Mängel im Bereich der Archivierung auf, vor allem im Bereich der Konvertierung.

Beispielsweise können Medienobjekte jeglicher Art einen Informationsverlust bei einer Konvertierung erleiden, da das Format eine geringere Anzahl an Formaten bei der Einbettung in das Dokument zulässt. Zusätzlich werden Digitale Signaturen bei Konvertierungen nicht erhalten. Weiters aktualisiert das Format beim Öffnen der Datei das Feld des Datums und der Zeit, wodurch das Originaldatum des Dokuments überschrieben wird, was aus archivischer Perspektive nicht vorteilhaft ist. Ebenso können bei der Bildung von SIPs Schwierigkeiten auftreten. Ein Vorteil dieses Speicherformates ist allerdings die Tatsache, dass durch die Komprimierung der Daten in einer ZIP (Zipper) – Datei eine hohe Speicherdichte aufgewiesen werden kann.³²

3.1.2.4. TXT (Text)

Laut der Webseite KOST ist das Textformat, welches als reines Textformat gilt, das stabilste Format. Da es sich aber nur um ein Textformat handelt, ist auch das Layout beschränkt. In weiterer Folge bedeutet das, dass eine Migration von anderen Formaten in das Textformat einen Formatierungsverlust verursacht, wodurch ein wichtiger Informationsgehalt des Dokumentes verloren geht. Neben dem Verlust des Layouts können natürlich auch keine anderen Medien, wie etwa Bild oder Video, in

³² KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/odf.html> [07.06.2024].

das Textformat migriert werden. Abgesehen davon erfüllt das Format archivische Anforderungen und wird daher durchaus als Archivformat verwendet.³³

3.1.3. Migration und Emulation

Migration und Emulation sind zwei große Maßnahmen der Datenerhaltung.

Bei der Migration handelt es sich um die Überführung von Daten von einem bestehenden alten Speichermedium auf das neuere Trägerformat, oder genauer gesagt um den „Wechsel auf einen Datenträger mit aktueller Technologie“³⁴. Medien-Migration unterscheidet sich in 4 unterschiedliche Arten, welche alle in Abstufung einen größeren Eingriff in die digitalen Daten bedeuten. Es gibt das Refreshment, die Replication, das Repackaging und die Transformation.

Bei der ersten Variante werden die digitalen Daten auf einen gleichartigen, aber neueren Datenträger gespielt. Bei der Transformation werden hingegen „Änderungen in den gespeicherten Archivpaketen“³⁵ vorgenommen. Bei der Migration besteht immer die Gefahr, dass durch den Vorgang ein Datenverlust entstehen kann. Und je größer und aufwendiger die Migration an sich ist, desto höher ist auch diese Gefahr des Datenverlustes oder der Beschädigung der Daten.³⁶ Neben der Emulation, welche gleich erörtert wird, bleibt die Migration allerdings die einfachere Variante, um einem Datenverlust entgegenzuwirken. Auch wenn man sich hier die Frage stellen kann, ob es sich nach durchgeführter Migration überhaupt noch um das Original des Dokumentes oder der digitalen Daten handelt.³⁷

Der Begriff der Emulation „focuses on the re-creation of the environment in which the digital resource was originally generated.“³⁸ Dabei geht es um die Replikation der Charakteristika der jeweiligen Software und, beziehungsweise oder der Hardware, welche die Funktionalität und das Aussehen der digitalen Ressource wieder

³³ KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/txt.html> [07.06.2024].

³⁴ HEISE, ULLRICH, Digitale Langzeitarchivierung, 104.

³⁵ HEISE, ULLRICH, Digitale Langzeitarchivierung, 104.

³⁶ HEISE, ULLRICH, Digitale Langzeitarchivierung, 104.

³⁷ HEISE, ULLRICH, Digitale Langzeitarchivierung, 106.

³⁸ BOUNTOURI, Archives in the Digitale Age, 43.

darstellen können. Die Methode der Emulation als digitale Präservationsstrategie kann sich daher auf unterschiedliche Aspekte konzentrieren: die Anwendung, das Betriebssystem oder die Hardware. Es können aber auch mehrere oder alle dieser Aspekte in der Emulationsstrategie bearbeitet werden. Auf welche Aspekte sich das Archiv in der Emulation konzentrieren möchte und welche Aspekte in der Umsetzung sinnvoll sind, muss vor der Umsetzung der Strategie überlegt und ausgearbeitet werden.³⁹ Die Methode der Emulation hat den Vorteil, dass es hierbei zu keiner Veränderung an den digitalen Daten selbst kommt. Andererseits ist es, im Gegensatz zur Migration, auch die aufwendigere Methode.⁴⁰

3.2. Vorteile der digitalen Archivierung

Digitale Archivierung kann einige Vorteile aufweisen. Einerseits können auf kleinerem Platz mehr Informationen gespeichert werden. Andererseits kann der Platz für weitere Daten leichter eingerichtet werden, als der Platz für zusätzliche Regalmeter. Das alleine lässt den digitalen Speicher schon als besseren Lösungsansatz für zukünftige Archivierung erscheinen.

Andererseits können digitale Daten leichter und schneller durchsucht werden als analoge Daten. Ein weiterer Vorteil für die Benützung digitaler Daten ist, dass die Bearbeitung der digitalen Daten nicht vor Ort im Archiv stattfinden muss, sondern auch von anderen Orten geschehen kann.⁴¹

Digitale Speicherung muss nicht immer direkt im Archiv stattfinden. Durch die Auslagerung der digitalen Daten an Fremdfirmen, zum Beispiel IT-Unternehmen, Cloud-Services und ähnliches, können die digitalen Daten besser aufbewahrt werden, da die Mitarbeiterinnen und Mitarbeiter dieser Firmen im Umgang mit digitalen Daten meist besser geschult sind.⁴²

³⁹ BOUNTOURI, Archives in the Digitale Age, 43.

⁴⁰ HEISE, ULLRICH, Digitale Langzeitarchivierung, 106.

⁴¹ Jonas KERSCHNER, Aspekte der Archivierung digitaler Geodaten und kartographischer Darstellungsformen (Wien 2013) 14.

⁴² Tobias BEINERT, Armin STRAUBE, Aktuelle Herausforderungen der digitale Langzeitarchivierung, in: Was bleibt?. Nachhaltigkeit der Kultur in der digitalen Welt, hg von Paul KLIMPEL - Jürgen KEIPER (Berlin 2013) 27-46, 35f.

Es ist dennoch immer notwendig, sich der schnellen Veränderung der Zustände im digitalen und technischen Bereich bewusst zu sein, um auf diese Veränderungen auch adäquat eingehen zu können.⁴³

3.3. Beispiele digitaler Archivierung in Österreich

Es sollen hier nicht alle größeren Archive Österreichs und deren Konzept für ihre digitalen Archive vorgestellt werden. Es soll hier lediglich ein kurzer Überblick über die grundsätzliche Lage der digitalen Archivierung in Österreich gegeben werden.

Das digitale Archiv des Österreichischen Staatsarchivs wurde im Jahr 2012 in einer Testversion eingeführt. Das System basiert auf den vorher schon vorgestellten OAIS-Referenzmodell. Das Österreichische Staatsarchiv übernimmt die Dokumente aus der Registratur und konvertiert diese dann in das Archivformat PDF/A. Gemeinsam mit den konvertierten Dokumenten werden auch die Originale und die Metadaten der Dokumente übernommen, ebenso wie die Daten, welche während des Konvertierungsprozesses entstanden sind. Diese werden in einem XML-Schema ins Archiv implementiert. Das digitale Archiv unterliegt einer ständigen Überwachung im Sinne des Preservation Plannings, damit rechtzeitig auf technische Veränderungen eingegangen werden kann. Seit 2018 wird das System an die neuesten technischen Standards angepasst, unter anderem auch die Hardware und Software des digitalen Archives. Für die Nutzung der digitalen Archivalien, welche den Benutzerinnen und Benutzern zur Verwendung offenstehen, soll ein digitaler Lesesaal eingerichtet werden, über den ein Zugriff auf diese Daten möglich sein soll. Unterstützt wird dieser Lesesaal auch durch das Archivinformationssystem des österreichischen Staatsarchives.⁴⁴

Das Niederösterreichische Landesarchiv implementierte sein digitales Archivsystem 2020. In dieses System werden laufend die elektronischen Akten der

⁴³ BEINERT, STRAUBE, Aktuelle Herausforderungen der digitalen Langezeitarchivierung, 38.

⁴⁴ Digitale Archivierung im Österreichischen Staatsarchiv, in: Webseite des Österreichischen Staatsarchiv, <https://www.oesta.gv.at/ueber-uns/digitales-archiv-oesterreich.html#:~:text=Die%20Verantwortung%20f%C3%BCr%20die%20Umsetzung%20der%20digitalen%20Archivierung%20obliegt%20gem%C3%A4%C3%9F> [07.06.2024].

niederösterreichischen Landesverwaltung übernommen. Bei der Archivierung der Daten werden diese automatisch aus dem laufenden System entfernt, damit dieses System entlastet wird und eine bessere Funktion gewährleistet werden kann. Zusätzlich kann durch die Löschung aus dem aktiven System während der Archivierung im digitalen Archiv des Niederösterreichischen Landesarchiv die Datenschutzverordnung besser eingehalten werden.⁴⁵

Auch in den Landesarchiven in den Bundesländern Oberösterreich und Vorarlberg kommt Digitale Archivierung schon zum Einsatz. Dabei wird wie auch in Niederösterreich das Produkt der Firma COSMOS Preservation verwendet.⁴⁶

Insgesamt lässt sich eine starke Heterogenität bei der Ausführung und Umsetzung der digitalen Archivierung in Österreich feststellen. Daher gibt es Versuche, an einheitlicheren Ansätzen und Implementierungen der digitalen Archivierung zu arbeiten, soweit dies die unterschiedlichen Verwaltungsorganisationen der einzelnen Landesarchive zulassen.⁴⁷

3.4. Problemstellungen der digitalen Archivierung

Neben der dauerhaften Zunahme an Datenmengen, welche potentiell von Archiven übernommen werden könnten oder sollten, kommen auch andere Probleme bei der Archivierung digitaler Daten auf. Diese sind unter anderem die Haltbarkeit der Datenträger und das dadurch hervorgerufene Problem der schwindenden Lesbarkeit der gespeicherten Daten oder die großteils fehlenden Konzepte zur Archivierung von Webseiten, Socialmedia-Plattformen und Mailaccounts und deren generelle Flüchtigkeit, da es nicht gesichert ist, dass diese Daten auch dauerhaft unter der gleichen Adresse aufrufbar bleiben.⁴⁸

⁴⁵ JP, NÖ Landesarchiv wird zum Digitalen Gedächtnis des Landes. Digitalisierung des Archivs ist Pionierprojekt in Österreich, in: Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, <https://voeb-b.at/noe-landesarchiv-wird-zum-digitalen-gedaechtnis-des-landes/#:~:text=Durch%20das%20digitale%20Archiv%20k%C3%B6nnen%20so%20elektronische%20Akten%20aus%20dem> [07.06.2024].

⁴⁶ Markus SCHMIDGALL, Zum Stand der digitalen Archivierung in den österreichischen Bundesländern. *Scrinium* 73, Zeitschrift des Verbandes der Österreichischen Archivarinnen und Archivare (2019), 86-89, 87.

⁴⁷ SCHMIDGALL, Zum Stand der digitalen Archivierung in den österreichischen Bundesländern, 87ff.

⁴⁸ ALTENHÖNER, BRANTL, CEYNOWA, Digitale Langzeitarchivierung in Deutschland, 184.

Zugrunde liegt diesen Problemen die Tatsache, dass es Menschen nicht möglich ist, digitale Daten ohne einem passenden System zur Darstellung der Daten lesen zu können. Im Gegensatz dazu, kann der Inhalt analoger Daten ohne Probleme wahrgenommen werden.⁴⁹

Vor allem um der ersten Problematik zu entkommen, müssen die Daten einer regelmäßigen Datenmigration unterzogen werden. Hierbei sollen die Daten von dem veralteten Trägermaterial oder dem veralteten Dateiformaten auf die neuere Version gespeichert werden.⁵⁰ Ein weiteres Problem, welches bei der digitalen Archivierung aufkommt, ist die Produktion von Elektroschrott. Dieser entsteht als Nebenproduktion der Datenmigration. Darunter fallen Magnetbänder, Festplatten, aber auch Server. Dieser Aspekt ist allerdings in der Literatur kaum bis gar nicht angeführt, wenn es um Nachteile der digitalen Archivierung geht.

3.4.1. Kosten

Das digitale Archiv ist bezüglich größerer Datenmengen einfacher zu handhaben als ein analoges Archiv, welches sich neben anderem um den Platz für mehrere Regalmeter kümmern müsste. Aber auch wenn der reine Speicher an sich günstiger ist, so entstehen für das digitale Archiv schon im Bereich der Übernahme der Daten höhere Kosten. Daher, „Digitale Archivierung ist nicht per se billiger als analoge“⁵¹

Aber nicht nur der reine Speicher an sich kostet Geld, sondern auch das Management dahinter und die Sicherung der Daten ist Teil des Kostenfaktors. Auch die Anschaffung neuer Möglichkeiten für den Zugriff auf die Daten muss regelmäßig überlegt und in die Kostenkalkulation eingerechnet werden. Diese Faktoren bedeuten im Gesamten größere Kostenfaktoren als bei der analogen Archivierung.⁵²

Gerade im Bereich der ständigen Datenmigration kommt es im digitalen Archiv laufend zu neuen Kosten.

⁴⁹ KERSCHNER, Aspekte der Archivierung digitaler Geodaten und kartographischer Darstellungsformen, 14.

⁵⁰ BEINERT, STRAUBE, Aktuelle Herausforderungen der digitale Langzeitarchivierung, 35.

⁵¹ BEINERT, STRAUBE, Aktuelle Herausforderungen der digitalen Langzeitarchivierung, 29f.

⁵² BEINERT, STRAUBE, Aktuelle Herausforderungen der digitalen Langzeitarchivierung, 30.

3.4.2. CO2-Verbrauch des digitalen Speichers

Digitalisierung wird immer wieder in Verbindung mit Nachhaltigkeit erwähnt, denn es werden weniger offensichtliche Ressourcen verwendet. Es braucht weniger Papier, es kann eine wesentlich größere Menge an Daten auf kleinerem Platz gespeichert werden und die Kommunikation wird wesentlich erleichtert.

Dennoch ist ein digitaler Speicher nicht unbedingt nachhaltiger als ein analoger Speicher. Denn auch hier werden Ressourcen bei der Produktion und auch bei der Erhaltung des digitalen Speichers verbraucht. Daher kann auch der weltweite digitale Speicher als hoher Energieverbraucher eingestuft werden. Dies bedeutet auch, dass durch den digitalen Speicher eine große Menge an Treibhausgasemissionen produziert wird.

Durch den stetigen Zuwachs des digitalen Speichers steigt auch jährlich dessen Energieverbrauch.⁵³

3.5. Fazit

Aufgrund der stetigen Entwicklungen und Verwendung digitaler Daten und Akten im Verwaltungsbereich, auf Social Media und Webseiten, ist eine Archivierung ebendieser Daten erforderlich. Diese müssen als Teil des historischen Gedächtnisses und laut den Archivgesetzen ins Archiv überführt werden und dort aufbewahrt und schließlich zugänglich gemacht werden.

Die Entwicklungen der digitalen Archive in Österreich sind noch an keinen Schlusspunkt gelangt und werden es aufgrund der Beschaffung der digitalen Daten auch nie sein. Denn um ein erfolgreiches und funktionierendes digitales Archiv

⁵³ Nicole FEGGER, Digitaler CO2-Fußabdruck. Digitalisierung und Nachhaltigkeit, in: Overlook, <https://overlook.one/blog/digitaler-CO2-Fu%C3%9Fabdruck> [08.06.2024].

aufzubauen und zu erhalten, muss sich dieses auch rechtzeitig und jederzeit an Änderungen in der digitalen Welt anpassen können, damit die Daten auch weiterhin verwendbar bleiben.

Dennoch stellt sich die Frage, ob ein digitales Archiv tatsächlich der Endpunkt der archivischen Entwicklung sein wird, oder ob es nicht doch andere Möglichkeiten der Archivierung gibt, welche für digitale Daten auf lange Sicht besser geeignet sein werden. Diesen Denkanstoß möchte ich als Überleitung zum nächsten Abschnitt meiner Arbeit verwenden.

Ich möchte anschließend auf andere Möglichkeiten der Speicherung digitaler Daten eingehen und mich mit deren Vor- und Nachteilen auseinandersetzen.

4. Optical Storages

Bei dem Namen Optical Storages handelt es sich um einen vorläufigen Namen von Speichermethoden, welche zu den "Next Generation Archival Big Data Storage" zählen. Diese Speichermethoden arbeiten mit dem WORM-Prinzip. Ausgeschrieben bedeutet die Abkürzung „Write Once, Read Forever“.

Hauptsächlich sollen diese Speichermethoden jenen Problematiken der digitalen Archivierung entgegenwirken, welche ich am Anfang meiner Arbeit in Kapitel 3 angesprochen haben.

Es gibt mittlerweile unterschiedliche Methoden, mit denen Optical Storages arbeiten können. Allen Projekten liegt zugrunde, dass es mit der wachsenden Datenmenge eine neue Art der langfristigen Datenspeicherung braucht, um einem Datenverlust entgegenzuwirken, beziehungsweise um eine erleichterte Arbeitsweise zu schaffen.

Genauer gesagt, allen Methoden ist gemein, dass es bei der Speicherung auf ihren Datenträgern keinen Bit-Rot, also keinen Bit-Verlust über die Zeit gibt. Die Daten, welche auf den jeweiligen Datenträgern gespeichert werden, bleiben auch genau in dem gespeicherten Zustand erhalten, ohne dass es eine regelmäßige Migration der Daten braucht, um den Bit-Verlust entgegenzuwirken. Um die Daten digital zu erhalten, müssen mit den heutigen Datenträgern regelmäßige Datenmigrationen durchgeführt werden, was einen zusätzlichen Aufwand darstellt. Findet diese Migration nicht regelmäßig statt, so kann es zu einem Verlust der Daten kommen. Die Verwendung von Optical Storages versucht hier eine einfachere Arbeitsweise zu etablieren.

Ein weiterer wichtiger Punkt, welcher für die Optical Storages spricht, ist die lange Haltbarkeit der Datenträger. Die einzelnen Projekte haben Datenträger mit unterschiedlicher Haltbarkeit, allerdings halten sie alle zumindest länger als 100 Jahre ohne zusätzliche Bearbeitung, also wesentlich länger als die heutigen technischen Datenträger. Aus der langen Haltbarkeit der Datenträger ergeben sich auch die günstigeren Kosten der Projekte über einen langen Zeitraum. Die Projekte brauchen zwar noch einen höheren Kostenaufwand bei der Herstellung der Datenträger, allerdings rentieren sich die Kosten im Laufe der Zeit.⁵⁴

⁵⁴ Martin STÜRZLINGER; Antrittsvorlesung FH Potsdam, 18.12.2023.

Ich möchte in diesem Kapitel die unterschiedlichen Ansätze kurz vorstellen und deren Funktionsweisen kurz vorstellen. Im Unterkapitel 4.7. habe ich bildliche Darstellungen der jeweiligen Optical Storages eingefügt, um ein besseres Verständnis der Speichermethoden geben zu können.

4.1. Archium

Das Projekt "Archium" arbeitet mit der Analogisierung digitaler Daten, um eine langfristige Speicherung der digitalen Daten zu ermöglichen. Das Prinzip beruht auf der Überlegung, dass analoge Archivierung auf Papier bereits erprobt ist und die Erhaltung von Daten auf Papier gut machbar ist. Diese Art der Archivierung erfüllt alle Vorteile der analogen Archivierung. Zum Beispiel wird kein zusätzlicher Strom, kein zusätzliches Personal für die Wartung, kein Internet und keine zusätzliche Technik vonnöten sein.

Die Daten werden auf ISO-genormten Papier gedruckt, platzsparend und klein, damit das Platzproblem nicht zum Tragen kommt. Ebenso werden die Metadaten der Archivalien mit einem QR-Code dazu gedruckt. Archium wirbt vor allem damit, dass ihre Methode vor Angriffen aus dem Netz sicher ist. Die Datenspeicherung ist von anderen Firmen, von Strom und ähnlichem unabhängig. Einen weiteren Vorteil, welchen Archium für ihre Methode darlegen, ist das Wegfallen von vielen Erhaltungskosten, die sich bei der Speicherung digitaler Daten auf technischen Datenträgern automatisch ergeben.⁵⁵

Archium wirbt auf ihrer Webseite mit einer Geschichte. Kinder fanden vor Jahren am Dachboden eine Kiste mit alten Fotos und konnten diese einfach anschauen. Kinder heutzutage finden alte Disketten. Allerdings können diese nicht mehr angeschaut werden, weil der dazugehörige Computer nicht mehr funktioniert. Und in Zukunft werden überhaupt keine alten Informationen mehr auf Dachböden gefunden werden, weil alle digitalen Daten in Cloudservices gespeichert und aufbewahrt werden und nicht garantiert werden kann, dass diese Daten noch erhalten bleiben.⁵⁶

⁵⁵ Microsheets und Mikrofilm, in: Archium, <https://www.archium.org/mikrofilm-microsheets.html> [19.08.2024].

⁵⁶ Über Uns, in: Archium, <https://www.archium.org/digital-definition.html> [19.08.2024].

Archium setzt daher auf analoge Daten. Einerseits können analoge Datenträger meist ohne zusätzliche Kenntnisse gelesen und interpretiert werden. Andererseits kann einem Datenverlust meist rechtzeitig entgegengewirkt werden, weil dieser langsamer vonstattengeht als bei digitalen Daten und es fällt der Faktor der regelmäßigen Datenmigration weg. Dadurch sind die Lagerung und die Speicherung von analogen Daten auch in diesem Aspekt günstiger als das Speichern digitaler Daten.

Als letzten Punkt führt Archium noch den Punkt der Authentizität an. Denn ihrer Meinung nach ist das Vertrauen in analoge Daten höher als das Vertrauen in digitale Daten. Und das werde, laut Archium, auch in Zukunft der Fall sein.⁵⁷

Gelagert werden die analogen Akten und Datenträger von Archium in einem Stollen in der Nähe von Thüringen. Archium verspricht seinen Kundinnen und Kunden, die von ihnen archivierten Daten vor Umweltkatastrophen und eventuellen Angriffen zu schützen.⁵⁸

Archium bringt viele Argumente, weswegen ihre Art der Archivierung digitaler Daten eine zukunftsorientierte Lösung ist. Die Daten werden auf einem kleineren Platz aufbewahrt und bleiben dennoch mit einer Lupe lesbar. Die dazugehörigen Metadaten werden dabei mit einem QR-Code daneben gespeichert. Diese sind dann auch vor Cyberangriffen und Datenverlusten im Netz geschützt. Weiters sind diese analogen Daten auch nachhaltiger als Daten auf digitalen Datenträgern, da für die Speicherung auf Papier weder seltene Erden gebraucht werden noch Sondermüll produziert wird.⁵⁹ Archium bietet aber auch Archivsoftwares und Archivdatenbanken zur Archivierung der digitalen Daten an, welche als Back-Up analog auf Papier archiviert werden sollen. Die Software basiert auf dem ISAD (G) (International Standard Archival Description (General)).⁶⁰

All diese Punkte sprechen für die Archivierungsmethode von Archium. Allerdings gibt es auch Punkte, welche gegen diese Art der Archivierung digitaler Daten spricht. Einerseits muss auch Papier unter den richtigen Umständen archiviert und gelagert werden, wofür klimatische Bedingungen zur besseren Erhaltung erzeugt werden müssen. Außerdem stellt sich die Frage, ob der QR-Code auch in Zukunft immer noch angewendet werden kann und ob die darin gespeicherten Metadaten in ein

⁵⁷ Was ist Analogisierung?, in: Archium, <https://www.archium.org/analogisierung.html#sec-3ac7> [19.08.2024].

⁵⁸ Archivdienstleistungen, in: Archium, <https://www.archium.org/archivdienstleister.html> [19.08.2024].

⁵⁹ Krisensicheres Backup, in: Archium, <https://www.archium.org/mikrofilm-microsheets.html> [19.08.2024].

⁶⁰ Archivdatenbank, in: Archium, <https://www.archium.org/archivdatenbank.html> [19.08.2024].

paar Jahren noch abgerufen werden können. Und auch der Zusatz der Lupe, welche zum Lesen verwendet werden muss, erschwert die Handhabung der mit Archium gespeicherten Daten. Denn in digitaler Form können Dokumente leichter und größer dargestellt werden.

Beim Thema Kostenfaktor wirbt Archium damit, dass nur die Archivierung auf Papier etwas kostet. Die bei digitaler Archivierung entstehenden laufenden Kosten bezüglich der Datenmigration fallen hier also weg.⁶¹

Allerdings muss auch Papier eventuell restauriert und gereinigt werden. Das heißt, auch wenn die Kosten nicht so regelmäßig anfallen und vom Umfang nicht so umfassend sind, so können dennoch auch bei Archium noch Kosten während der Lagerung und der Erhaltung der Daten entstehen.

4.2. MOM-Memory of Mankind

Das Projekt "Memory of Mankind" setzt auf das Prinzip, dass Daten auf Microfilmpplatten aus Keramik innerhalb des Salzbergwerkes in Hallstatt aufbewahrt werden sollen. Das Projekt startete 2012 mit dem Ziel, ein Abbild der jetzigen Gesellschaft zu schaffen und an einem Ort zu lagern, welcher die Haltbarkeit gewährleistet. Vorerst möchte das Projekt 1000 Bücher, welche als archivwürdig und bezeichnend für die Gesellschaft sind, auf den vorher genannten Platten speichern, welche im Salzbergwerk gelagert werden sollen. Die Bücher können von jedem und jeder auf der Homepage des Projektes vorgeschlagen werden, damit ein umfassendes Bild wiedergegeben werden kann.⁶²

Der treibende Gedanke hinter MOM war die Überlegung, dass die technischen Speichermedien sehr fragil sind im Vergleich zu anderen Datenträgern. Man spricht auf der Homepage neben den fragilen Speichermedien auch von einem Digital Dark Age, da sie die Frage aufwerfen, ob die digitalen Daten auch noch in 100 oder mehr Jahren aufrufbar oder verwendbar sein werden, wenn wir bei den heutigen Speichermethoden bleiben. Wenn dem nicht der Fall sein sollte, so würden wir

⁶¹Über Uns, in: Archium, <https://www.archium.org/digital-definition.html> [19.08.2024].

⁶²A Thousand of Books for a Million Years, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/1000books/#:~:text=A%20Thousand%20Books%20for%20a%20Million%20Years.%20Das%20Memory%20of> [24.01.2024].

riskieren, dass eine große Menge an Daten aufgrund dieser Datenspeicherung verloren gehen.

MOM arbeitet mit einem Datenträger, welcher die bis jetzt ältesten gefundenen Daten überliefert. Diese befinden sich auf Tontafeln, daher arbeitet auch MOM mit keramischen Mikrofilm Tafeln. Laut eigenen Angaben des Projektes können auf einer dieser Tafeln bis zu 1000 Seiten eines Buches gespeichert werden. Die Tafeln haben eine Größe von 20x20cm und sind widerstandsfähig gegenüber Wasser, Säuren und hohen Temperaturen von bis zu 1.500C°. ⁶³

Die Keramiktafeln von Memory of Mankind sollen in Salzbergwerken gelagert werden. Sie werden von MOM als „unverwüstliche Datenträger“⁶⁴ bezeichnet. Die Daten sollen direkt lesbar und interpretierbar sein. Die Texte werden, wie auch bei Archium, minimalisiert auf den Keramiktafeln aufgetragen und können mit Hilfsmitteln gelesen werden. Bei Bildern ist die Kapazität der Tafeln durch die hohe Auflösung beschränkt. ⁶⁵

Die Vorteile der Methode sind ebenfalls ähnlich, wie bei Archium. Der digitale Aspekt der Archivierung fällt weg und durch die Verwendung von Keramiktafeln entgeht man der regelmäßigen Datenmigration. Und auch dem Datenverlust wird durch die Verwendung analoger Datenträger entgegengewirkt.

Allerdings kann man auch hier ein oder zwei Punkte gegen diese Methode vorbringen. Die Lagerung in einem Salzbergwerk erscheint auf den ersten Blick praktisch bezüglich der klimatischen Verhältnisse. Allerdings wird das Archiv nicht als tagtäglich benutztes Archiv aus- und angelegt, denn der Lagerort erleichtert den Zugang zu den Daten nicht. MOM möchte die Daten für einen wesentlich späteren Zeitpunkt gespeichert wissen⁶⁶, wodurch sich diese Methode als nicht sehr vorteilhaft für ein Archiv erweist, welches regelmäßig benutzt wird und auch benutzt werden soll, in dem täglich neue Daten hinzukommen und auch für Forschungszwecke und Ähnlichem verfügbar sein soll.

⁶³ A Thousand of Books for a Million Years, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/1000books/#::~:~:text=A%20Thousand%20Books%20for%20a%20Million%20Years.%20Das%20Memory%20of> [24.01.2024].

⁶⁴ Home, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/> [22.05.2024].

⁶⁵ Wie bewahrt man Informationen dauerhaft für 1 Million Jahre auf?, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/how-is-information-kept-legible-for-1-million-years/>, [22.05.2024].

⁶⁶ Wer soll MOM finden?, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/who-is-supposed-to-find-mom/>, [22.05.2024].

4.3. WORF (Write Once, Read Forever)

WORF arbeitet mit dem in der Einleitung schon genannten Namen "Write Once, Read Forever".

Dieses Projekt arbeitet mit Platten aus Silberhalogenid, welche laut dem Projekt unter normalen, klimatischen Umständen mehr als ein Jahrhundert halten sollen. Damit braucht dieses Projekt nur mehr für das Lesen der gespeicherten Daten Strom, um diese wieder zugänglich machen zu können.

Geschrieben werden die Daten als stehende Wellen von Licht im Nanobereich im jeweiligen Farbbereich der Wellen auf den photosensitiven Platten aus Silverhalogenid. Die erzeugten Muster der stehenden Wellen sind spezifisch einer bestimmten Wellenlänge zuzuordnen.⁶⁷ Durch das Schreiben der Daten in stehenden Wellen können diese auch nicht, wie die Farben moderner Farbfilme, verblassen.⁶⁸ Weitere Vorteile dieser Speichermethode liegen darin, dass der Datenträger bei der Lagerung einfach handzuhaben ist, keine aufwendigen, klimatischen Gegebenheiten braucht und zusätzlich auch etlichen Widrigkeiten des Weltraumes standhalten kann. Daher wird diese Methode auch darauf getestet, bei längeren Aufenthalten im Weltraum für Missionen als Datenspeicher in Verwendung zu kommen.⁶⁹

In der Literatur finden sich keine genauen Angaben bezüglich der Kosten, welche während des Prozesses entstehen.

4.4. Microsoft Silica

Das Projekt von Microsoft Silica arbeitet im Gegensatz zu WORF mit Quarzglas, welches ebenfalls eine lange Lebensdauer vorweisen kann. Laut der Webseite von Projekt Silica soll dieser Datenträger 10.000 bis 100.000 Jahre halten. Es soll, wie die anderen Methoden auch, einmal geschrieben und öfter gelesen werden können und andererseits der Datenmigration, welche auf magnetischen Datenträgern notwendig sind, um keine Daten zu verlieren, entgegenwirken.

⁶⁷ WORF, in: Creative Technology, <http://www.creative-technology.net/CTECH/WORF.html>, [24.01.2024].

⁶⁸ Richard J. SOLOMON, Eric ROSENTHAL, Pedro G.C. DE OLIVEIRA et al., WORF (Write Once, Read Forever). Next Generation Archival Big Data Storage, *IEEE Aerospace Conference* (2021)3.

⁶⁹ SOLOMON, ROSENTHAL, DE OLIVEIRA et al., WORF, 3.

Die Daten werden mit einem Laser auf die Platten aus Quarzglas geschrieben und können mit einem Mikroskop, speziell mit einem polarisationsempfindlichen Mikroskop, wieder gelesen werden. Dadurch garantiert Microsoft, dass es während dem Lesen zu keiner unabsichtlichen Überschreibung der Daten kommen kann.⁷⁰ Mit dieser Methode der Speicherung schafft es Microsoft eine hohe Dichte an Daten auf kleinstem Raum zu speichern. Genauer gesagt werden laut Microsoft 7 Terabyte auf der Größe einer DVD gespeichert.⁷¹

Das Project Silica arbeitet mit Glas, weil es sich hierbei um eine günstige Methode handelt, welche gleichzeitig aufgrund der chemischen Zusammensetzung auch lange haltbar ist. Weiters fallen die auf dem Glas gespeicherten Daten nicht elektromagnetischen Wellen zum Opfer.⁷² Neben der Speicherdichte sind auch die Erhaltungskosten der Glasplatten recht gering, sobald diese einmal gelagert werden.⁷³

Die Speicherung auf Glas ist immer noch in der Entwicklung, vor der kommerziellen Nutzung muss es laut Expert*innen noch drei bis vier Entwicklungsstufen durchlaufen.⁷⁴

Es stellt sich hierbei dennoch die Frage, ob es sinnvoll ist, die Zukunft der digitalen Archivierung in Händen eines einzelnen Konzerns zu lassen und ob es hierbei nicht in Zukunft zu Problemen kommen könnte, wenn es um Platz und/ oder Lizenzen geht.

4.5. DNA Data Storage

Auch bei DNA Data Storage handelt es sich um eine Speicherung der Daten nach dem WORM-Prinzip. Die einmal auf der DNA gespeicherten Daten können ohne

⁷⁰ Project Silica Overview, in: Microsoft, <https://www.microsoft.com/en-us/research/project/project-silica/>, [24.01.2024].

⁷¹ Project Silica Overview, in: Microsoft, <https://www.microsoft.com/en-us/research/project/project-silica/>, [24.05.2024].

⁷² Research, in: Microsoft, <https://www.microsoft.com/en-us/research/video/research-talk-storing-data-for-millennia/>, [24.05.2024].

⁷³ Sealed in Glas, in: Unlocked Microsoft, <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.,> [24.05.2024].

⁷⁴ Sealed in Glas, in: Unlocked Microsoft, <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.,> [24.05.2024].

Datenmigration lange Zeit aufbewahrt werden. Zusätzlich hat DNA eine lange Haltbarkeit, als Speichermedium ist es das bis heute am längsten haltbare Material. Zusätzlich handelt es sich im Vergleich zu allen anderen Datenträgern bei DNA Data Storage um die Methode, welche am platzsparendsten ist.⁷⁵

4.6. Fazit

Die Kosten der Optical Storages halten sich besonders nach der Anschaffung in Grenzen, da es bei keiner der Methode notwendig ist, eine regelmäßige Datenmigration durchzuführen. Ebenso werden die zu lagernden Daten auf einem kleineren Platz gespeichert als herkömmliche Datenträger, wie Magnetbänder, CDs, DVDs und Ähnliches. Die Träger des Speichermaterials sind, wie bei MOM, bei Microsoft Silica und DNA, von besonders haltbarem Material und können auch widrigere klimatische Bedingungen und Umstände aushalten. Daher sprechen einige Punkte für die Verwendung von Optical Storages als zukünftige Lösung für die digitale Archivierung.

Aber auch wenn die Optical Storages als Zukunft der digitalen Archivierung gehandhabt werden und diese vielleicht auch, in welcher Form auch immer, werden könnten, dürfen die Nachteile der jeweiligen Speichermethode nicht außer Acht gelassen werden. Auch wenn die Daten platzsparender gelagert werden können, so sind nicht alle der vorgestellten Methoden sinnvoll für einen Archivalltag. So zum Beispiel ist die Lagerung von Daten in einem Salzbergwerk maximal als Back-Up Speicher umsetzbar. Auch die Archivierung auf Papier, wenn auch in minimalisierter Form, braucht im Vergleich zu anderen Speichermethoden mehr Platz.

Die Überlegungen der Optical Storages öffnen jedoch die Möglichkeit für eine interessante Diskussion, wie digitale Archivierung in Zukunft gestaltet werden könnte und bietet Ansätze, wie den Problemen derselben entgegengewirkt werden könnte.

⁷⁵ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 1f.

In weiterer Folge dieser Arbeit werde ich näher auf die DNA Datenspeicherung eingehen, sowie die Vorteile, aber auch Nachteile dieser Speichermethode erläutern.

4.7. Abbildungen der vorgestellten Optical Storages



Abbildung 2, Archivierung auf Papier.⁷⁶

⁷⁶ Archivierung auf Papier, in: Archium, <https://archium.org/mikrofilm-microsheets.html###>, [11.12.2024].

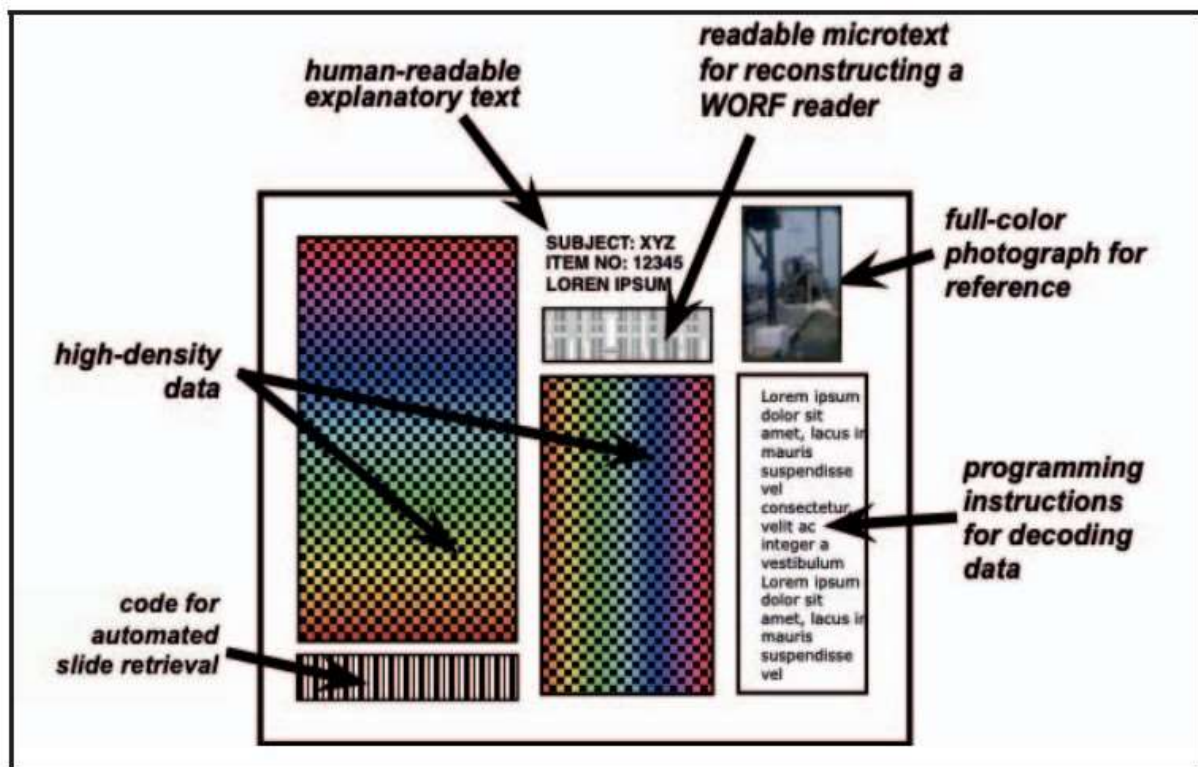


Abbildung 3, Ein Beispiel von WORF Medien, welche aus mehrfach lesbaren Informationen bestehen.⁷⁷



Abbildung 4, Ceramic data carrier.⁷⁸

⁷⁷ SOLOMON, ROSENTHAL, DE OLIVEIRA et al., WORF, IEEE Aerospace Conference (2021) 3.

⁷⁸ Ceramci data carrier, in: Memory of Mankind. <https://www.memory-of-mankind.com/subwebsite/> [11.12.2024].



Abbildung 5, Schritt 1 und 2 des Projekt Silicas.⁷⁹

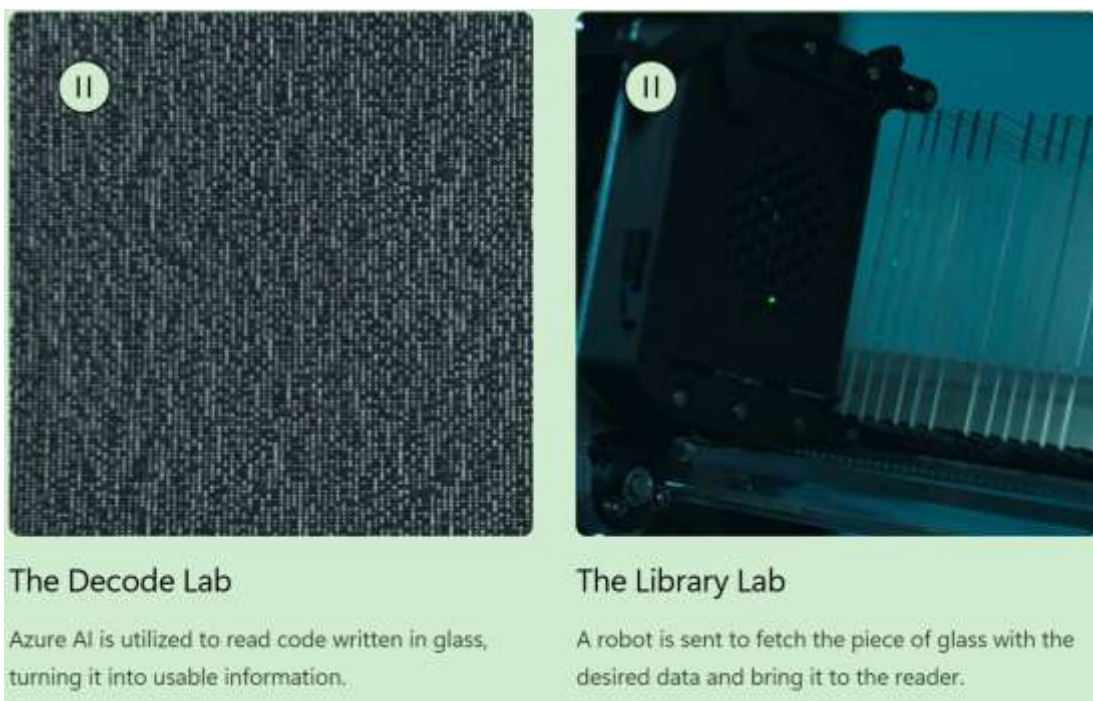


Abbildung 6, Schritt 3 und 4 von Project Silica.⁸⁰

⁷⁹ Sealed in Glas, in: Unlocked Microsoft. <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.> [11.12.2024].

⁸⁰ Sealed in Glas, in: Unlocked Microsoft. <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.> [11.12.2024].

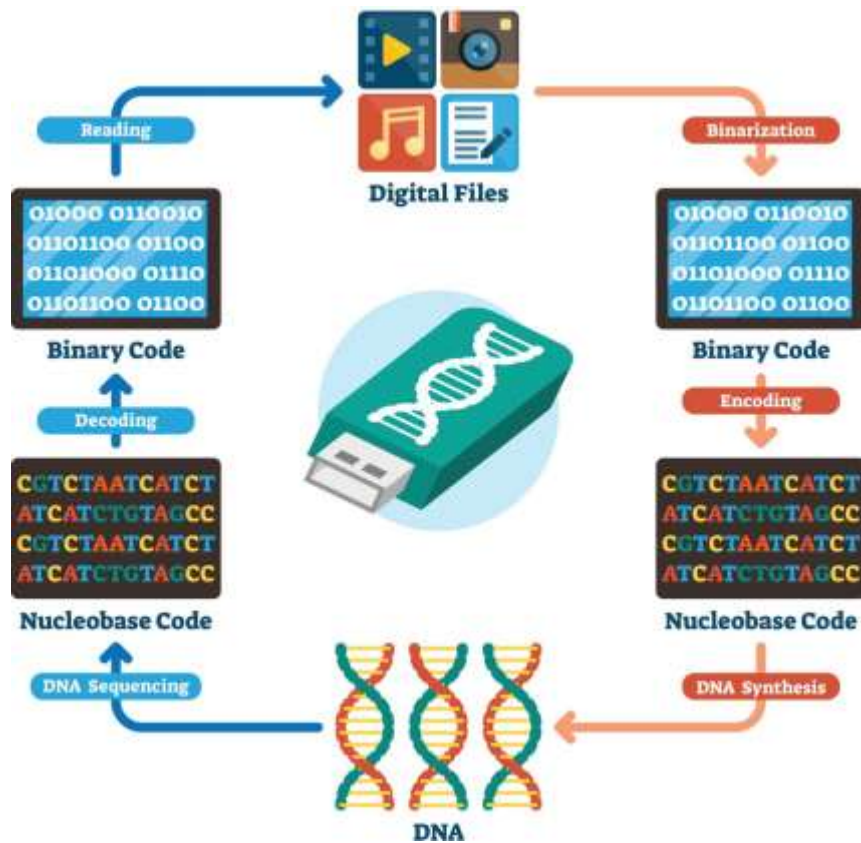


Abbildung 7, Darstellung des Ablaufes der Schaffung eines DNA Data Storage.⁸¹

⁸¹ KIM, PSCHETZ, LINEHAN (Hg.), Archives in DNA, 2.

5. DNA Data Storage

Die nachfolgenden Kapitel beschäftigen sich genauer mit dem DNA Data Storage, den technischen Hintergründen der Herstellung der synthetischen DNA und der Codierung und Speicherung von digitalen Daten auf den DNA-Strängen. Weiters werden die unterschiedlichen Ansätze, mit denen Daten auf DNA gespeichert werden können und die eventuelle praktische Anwendung der Methode aufgezeigt. In dieser Arbeit wird kritisch hinterfragt, ob und wie diese Methode tatsächlich im Archiv angewandt werden kann. Ebenso werden mögliche Nachteile bei all den Vorteilen dieser Methode aufgezeigt.

5.1. Entwicklung der DNA Datenspeicherung

Im Jahr 1953 wurde die Struktur der DNA als Informationsträger des genetischen Materials von James Watson und Francis Crick in einem Artikel veröffentlicht⁸². Seitdem gab es unterschiedliche Konzepte, mit denen die Speicherung von Daten auf DNA vorgeschlagen wurde.⁸³

Erste Versuche der Synthetisierung von DNA gab es schon in den 1970er Jahren. Die tatsächliche Entschlüsselung der DNA gelang dann Ende der 1990er, wodurch die Grundlage der DNA Speicherung gelegt wurde.⁸⁴ Diese ersten Versuche starteten damit, dass ungefähr 20 Nukleotide synthetisiert wurden.⁸⁵

Seit den 1990er Jahren hat sich die Speicherung auf DNA rasant weiterentwickelt. Während im Jahr 1999 das Speichern und Auslesen einer Nachricht mit 23 Zeichen der Standard war, so konnte man im Jahr 2013 dasselbe mit einem Datensatz von 739 KB durchführen.⁸⁶

⁸² Edward EDELSON, Francis Crick and James Watson and the building blocks of life (New York 1998) 9.

⁸³ DONG, SUN, PING et al., DNA storage, 1092.

⁸⁴ Leander, BEIL, Gigantische Datenmengen auf DNA, in: ARD Alpha, <https://www.ardalpha.de/wissen/speichermedien-der-zukunft-dna-speicher-daten-100.html> [22.11.2024].

⁸⁵ BUKO, TUCZKO, ISHIKAWA, DNA Data Storage, 7.

⁸⁶ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 1f.

5.2. Vorteile der Methode

Der erste Vorteil ist die Tatsache, dass DNA die Möglichkeit bietet, große Mengen an Daten auf kleinstem Raum zu speichern. Die ungefähre Grenze der Daten, welche auf 1mm³ an DNA gespeichert werden kann, liegt laut Bornholt et al. bei einem Exabyte. Dies entspricht 109 Gigabyte pro Kubikmillimeter.⁸⁷ Es finden sich allerdings unterschiedliche Angaben zur tatsächlichen Speicherkapazität von DNA. Zum Beispiel wird die Speicherkapazität von DNA im Paper von Dong et al. bei 103 Gigabyte pro Kubikmillimeter angegeben.⁸⁸

Ein weiterer Vorteil des Speichermediums DNA ist die Tatsache, dass DNA als solches eine hohe Stabilität hat, was in weiterer Folge auch dazu führt, dass sie eine lange Haltbarkeit aufweist. Diese wird in der Literatur mit Jahrtausenden angegeben. Neben der langen Haltbarkeit kann DNA auch bei widrigen Umständen der klimatischen Bedingungen weiterhin intakt bleiben.⁸⁹

Aufgrund dieser Vorteile des Datenträgers DNA ergeben sich auch geringe Kosten bezüglich des Erhalts der DNA sobald die digitalen Daten darauf gespeichert sind⁹⁰, wodurch die weiteren Erhaltungskosten, welche bei digitalen Archiven entstehen, minimiert werden.

Bezüglich des Speicherplatzes gibt es ebenfalls unterschiedliche Angaben, wie viel Platz die Speicherung der weltweiten Daten auf DNA benötigen würden. Beispielsweise wird vermutet, dass die gesamten digitalen Daten weltweit auf einem Kilo DNA gespeichert werden könnten⁹¹. Andere wiederum vermuten, dass die gesamte Information der Firma Dropbox auf nur 500 Gramm DNA aufbewahrt werden könnte.⁹²

⁸⁷ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 1f.

⁸⁸ Yiming w, Fajia SUN, Zhi PING et al., DNA storage. Research landscape and future prospects. *National Science Review* 7 (2020) 1092-1107, 1092.

⁸⁹ HECKEL, An archive written in DNA, 236.

⁹⁰ Raphael KIM, Larissa PSCHETZ, Conor LINEHAN et al., Archives in DNA. Workshop Exploring Implications of an Emerging Bio-Digital Technology through Design Fiction. *Academic Mindtrek*. Proceedings of the 24th International Academic Mindtrek Conference (2021) 102-105, 102.

⁹¹ KIM, PSCHETZ, LINEHAN (Hg.), Archives in DNA, 102.

⁹² Wolfgang STIELER, Petabyte pro Gramm, in: Heise Online, <https://www.heise.de/hintergrund/Petabyte-pro-Gramm-3664504.html#:~:text=Dr.%20Wolfgang%20Stieler.%20DNA-Molek%C3%BCle%20sind%20gewisserma%C3%9Fen%20Festplatten%20%E2%80%93%20nur%20eben> [26.01.2024].

Die oben beschriebenen Fakten seiner kostengünstigen und effizienten Speichermöglichkeit sprechen für die Implementierung einer DNA basierten Speichermethode in den Archiven.

5.3. Technische Hintergründe

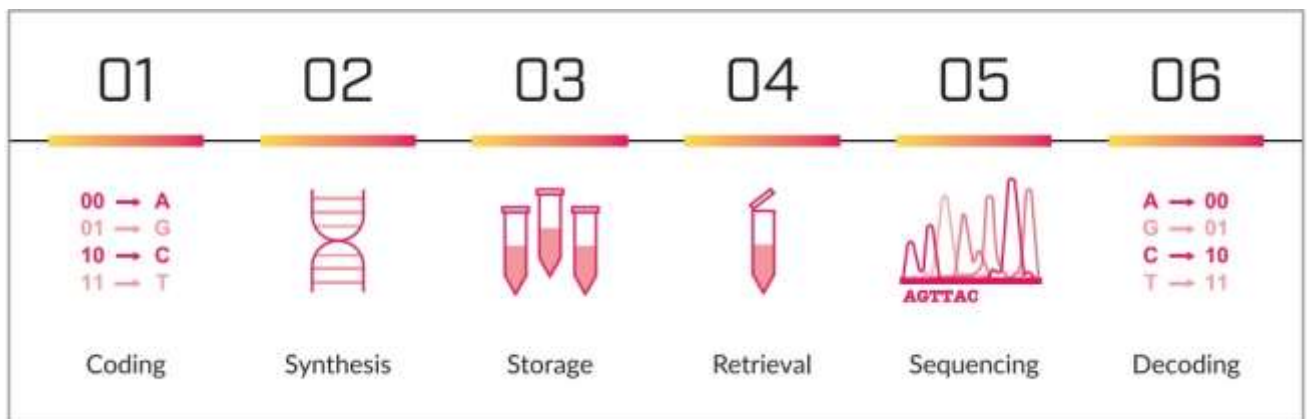


Abbildung 8, Ablauf des DNA Data Storages.⁹³

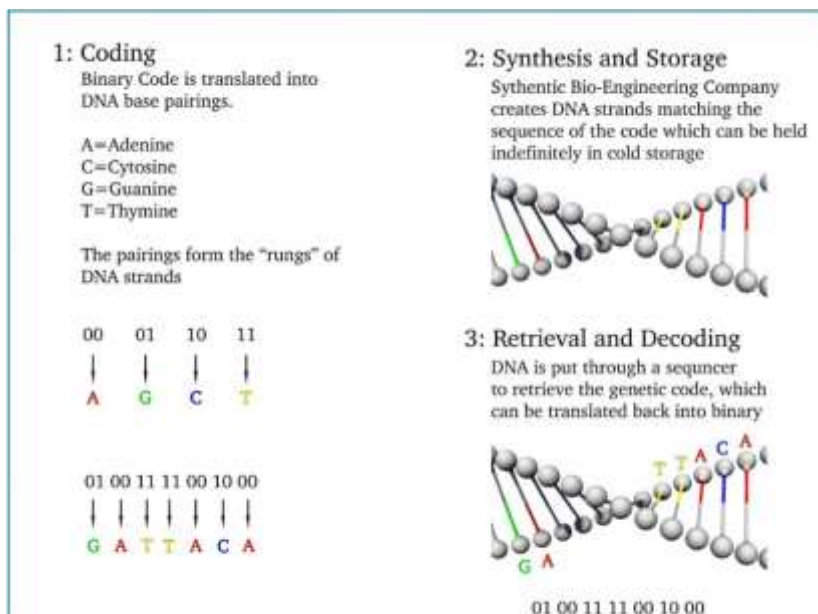


Abbildung 9, Ablauf des DNA Data Storages.⁹⁴

⁹³ DNA Data Storage Alliance, Preserving our digital legacy. An Introduction to DNA Data Storage (2021) 15.

⁹⁴ Potomac Institute für Policy Studies, The Future of DNA Data Storage, 9.

Die meisten Forschungsprojekte arbeiten mit der Methode der Codierung des Binären Codes auf die einzelnen Basen der DNA und der anschließenden Sequenzierung⁹⁵ und Synthetisierung⁹⁶ der DNA, welche dann den binären Code darstellt und als solcher gespeichert und gelagert werden kann. Dieser Ablauf ist in Abbildung 9 bildlich dargestellt. Der generelle Ablauf der Speichermethode des DNA Data Storages wird in Abbildung 8 verdeutlicht.

Unterschiede entstehen dann vor allem bei der Codierung der jeweiligen Codes auf die Basen, welche in einfacherer oder komplizierterer Form ausgeführt werden können.

5.3.1. Codierung

Wie schon erwähnt unterscheiden sich die Codierungen des binären Codes zu den Nukleinbasen je nach Forschungsprojekt. Diese sind notwendig, um lange Homopolymere zu verhindern. Homopolymere sind Ansammlungen an Basen, welche mehr als drei aufeinanderfolgende gleiche Nukleotiden aufweisen. Beim Auftreten solche Homopolymere kann es zu einer erhöhten Fehlerquote bei den darauffolgenden Arbeitsschritten kommen.⁹⁷

Die einfachste Art der Codierung ist das Singel Transcoding, in Form eines Two to One. Hierbei werden zwei Stellen des binären Codes jeweils einer Stelle des Basen Codes der DNA zugeordnet.

Für das bessere Verständnis der Codierung und Decodierung werde ich ein praktisches Beispiel anführen, welches im Unterkapitel Decodierung in gleicher Weise wieder in Binären Code umgewandelt wird. Die Codierung funktioniert auf die Art und Weise, dass 00-A, 01-G, 10-C und 11-T entspricht.

⁹⁵ Sequenzierung: Bei der Sequenzierung handelt es sich um das Auslesen der Abfolge der einzelnen Genome.

⁹⁶ Synthetisierung: Die künstliche Herstellung von DNA-Strängen, welche zur Speicherung der digitalen Daten verwendet werden.

⁹⁷ Thomas BUKO, Nella TUCZKO, Takao ISHIKAWA, DNA Data Storage. *Bio Tech* 12 (2023) 1-17, 2.

01 10 01 00 01 10 00 01 01 11 01 00 01 10 00 01 00 11 10 10

G C G A G C A G G T G A G C A G A T C C

Die Codierungsformen, welche in weiterer Folge des Kapitels vorgestellt werden, gestalten sich insgesamt ein bisschen komplizierter als das einfache Singel Transcoding bezüglich des Weges, wie der binäre Code erreicht wird oder umgewandelt werden kann.

Binary Transcoding kann in unterschiedlichen Formen auftreten. In Abbildung 10 werden diese bildlich dargestellt. Einerseits können zwei Basen einem binären Code zugeordnet werden, oder, wie im oberen Beispiel dargestellt, eine Base pro zwei Stellen des binären Codes.⁹⁸ Das bedeutet, dass Symbol für Symbol gemappt wird, also jedem Datenbit eine von zwei der vier Basen zugeordnet werden kann. Für den Datenbit 0 verwendet das Forscherteam die Basen Adenin (A) und Cytosin (C) und für den Datenbit 1 wurden die Basen Thymin (T) und Guanin (G) zugeordnet.⁹⁹ Diese Art der Codierung kann als die einfachste Methode der Codierung angesehen werden. Die drei anderen dargestellten Versionen der Codierungen gestalten sich etwas komplizierter.¹⁰⁰

⁹⁸ Zhi PING, Shihong CHEN, Sha Joe ZHU, Carbon-based archiving. Current progress and prospects of DNA-based data storage. *GigaScience* (2019), 1-10, 3.

⁹⁹ S.M. Hossein Tabatabaei YAZID, Han Mao KIAH, Eva Ruiz GARCIA, DNA-Based Storage, Trends and Methods (2015) 1-20, 8.

¹⁰⁰ PING, CHEN, ZHU, Carbon-based archiving, 3.

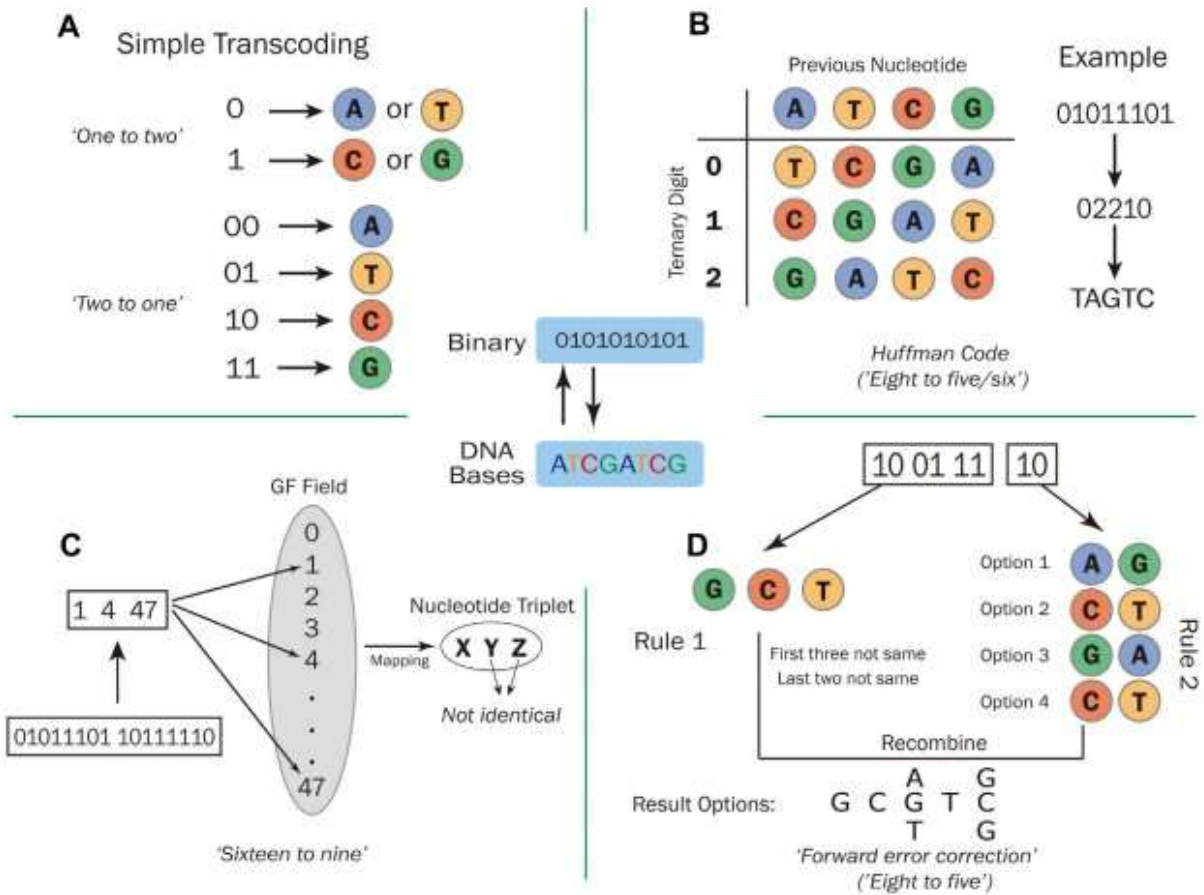


Abbildung 10, Binary Transcoding.¹⁰¹

Eine der Methoden wurde von Church (et al.) eingeführt. Hierbei handelt es sich um die Codierung des binären Codes mit dem ASCII-Coding, welcher nachher mit einem Single Transcoding in die DNA-Basen gewandelt wurde. Um einzelne Datensätze schnell innerhalb des Pools erkennen zu können, wurde am Anfang und am Ende des DNA-Blocks ein Primer angefügt.¹⁰² Zusätzlich kann der binäre Code auch als Code-Paar zu den jeweiligen Basen zugeordnet werden.¹⁰³

ASCII-Coding ist eine häufig angewandte Methode der Codierung. In Abbildung 11 ist dargestellt, inwiefern eine Codierung mit dem ASCII-Coding aussehen könnte. ASCII steht für „American Standard Code for Information Interchange“. Dabei werden die codierten Zeichen einem Bit-Code zugeordnet, welcher aus 7 Bits besteht. Diese Bits werden also in Form eines binären Codes dargestellt, jede der sieben Stellen kann

¹⁰¹ PING, CHEN, ZHU, Carbon-based archiving, 2.

¹⁰² YAZID, KIAH, GARCIA, DNA-Based Storage, Trends and Methods, 8.

¹⁰³ Potomac Institute für Policy Studies, The Future of DNA Data Storage, 8.

einer 0 oder einer 1 entsprechen. Jeder Buchstabe des Alphabets, den einzelnen Sonderzeichen und den einzelnen Zahlen wird somit ein ASCII-Code zugeordnet.¹⁰⁴

message	r	a	m	y
ASCII	114	97	109	121
binary sequence	01110010	01100001	01101101	01111001
DNA sequence	CTAG	CGAC	CGTC	CTGC

Abbildung 11, ASCII Codierung der Nachricht.¹⁰⁵

Ein weiterer Zugang ist die Codierung mit dem Huffman Code. Der Huffman Code wird verwendet, um binäre Daten verlustfrei zu komprimieren. Es wird hier mit der Wahrscheinlichkeit des Auftretens der einzelnen Zeichen gearbeitet. Die Pfade des Huffman-Baums, in dem der Huffman Code angeordnet wird, werden binär angeordnet und durch das jeweilige Bitmuster beschrieben. Je häufiger ein Zeichen auftritt, desto

¹⁰⁴ ASCII-Code und ASCII-Tabellen, in: heise online, <https://www.heise.de/tipps-tricks/ASCII-Code-und-ASCII-Tabellen-4558152.html#:~:text=Wir%20zeigen%20Ihnen%20im%20folgenden,%20was%20es%20mit%20diesem%20Code> [26.01.2024].

¹⁰⁵ Thomas BUKO, Nella TUCZKO, Takao ISHIKAWA, DNA Data Storage. Bio Tech 12, (2023), 3.

näher befindet sich dieses an der Wurzel des Huffman-Baumes.¹⁰⁶

Binary data	P	o	l	y	a	;
	01010000	01101111	01101100	01111001	01100001	00111011
Base 3	12011	02110	02101	222111	01112	222021
Huffman code						
DNA nucleotides	GCGAG	TGAGT	ATCGA	TGCTCT	AGAGC	ATGTGA

(a) Translating binary data to DNA nucleotides via a Huffman code.

		Previous Nucleotide			
		A	C	G	T
Ternary Digit To Encode	0	C	G	T	A
	1	G	T	A	C
	2	T	A	C	G

Abbildung 12, Huffman Codierung.¹⁰⁷

Wie in Abbildung 12 zu sehen ist, werden die binären Codes mithilfe des Huffman Codes in ihre Wahrscheinlichkeit aufgeteilt, welche wiederum anschließend in den Basencode umgewandelt werden können.¹⁰⁸

Je nachdem, welche Methode der Codierung im zweiten Schritt der Arbeit verwendet wurde, muss auch die gleiche Methode wieder bei der Recodierung der Daten verwendet werden, um das richtige Ergebnis des binären Codes und somit auch das richtige Ergebnis des digitalen Files wieder erhalten zu können.

Bei den vorgestellten Prozeduren der Codierung handelt es sich um die in der Literatur am häufigsten angewandten Codierungsmethoden. Allerdings werden auch immer

¹⁰⁶ Huffman-Codierung, in: Informatik Verstehen, <https://www.informatik-verstehen.de/lexikon/huffman-codierung/#:~:text=Huffman-Codierung.%20Die%20Huffman-Codierung%20ein%20Algorithmus%20zur%20verlustfreien%20Komprimierung%20bin%C3%A4rer%20Daten,> [26.01.2024].

¹⁰⁷ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 5.

¹⁰⁸ BORNHOLT, A DNA-Based Archival Storage System, 5.

wieder andere Anwendungen angeführt, auf die ich in dieser Arbeit aber nicht genauer eingehen werde.¹⁰⁹

5.3.2. Synthetisierung (Schreiben von DNA)

„DNA synthesis is the chemical fabrication of sequences of nucleic acids.“¹¹⁰ Dieser Satz fasst diesen Arbeitsschritt innerhalb des DNA Data Storage kurz und prägnant zusammen. Allerdings kann dieser Vorgang der Synthese der DNA in unterschiedlicher Form ausgeführt werden. Zum einen gibt es eine Base to Base-Synthetisierung, die auf chemischen oder auf enzymatischer Basis funktioniert. Wie in Abbildung 13 dargestellt, wird der Anfang mit einer Base und einem Blocker gestartet, welcher die Base während des Prozesses schützen soll und in Form eines chemischen Elementes auftritt, das an die Base angehängt werden kann. In weiterer Folge wird dieser Blocker immer wieder von der letzten Base gelöst. In der Abbildung 13 werden diese Schritte als „Deblock“ bezeichnet. Nach diesem Schritt wird die nächste benötigte Base mit einem Blocker an die vorherige Base angehängt. In Abbildung 13 sind dies die Schritte Zwei und Drei.¹¹¹ Diese Vorgänge werden anschließend so lange wiederholt, bis das Ergebnis dem DNA Strang entspricht, welchen man im vorherigen Schritt aus dem binären Code codiert hat.

¹⁰⁹ Weiterführende und tiefergreifende Ausführung zu anderen Codierungsmethoden finden sich hier: Philipp ANTOWIAK, Jory LIETARD, Mohammad DARESTANI et al., Low cost DNA data storage using photolithographic synthesis and advanced information reconstruction and error correction. *Nature Communications* (2020) 1-10.; Ryan GABRYS, Hans Mao KIAH, Olga MILENKOVIC, Asymmetric Lee Distance Codes for DNA-Based Storage. *IEEE Transactions of Information Theory* 36/8 (2017) 4982-4995.; Nick GOLDMAN, Paul BERTONE, Siyuan CHEN (Hg.), Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA, In: *Nature* (2013) 3-9.

¹¹⁰ DNA Data Storage Alliance, Preserving our digital legacy. An Introduction to DNA Data Storage (2021) 1-37, 24.

¹¹¹ DNA Data Storage Alliance, Preserving our digital legacy, 24.

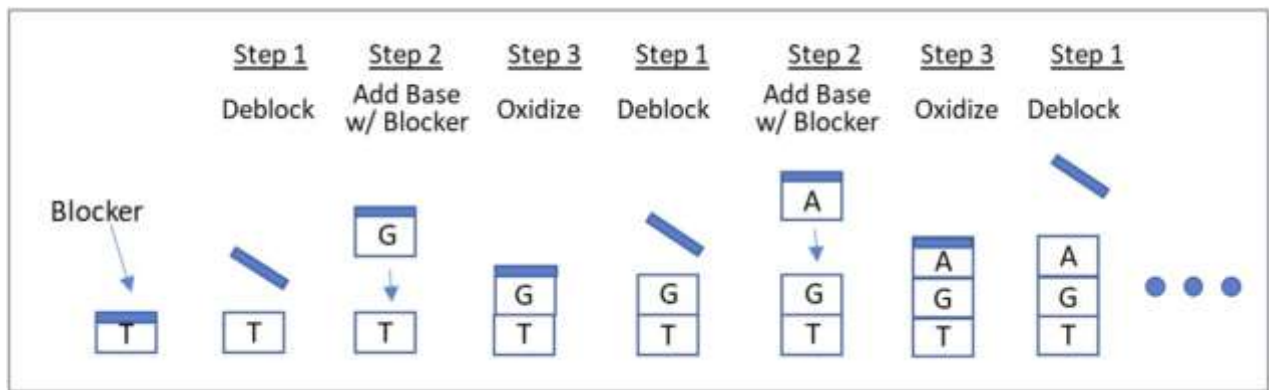


Abbildung 13, Ablauf der chemischen Synthesisierung von DNA.¹¹²

Die meistverwendete Methode der chemischen Synthese ist die Phosphoramidit Synthese. Diese Methode wird seit dem Ende der 1980er Jahre weiterentwickelt und verwendet. Daher handelt es sich hierbei um eine robuste, gut getestete Methode der Synthesisierung. Allerdings sprechen der Kostenfaktor und auch der Zeitfaktor gegen diese Methode. Andere und neuere Methoden weisen daher mehr Vorteile auf.¹¹³

Als Beispiel wäre dafür die enzymatische Synthese zu nennen. Diese verwendet Enzyme in einem ähnlichen Ablauf wie die chemische Synthese. Als Vorteile dieser Methoden gelten die Herstellungsdauer der Synthese, welche schneller von statten geht als bei der chemischen Synthese. Dadurch können die Kosten der Herstellung des Speichers reduziert werden. Außerdem entstehen bei dieser Methode weniger Nebenprodukte. Allerdings ist diese Methode noch nicht auf dem Markt und daher noch nicht in Anwendung.¹¹⁴

Neben der *Base to Base*-Synthesisierung gibt es die Methode der Synthesisierung mit Ligation. Grundsätzlich werden die Basen auch wieder mit einer Base to Base-Synthesisierung hergestellt, wobei die hergestellten Abschnitte kürzer sind. Das bedeutet, dass die zu speichernden Daten schon im Vorhinein in diese kleineren Abschnitte unterteilt werden müssen, bevor eine Synthese der DNA Stränge stattfinden kann. Die Methode arbeitet anschließend mit dem Zusammenstecken der einzelnen, kürzeren DNA Strängen. Dies ist in der Abbildung 14 bildlich dargestellt.

¹¹² DNA Data Storage Alliance, Preserving our digital legacy, 25.

¹¹³ DNA Data Storage Alliance, Preserving our digital legacy, 25.

¹¹⁴ DNA Data Storage Alliance, Preserving our digital legacy, 25.

Die einzelnen Stränge sind als Doppelhelix synthetisiert, welche auf einer Seite einen einzelnen Strang an Basen haben, welcher länger ist als die Doppelhelix. Mit diesen Seiten werden die kurzen Abschnitte der DNA dann zusammengehängt und bilden im Abschluss das vollständige digitale Dokument in DNA Form. Diese Methode wird vor allem deswegen angewandt, weil es bei der Herstellung kürzerer Oligonukleotiden tendenziell zu weniger Fehlerauftritten kommt. Daher wird durch diese Methode auch die Fehlerquote des gesamten Dokumentes verringert.¹¹⁵

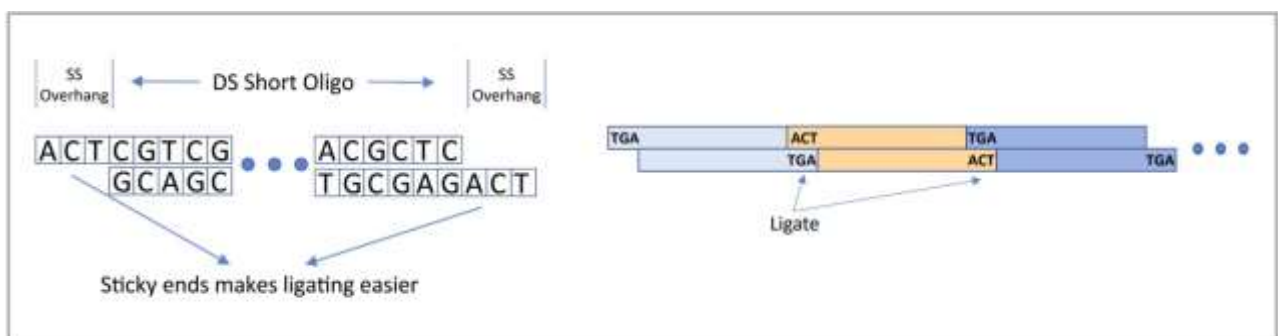


Abbildung 14, Synthese mit Ligation.¹¹⁶

5.3.3. Speicherung der Daten

Für die Speicherung der Daten müssen unterschiedliche Faktoren berücksichtigt werden.

Darunter fallen unter anderen die Kosten für die Behälter und die Kosten des Auslesens. Für die Kosten der Behälter muss die Datenmenge einberechnet werden, die pro Behälter aufbewahrt werden kann.

Ein weiterer Aspekt ist der physische Erhalt der DNA. Auch wenn DNA ein robustes Material ist, so wird es durch Kontakt mit unterschiedlichen Materialien, wie Wasser, UV-Licht, Ozon, Sauerstoff und Ähnlichem, zum Teil zerstört. Zum Schutz der DNA

¹¹⁵ DNA Data Storage Alliance, Preserving our digital legacy, 26.

¹¹⁶ DNA Data Storage Alliance, Preserving our digital legacy, 26.

gibt es unterschiedliche Ansätze. Einerseits die chemische Kapselung der DNA, und andererseits die physische Kapselung der DNA.¹¹⁷

Auch hier gibt es wieder unterschiedliche Ansätze, wie diese Kapselung ausschauen könnte. Eine der Methoden der Verkapselung kann in Siliziumdioxid stattfinden, um die DNA langfristig zu schützen. Bei dieser Speichermethode ist auch das Auslesen der Daten einfach möglich.¹¹⁸ Die Forschungsgruppe um Robert Grass hat herausgefunden, dass die Aufbewahrung von DNA in Siliziumdioxid zu einer besseren Speicherung führt als in Flüssigkeiten oder als „pure solid-state DNA“¹¹⁹. Die Aufbewahrung der DNA in Siliziumdioxid führt bei niedrigen Temperaturen außerdem zu einer langen Haltbarkeit der Daten. Bei einer Lagerungstemperatur von ca. Minus 18 Grad Celsius geht die Forschungsgruppe von einer Haltbarkeit von bis zu 2 Millionen Jahren aus.¹²⁰ Die Methode der Verkapselung der DNA in Siliziumdioxid wird in mehreren Papers als die gängigste Methode angegeben. Wie diese Verkapselung genau aussieht, unterscheidet sich allerdings in den Papers. Bei einigen Forschungsteams wird die DNA in sogenannten „silica nanobeads“¹²¹ aufbewahrt. Bei der Auslesung der Daten wird eine chemische Lösung hinzugefügt, welche die Nanobeads auflöst, während die DNA davon nicht beschädigt wird.¹²²

Andererseits kann die DNA auch in alkalischen Salzen, Polymer, Zucker, Seidenproteinen und Calcium Phosphat Präzipitationen verkapselt werden.¹²³ Auch die Verkapselung in Glas oder Edelstahl ist eine Möglichkeit, welche in der Literatur vorgestellt wird.¹²⁴

Eine weitere Methode ist die Aufbewahrung der DNA-Stränge in chemischen Lösungen, welche mit der DNA in kleinen Röhrchen gefüllt werden. Hauptsächlich handelt es sich bei diesen Lösungen um einen sogenannten TE-Puffer (TRIS-EDTA-Puffer), einer Lösung bestehend aus Trishydroxymethylamionmethan (TRIS) und

¹¹⁷ DNA Data Storage Alliance, 27f.

¹¹⁸ HECKEL, An Archive written in DNA, 236.

¹¹⁹ DONG, SUN, PING et al., DNA storage, 1102.

¹²⁰ DONG, SUN, PING et al., DNA storage, 1102.

¹²¹ Chisom EZEKANNAGHA, Anke BECKER, Dominik HEIDER et al., Design considerations for advancing data storage with synthetic DNA for long-term archiving. *Materials Today Bio* (2022) 1-11, 4.

¹²² EZEKANNAGHA, BECKER, HEIDER et al., Design considerations for advancing data storage with synthetic DNA for long-term archiving, 4.

¹²³ Andrea DORICCHI, Casey PLATNICH, Andreas GIMPEL et al., Emerging Approaches to DNA Data Storage. Challenges and Prospects, In: *ACS Nano* (2022) 17552-17571, 17555.

¹²⁴ Steve NADIS, Tough microbe's DNA could be a lasting archive. *Microbiology* 376/ 6480 (2020) 840-841, 840.

Ethylendiamintetraessigsäure (EDTA).¹²⁵ Da sich pro Behälter viele DNA Stränge mit gespeicherter Information befinden, werden diese in einer Lösung aufbewahrt, welche eine Reaktion zwischen den einzelnen Strängen verhindern soll, beziehungsweise eine Ordnung der Stränge schafft.¹²⁶ Zusätzlich würde sich die aktive Speicherung der DNA mit den oben genannten Verfahren laut dem heutigen Stand der Forschung im Bereich von Fremdfirmen befinden. Daher ist auch das Wissen bezüglich der genauen Funktion der Verkapselung der DNA auf Seiten der Archive nicht unbedingt notwendig.¹²⁷

5.3.4. Sequenzierung (Lesen von DNA)

Bei der Sequenzierung der DNA handelt es sich um das Auslesen der einzelnen Basen der DNA, beziehungsweise „a variety of techniques that detect and report the order of the bases in a strand of DNA.“¹²⁸ Auch hier gibt es unterschiedliche Methoden, wie diese Sequenzierung vorgenommen werden kann. Neben der älteren Methode des „Chain Termination Sequencing“ gibt es seit den 1990er Jahren die Technologie des „Next Generation Sequencing“ (NGS). Diese Technologie schafft es, durch gleichzeitiges Auslesen der DNA eine große Zeitersparnis im Gegensatz zu den Methoden der ersten Generation des Sequenzierens zu schaffen. Innerhalb des NGS können wiederum verschiedene Versionen der Sequenzierung vorgenommen werden. Es gibt allerdings zwei Methoden, welche besonders häufig verwendet werden.¹²⁹

¹²⁵ Storage of DNA, in: Qiagen, <https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/bench-guide/dna/handling-dna/storage-of-dna#:~:text=Storage%20of%20DNA.%20Purified%20DNA%20should%20be%20stored%20at%20%E2%80%9320%C2%B0C> [18.04.2024].

¹²⁶ S.V. RAMANAMURTHY, P.L. KIRANMAYI, S.D. SARAYU et al., DNA Data Storage. *Journal for Research* 3/11 (2018) 32-37, 33.

¹²⁷ Für eine detailliertere Beschreibung des chemischen Prozesses siehe: Reinhard HECKEL, An archive written in DNA. *Nature Biotechnology* 36/3 (2018) 236-237.; Andrea DORICCHI, Casey PLATNICH, Andreas GIMPEL et al., Emerging Approaches to DNA Data Storage. Challenges and Prospects. *ACS Nano* (2022) 17552-17571.; James BANAL, Tyson SHEPHERD, Joseph BERLEANT et al., Random access DNA memory using Boolean search in an archival file storage system. *Nature Materials* 20 (2021) 1272-1280.

¹²⁸ RAMANAMURTHY, KIRANMAYI, SARAYU et al., DNA Data Storage, 29.

¹²⁹ RAMANAMURTHY, KIRANMAYI, SARAYU et al., DNA Data Storage, 29.

Sequencing by synthesis (SBS) arbeitet damit, dass ein einzelner Strang der DNA als Sample genommen wird. Dieser wird dann mithilfe von Synthetisierung mit den dazugehörigen Basen vervollständigt. Dafür wird Polymerase verwendet. Der genaue Vorgang wird in Abbildung 15 dargestellt.

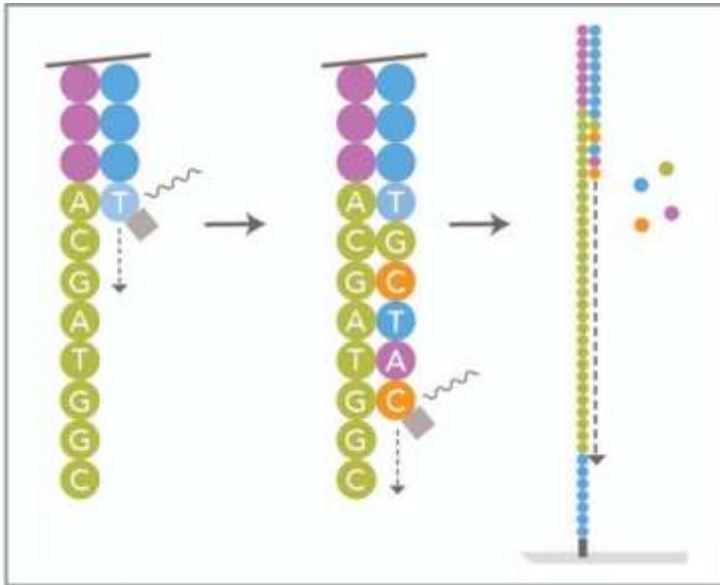


Abbildung 15, Sequencing by Synthesis.¹³⁰

Die zweite Methode heißt Nano Sequencing. Hier wird der zu lesende DNA Strang durch eine Membran gefädelt, welche von einer Elektrolytlösung umgeben ist. Die Lösung wird unter elektrische Spannung gesetzt. Jedes Mal, wenn eine Base des DNA Stranges durch die Membran läuft, wird dies aufgrund dieser Spannung in der Lösung registriert. Je nachdem, um welche Base es sich dabei handelt, wird eine andere Spannung registriert, wodurch der genaue Ablauf des Stranges determiniert werden kann. Ein Vorteil dieser Methode ist, dass der Vorgang in Echtzeit beobachtet werden kann, man somit einen direkten Zugriff zu den Ergebnissen der Sequenzierung bekommt.¹³¹ Eine Darstellung des Nano Sequencing wird in der Abbildung 16 visualisiert.

¹³⁰ LANDSMAN, STRAUSS, The DNA Data Storage Model, 82.

¹³¹ RAMANAMURTHY, KIRANMAYI, SARAYU et al., DNA Data Storage, 30f.

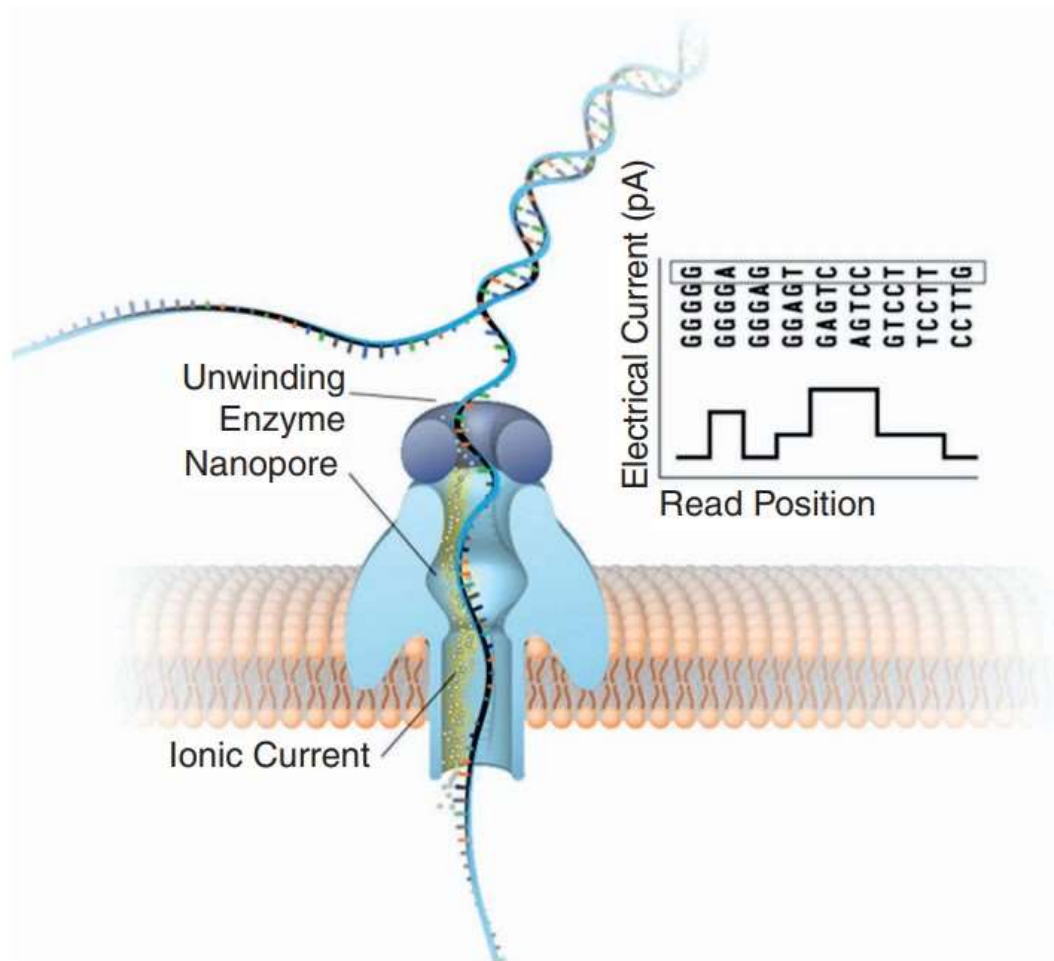


Abbildung 16, Nano Sequencing.¹³²

5.3.5. Decodierung

Nachdem die Sequenzierung abgeschlossen ist, muss die Decodierung der Basen vorgenommen werden, damit die digitale Datei wiederhergestellt und lesbar gemacht werden kann.

Besonders wichtig bei diesem Schritt ist, dass der gleiche Prozess wie bei der Codierung angewandt wird, damit auch die richtigen Daten ausgelesen werden.

¹³² LANDSMAN, STRAUSS, The DNA Data Storage Model, 83.

Sollte man bei der Codierung einen Huffman Code angewandt haben, so muss dieser wieder in umgekehrter Version bei der Decodierung angewandt werden.

Beim Prozess der Decodierung kann es auch passieren, dass DNA Stränge zerstört werden, weswegen es von besonderer Wichtigkeit ist, dass mehrere Kopien von den Daten erstellt werden. Es hat höchste Priorität, dass Fehlern bei der Codierung entgegengewirkt wird, wie auch die Gewährleistung, dass es bei einer eventuellen Zerstörung der Daten beim Auslesen zu keinen Problemen kommt.

Dies bedeutet für das kleine praktische Beispiel, welches ich am Anfang des Kapitels 4.2.1. angeführt habe, folgendes:

G C G A G C A G G T G A G C A G A T C C

01 10 01 00 01 10 00 01 01 11 01 00 01 10 00 01 00 11 10 10

Der Basen Code muss wieder mit dem gleichen Schema behandelt werden, wie bei der Codierung, damit auch wieder der gleiche binäre Code als Ergebnis vorhanden ist. Also gilt hier in umgekehrter Form A-00, G-01, C-10, T-11.

Bei der praktischen Anwendung von DNA Datenspeicherung werden die umgewandelten DNA-Stränge auch immer mit einem Primer versehen, um Anfang und Ende des jeweiligen Dokumentes zu kennzeichnen. Dieser Primer muss auch bei der Decodierung beachtet werden, um nicht das falsche Dokument zu rekonstruieren.

5.3.6. Error Correction

Bevor ich die Nachteile der Verwendung von DNA als Speicher ausführe, möchte ich noch auf ein Konzept eingehen, welches bei der DNA-Speicherung von besonderer Bedeutung ist. Hierbei handelt es sich um das Prinzip der Error Correction. Gerade bei der Synthetisierung und Sequenzierung von DNA können häufig Fehler auftreten. Unter anderem durch eventuell vermehrtes Auftreten von den Basen C und G, also

Homopolymere. Daher braucht es eine Error Correction, welche hilft, diese Fehler auszubessern oder vorzubeugen.

Hierbei wird zwischen logischer und physischer Redundanz der DANN unterschieden. Bei physischer Redundanz werden mehrere Kopien der zu speichernden Daten auf DNA hergestellt, um die Chancen zu erhöhen Fehlern entgegenzuwirken. Allerdings wird bei dieser Methode nicht komplett verhindert, dass keine potentiellen Fehler im Prozess auftreten können. Anders wiederum bei der logischen Redundanz. Hier werden zusätzliche Informationen eingearbeitet, also Metadaten zu dem Prozess der Datenspeicherung und der Herstellung der DNA.¹³³

Verschiedene Forschungsteams haben hier unterschiedliche Herangehensweisen angewendet, um eine Methode zu finden, mit der die geringste Fehlerquote erreicht werden kann. Das Team um Church arbeitet mit der Methode, dass einzelne Sequenzen öfter synthetisiert werden. Damit können durch Vergleiche Fehler schnell gesehen werden. Allerdings gibt es keine extra Methode für die Error Correction.

Das Team um Goldman arbeitet in der Error Correction mit dem Huffman Code, um Homopolymere zu vermeiden. Zusätzlich werden regelmäßige Abschnitte der DNA überlappt, wodurch Fehler im Vergleich mit diesen Sequenzen schneller erkannt werden können.¹³⁴

Jede dieser Methoden nähert sich dem Konzept der Error Correction mit einem anderen Zugang, aber allen ist gleich, dass ein Informationsverlust damit ganz oder teilweise verhindert werden soll.¹³⁵

5.4. Nachteile

Es gibt natürlich nicht nur Vorteile dieser alternativen Methode zur Speicherung von digitalen Daten. Einer der angeführten Nachteile in der Literatur ist die Tatsache,

¹³³ EZEKANNAGHA, BECKER, HEIDER et al., Design Considerations for advancing data storage with synthetic DNA for long-term archiving, 5f.

¹³⁴ Thomas HEINIS, Jamie ALNASIR, Survey of Information Encoding Techniques for DNA. *ResearchGate* (2022) 1-28, 15.

¹³⁵ EZEKANNAGHA, BECKER, HEIDER et al., Design Considerations for advancing data storage with synthetic DNA for long-term archiving, 8.

dass die digitalen Daten, transformiert in ihren binären Code, Millionen an Bits lang sein können, gerade, wenn es sich um große Datenmengen handelt. Und auch wenn sich die Technik der Herstellung synthetischer DNA laufend weiterentwickelt, so steht sie momentan noch auf dem Stand, dass keine längeren DNA-Stränge als 100 bis 200 Nukleotide in zusammenhängender Weise synthetisiert werden können. Daher müssten große digitale Daten über mehrere Fragmente von DNA gespeichert werden. Dadurch wird die praktische Arbeit mit diesen DNA-Pools nicht erleichtert und auch eine Anwendung im Archiv massiv erschwert. Denn dies bedeutet, dass die einzelnen Teile der Daten gut gekennzeichnet werden müssen, damit sie auch wieder auffindbar und richtig zusammensetzbar sind.¹³⁶

Weiters entstehen gerade bei der Synthetisierung und Sequenzierung der DNA immer noch Fehler bei der Anordnung der Nukleotide. Der Anteil dieses Fehlers liegt meist bei einem Prozent der Synthetisierung und der Sequenzierung, allerdings häufen sich die fehlerhaften Nukleotide bei längerer Laufzeit der Herstellung der DNA.¹³⁷ Dadurch werden auch die digitalen Daten nicht mehr korrekt gespeichert und dargestellt.

Ein weiteres Problem des jetzigen Forschungsstandes ist die noch fehlende Möglichkeit, aus dem Pool der gespeicherten Daten auf DNA spezifische Daten separat anzuschauen. Das heißt, es muss für den Leseprozess meist noch immer der gesamte Pool der gespeicherten Daten wieder decodiert werden. Dadurch entsteht ein Mehraufwand an Arbeit und Kosten.¹³⁸

Neben den technischen Nachteilen ist auch der Faktor der Kosten ein Nachteil, welchen man nicht außer Acht lassen darf. Die Herstellung von synthetischer DNA wurde seit ihrer ersten Entwicklung laufend günstiger. Die Kosten unterscheiden sich je nach Institution, welche das jeweilige Projekt durchführten. Die Reduktion der Kosten folgte bis 2008 ungefähr der Moore's Kurve¹³⁹, seit 2008 sinken die Kosten

¹³⁶ HECKEL, An archive written in DNA, 236.

¹³⁷ BORNHOLT, A DNA-Based Archival Storage System, 2.

¹³⁸ BORNHOLT, A DNA-Based Archival Storage System, 2.

¹³⁹ Moore's Law: Die Gesetzmäßigkeit des Moorschen Gesetz besagt, dass die Nummer der Transistoren, also ein Verstärker elektrischer Signale, sich alle zwei Jahre auf einem Mikrochip verdoppeln. Dadurch geht man davon aus, dass sich die Leistung des Computers alle zwei Jahre ebenfalls verbessert und zusätzlich die Kosten nicht in gleicher Form dazu steigen. Man zahlt also im Laufe der Zeit weniger für einen Computer, der mehr Leistung abrufen kann, genauer gesagt verdoppeln sich die Transistoren auf den Chips, während die Kosten halbiert werden. Die Kurve dieses Gesetzes wächst exponentiell und wurde von Gordon Moore im Jahr 1965 formuliert. (What Is Moore's Law and Is It Still True? (investopedia.com) [26.01.2024].

jedoch schneller, als diese Kurve vorhersagen würde, da sich die Technik schneller weiterentwickelt.¹⁴⁰

Dennoch können die Kosten der Herstellung und der Speicherung der digitalen Daten auf DNA noch immer bei 3.500 Dollar pro Megabyte liegen. Dies ist der Stand aus dem Jahr 2017. Man kann davon ausgehen, dass diese Kosten auch in den kommenden Jahren weiter sinken werden. Dennoch muss der Kostenfaktor bei der Implementierung eines DNA-Archives in praktischer Hinsicht mitbedacht werden.¹⁴¹ Dies bedeutet, dass der Aufbau eines Archives, dessen Daten auf DNA gespeichert werden, in den ersten Schritten einen größeren Kostenaufwand benötigt. Allerdings kann sich das Archiv nach der Implementierung des Systems laufende Kosten, wie die der Datenmigration und Ähnlichem, sparen.

¹⁴⁰ Paul MUIR, Shantao LI, Shaoke LOU et al., The real cost of sequencing. Scaling computation to keep pace with data generation. *Genome Biology* (2016) 4-9, 4.

¹⁴¹ Potomac Institute für Policy Studies, The Future of DNA Data Storage, 16.

6. DNA Data Storages als Möglichkeit der digitalen Archivierung

6.1. Theoretische Umsetzungen

Es gibt viele unterschiedliche Projekte, die sich mit verschiedenen Methoden und Herangehensweisen an der Herstellung von DNA, an der Codierung der digitalen Daten und der Lagerung der fertigen, auf DNA gespeicherten Daten, auseinandersetzen. Anschließend sollen ein paar dieser theoretischen Konzepte und Projekte, deren Besonderheiten und Vorgehensweisen vorgestellt werden.

6.1.1. Archival Storage System nach Church

Das Team um George Church arbeitete schon 2012 an einem Speichersystem auf DNA Basis. Im Gegensatz zu vorherigen Ansätzen unterscheidet sich der Ansatz von Church und seinem Team in fünf Punkten. Es wurde die Codierung von zwei Bits per Base auf ein Bit per Base geändert. Dadurch wird es möglich, bei der Codierung zum Beispiel ein erhöhtes Vorkommen von C- und G-Basen zu vermeiden. Durch den neuen Ansatz können auch lange DNA-Stränge vermieden werden, welche bei der Synthetisierung Schwierigkeiten verursachen. Um vermehrte Fehler bei der Synthetisierung entgegenzuwirken, sollen mehrere Kopien die jeweiligen Fehler der anderen Kopien korrigieren. Mit der Anwendung des Next Generation Sequencing können größere Mengen an Informationen mit weniger Geld codiert und gespeichert werden. Die Methode von Church und seinen Kolleginnen und Kollegen ist, wie andere Methoden und Projekte auch, vorerst nur ein Zwischenschritt der Technologie und es ist mit Weiterentwicklungen der Methode zu rechnen.¹⁴²

¹⁴² George M. CHURCH, Yuan GAO, Sriram KOSURI, Next-Generation Digital Information Storage in DNA. *Science* (2012), 1-2, 1.

6.1.2. Archival Storage System nach Bornholt

Das theoretische Konzept, welches das Team von James Bornholt entwickelt hat, versucht die bisherigen Konzepte und Herangehensweisen der DNA Datenspeicherung zu verbessern. Einerseits sollen mit dieser Methode Redundanzen verringert werden und andererseits soll dadurch eine Methode geschaffen werden, mit der ein Direktzugriff auf Daten ermöglicht werden soll. Damit soll erreicht werden, dass für ein Auslesen von Informationen keine größeren Datenmengen decodiert und ausgelesen werden müssen, als notwendig oder gebraucht ist.¹⁴³ Wie im vorherigen Kapitel im Abschnitt der Nachteile der DNA Speichermethode schon erwähnt wurde, handelt es sich hierbei um einen der Defizite, welche die praktische Anwendung des Speichers erschwerten.

Das Speichersystem zeichnet sich dadurch aus, dass die Dichte des Speichers sehr hoch ist und das Speichermaterial sehr haltbar. Es soll dabei aber auch als Teil einer Speicherhierarchie angewandt werden, bei der der DNA Speicher die unterste Basis bildet, wie in Abbildung 17 zu sehen ist.

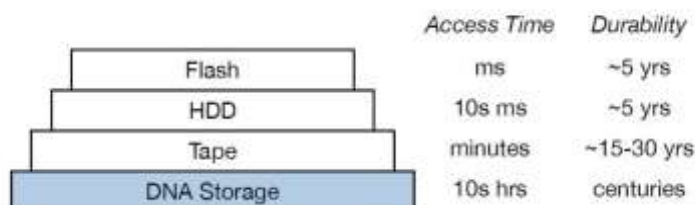


Abbildung 17, Speicherhierarchie nach Bornholt et al.¹⁴⁴

Das System besteht aus dem Synthesizer der DNA, einem Speichercontainer für die DNA, welcher DNA in einzelnen Abteilungen speichert, und einem Sequenzer, welcher die DNA wieder in digitale Daten umwandelt. Das Speichersystem selbst ist in Abbildung 18 zu sehen.

¹⁴³ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 2.

¹⁴⁴ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 3.

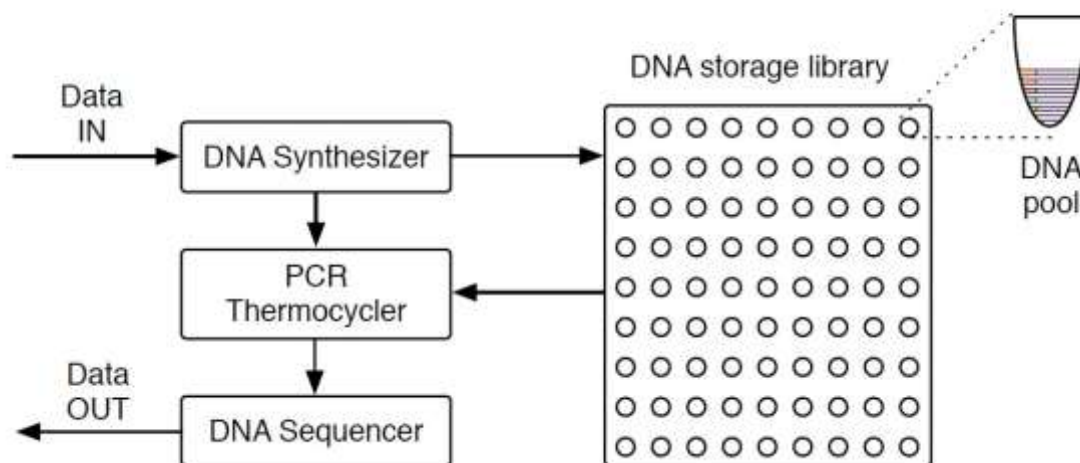


Abbildung 18, Darstellung des Speichersystems von Bornholt et al.¹⁴⁵

Auch wenn die DNA in einzelnen Abteilungen gespeichert wird, müssen die notwendigen Informationen zum gespeicherten Inhalt auf der DNA gespeichert werden. Diese werden von Bornholt et al. in Form von Identifikationstags umgesetzt, also einer „simple key-value architecture“¹⁴⁶. Der Key wird einem DNA-Pool zugeordnet und kann mit einem Mechanismus die spezifischen und gewünschten Daten aus dem vorher festgelegten Pool herausholen. Der genannte Key ist in diesem Fall ein PCR (Polymerase Chain Reaction) -Primer. Dieser wird im Schreibprozess der synthetischen DNA an die jeweiligen Stränge gehängt, wie in Abbildung 19 dargestellt wird.

Bei der sogenannten Polymerase Kettenreaktion, oder Polymerase Chain Reaction im Englischen, von dem die Abkürzung abgeleitet werden kann, handelt es sich um eine Methode, mit der mehrere Kopien der vorliegenden DNA-Stränge hergestellt werden können. Damit können Milliarden Kopien der gewünschten DNA-Abschnitte innerhalb von einigen Stunden hergestellt werden.¹⁴⁷

¹⁴⁵ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 3.

¹⁴⁶ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 3.

¹⁴⁷ Polymerase Chain Reaction, in: Britannica, [https://www.britannica.com/science/polymerase-chain-reaction#:~:text=polymerase%20chain%20reaction%20\(%20PCR\),%20a%20technique%20used%20to%20make](https://www.britannica.com/science/polymerase-chain-reaction#:~:text=polymerase%20chain%20reaction%20(%20PCR),%20a%20technique%20used%20to%20make) [08.08.2024].

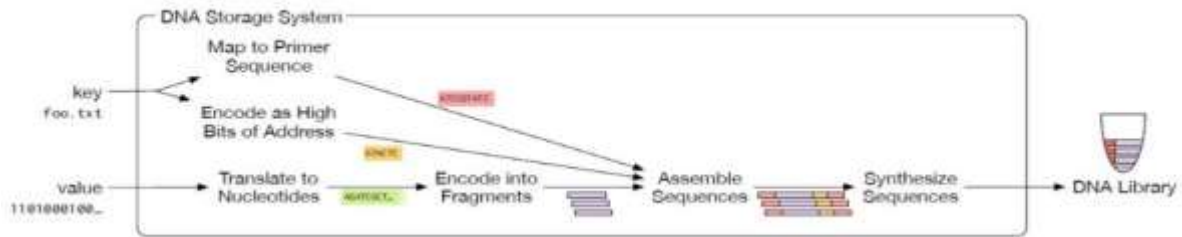


Abbildung 19, Darstellung des Hinzufügens des PCR-Primers als Key.¹⁴⁸

Beim Leseprozess können durch den gleichen Primer die angebrachten Keys verstärkt werden und sind dadurch leicht aufzufinden. Dies ist in Abbildung 20 zu sehen. Durch das Auslesen werden die DNA-Stränge mit der gewollten Information physisch aus dem Pool herausgezogen. Dadurch wird die Quantität der vorhandenen DNA durch jeden Leseprozess verringert. Für das Forschungsteam stellt dies allerdings kein Problem dar, da die Pools durch einfache Prozesse wieder mit neuen DNA-Strängen aufgefüllt werden können.¹⁴⁹

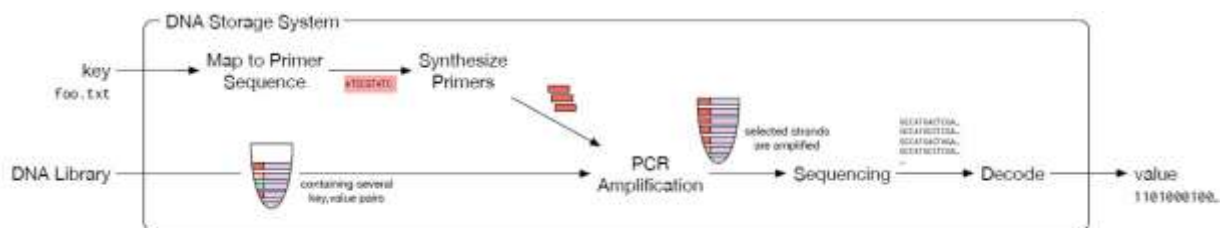


Abbildung 20, Darstellung des Auslesens und der Verstärkung des Primers.¹⁵⁰

Die Entscheidung, die DNA Stränge in einzelnen Pools zu speichern, basiert laut dem Forschungsteam darauf, dass durch diese Maßnahme eine vereinfachte Handhabung der Speicherung und des Auslesens ermöglicht wird. Sollten alle Daten in einem einzigen Pool aufbewahrt werden, so können unterschiedliche Komplikationen auftreten. Einerseits können Primer aufeinander reagieren,

¹⁴⁸ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 4.

¹⁴⁹ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 4.

¹⁵⁰ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 4.

andererseits kann es dadurch passieren, dass im Leseprozess nicht alle gewünschten Daten ausgelesen werden.¹⁵¹

Um die theoretischen Überlegungen des Forschungsteams auch praktisch zu überprüfen, wurden vier digitale Bilder in unterschiedlicher Größe auf DNA gespeichert. Für die Umwandlung von binärem Code in die DNA Basen wurden zwei unterschiedliche Codierungsmethoden angewandt, einerseits die Goldman-Codierung und andererseits die XOR (Exclusive OR)-Codierung.¹⁵² Das Experiment sollte darstellen, ob es mit der vorher erklärten Methode der Keys möglich ist, nur die gewünschten Daten aus einem Pool von DNA herauslesen zu können. Die Verstärkung der gewünschten Daten durch den jeweiligen Primer hat laut dem Forschungsteam funktioniert.¹⁵³

Laut dem Forschungsteam hat DNA-Datenspeicherung das Potential, als zukünftige Methode der Speicherung sinnvoll eingesetzt werden zu können. Aber wie auch andere Paper und Forschungsarbeiten aus diesem Bereich sagen, sind sie der Meinung, dass es noch Forschungsarbeit benötigt, bis es zu einem tatsächlichen Einsatz kommen kann.¹⁵⁴

6.1.3. Archival Storage System nach Goldman

Nick Goldman und sein Team haben eine neue Herangehensweise an die Synthetisierung der DNA und der Speicherung der digitalen Daten getestet, indem sie mit dem sogenannten in vitro Ansatz gearbeitet haben. Das bedeutet, dass die zu speichernden Daten in einen hypothetischen, langen Strang von DNA umgewandelt werden. Dieser Strang wird dann allerdings in kürzeren, leichter zu synthetisierenden

¹⁵¹ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 3.

¹⁵² XOR-Codierung: Hierbei handelt es sich um eine Entweder-Oder-Codierung. Für DNA-Codierung bedeutet dies folgendes: „The XOR gate evaluates “true” only if exactly one of the input sequences evaluates “true.” With binary inputs, XOR is defined as evaluating “true” if input values are opposite. In DNA-based logic gates, the XOR gate is the most simplistic design in that no base sequence needs to be supplied to the gate. Opposite input sequences are complementary and will bind together to form a double stranded sequence. If the inputs are not opposite, then the sequences will not be able to bind to one another and DNA Base will destroy both input sequences.” (Christy GEARHEART, Eric ROUCHKA, Benjamin ARAZI, DNA-Based Active Logic Design and its Implications. *Journal of Emerging Trends in Comouting and Information Science* 3/5 (2021), 756-766, 758.)

¹⁵³ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 7.

¹⁵⁴ BORNHOLT, LOPEZ, CARMEAN et al., A DNA-Based Archival Storage System, 11.

Strängen von DNA tatsächlich hergestellt. Getestet hat das Team diese Methode an fünf unterschiedlichen digitalen Datenformaten, unter anderem PDF, JPEG, MP3 (Moving Picture Experts Group) und Text.¹⁵⁵ Die praktische Umsetzung arbeitet mit Überlappungen bei der Synthetisierung der DNA. Damit soll einem Datenverlust vorgebeugt werden.¹⁵⁶ Das Projekt hat auch einen Blick darauf geworfen, wie sich eine DNA Datenspeicherung in Bezug auf größere Datenmengen umsetzen lassen würde. Die Anzahl der Basen synthetisch hergestellter DNA würde mit der Methode von Goldman (et al.) linear zu den zu codierenden Daten ansteigen. Zusätzlich müssen aber auch die Indizierungsinformationen zur Wiederherstellung der gesamten Dateien aus den kurzen DNA-Strängen miteinander berechnet werden. Das Team geht davon aus, dass sich die Effizienzrate der Codierung auch bei größeren Datenmengen noch bei über 70 Prozent befinden wird. Das bedeutet, dass DNA Datenspeicherung auch noch im Bereich von Petabyte umsetzbar bleibt. Ebenso verhält es sich laut Goldman (et al.) auch mit den Kosten, welche in Bezug auf das gespeicherte Datenvolumen nur wenig ansteigen.¹⁵⁷

6.1.4. DNA MemoChip

Das Team um George Stefano beschäftigt sich mit einer neuen Methode, wie auf DNA gespeicherte Daten auf vereinfachte Weise aufbewahrt und abgerufen werden können. Denn gerade der Aspekt des Abrufens der Daten aus der DNA ist ein zeitaufwendiger und fallspezifischer Prozess.¹⁵⁸

Um diese Schritte zu vereinfachen und das Abrufen der Daten schneller und leichter zu gestalten, soll die DNA auf einen Chip gespeichert werden, welcher in weiterer Folge vielleicht auch mit technischen Geräten, wie Handys und Laptops kompatibel

¹⁵⁵ Nick GOLDMAN, Paul BETRONE, Siyuan CHEN et al., Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. *Nature* (2013), 77-80, 77f.

¹⁵⁶ GOLDMAN, BETRONE, CHEN et al., Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA, 78f.

¹⁵⁷ GOLDMAN, BETRONE, CHEN et al., Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA, 79.

¹⁵⁸ George B. STEFANO, Fuzhou WANG, Richard M. KREAM, DNA MemoChip. Long-Term and High Capacity Information Storage and Select Retrieval. *medical Science Monitor* (2018), 1185-1187, 1186f.

gemacht werden sollte. Mit dieser Methode sollen Ansprüche der Datenspeicherung in technischer und informationsspezifischer Hinsicht erfüllt werden.¹⁵⁹

Allerdings gibt es mit dem jetzigen Stand der Forschung und der Technik noch einige Probleme, welche für die Umsetzung des MemoChips gelöst werden müssen, bevor es zu einer praktischen Anwendung kommen kann.¹⁶⁰

Dennoch handelt es sich hierbei um einen interessanten Ansatz, der auch für Archive verfolgt werden könnte. Diese Methode der DNA Datenspeicherung könnte aber auch eine praktikable Speichervariante für den Alltag werden. Eine breitere Anwendung dieser Methode kann auch für Archive von Vorteil sein, sei es auf finanzieller Ebene oder in Bezug auf Anwendung und Handhabung der neuen Speicherform.

6.2. Projekte zur praktischen Umsetzung des DNA Data Storage

Auch wenn in der Wissenschaft noch einige Schritte gesetzt werden müssen, damit ein DNA Datenspeicher in praktische Anwendung in weiten Bereichen kommen kann, so beschäftigen sich einige Projekte mit verschiedenen Konzepten, wie diese praktische Umsetzung ausschauen könnte. Vor allem in der Europäischen Union finden sich unterschiedliche Projekte, welche in die Zukunft mit DNA Datenspeicher investieren und an einer vereinfachten und verbreiteten Einsetzung desselben arbeiten. Auf den folgenden Seiten möchte ich mich einer ausgewählten Zahl dieser Projekte widmen und diese kurz vorstellen.

6.2.1. NEO-Projekt – Next Generation Molecular Data Storage

Das NEO-Projekt wurde durch die Europäische Union ins Leben gerufen. Der genaue Titel des Projektes lautet „Molekulare Datenspeicherung der nächsten

¹⁵⁹ STEFANO, WANG, KREAM, MemoChip, 1186f.

¹⁶⁰ STEFANO, WANG, KREAM, MemoChip, 1186f.

Generation“. Die Leitung des Projektes hat die Universität Paderborn inne, weitere beteiligte Universitäten sind die Technische Universität Graz, die Katholische Universität Leuven, das Imperial College London und die Universität Surrey.

Das Projekt arbeitet mit DNA in Nanostrukturen, welche in unterschiedlichen Formen hergestellt werden können und als "mikroskopisch kleine, komplexe Geflechte aus DNA-Strängen"¹⁶¹ auf der Webseite der Universität Paderborn beschrieben werden.

Die Daten werden auf den Nanostrukturen gespeichert und dann innerhalb einer Lösung in kleinen Röhrchen aufbewahrt. Das Auslesen erfolgt mit Hilfe eines Artificial-Intelligence Modells (AI-Modell), welches ein Mikroskopbild der Nanostrukturen ausliest und eine Umwandlung in den ursprünglichen Bit-Stream erstellt. Für diesen Teil des Projektes ist die Technische Universität Graz verantwortlich, welche sich mit dem Anlernen des AI-Modells und der Error Correction beschäftigt.

Mit dem Leiter des Projektes, Dr. Adrian Keller von der Universität Paderborn und Dr. Robert Peharz von der Technischen Universität Graz durfte ich ein Interview führen und Fragen zu dem Projekt und auch archivspezifische Fragen stellen. Der Fragebogen zu den besagten Interviews findet sich am Ende der Arbeit im Anhang.

Ziel des Forschungsprojektes ist es, Strukturen zu schaffen, welche es ermöglichen, Daten so lange wie möglich zu speichern. Und dies soll in einer Form umgesetzt werden, welche es schafft, die Daten so schnell wie möglich zu verändern, also als dynamische Daten zu speichern. Die lange Haltbarkeit der gespeicherten Daten auf DNA habe ich bereits im vorherigen Kapitel erwähnt und genauer erläutert. Diese ist auch in diesem Projekt gegeben, weil die Haltbarkeit der DNA außerhalb der Zelle auch von diesem Projekt auf 100-1000 Jahre geschätzt wird. Der Vorteil der Verwendung von DNA-Nanotechnologie liegt darin, dass durch das Steckkastenprinzip der Nanostrukturen der DNA jegliche Veränderungen der zu speichernden Daten schneller vorgenommen werden können als bei der typischen

¹⁶¹ Mit DNA zur nachhaltigen Datenspeicherung, in: Universität Paderborn, <https://www.uni-paderborn.de/nachricht/127311>, [13.03.2024].

Herangehensweise mit synthetisch hergestellter DNA. Diese muss jedes Mal neu geschrieben werden, wodurch ein Mehr an Kosten und ein größerer Arbeitsaufwand entsteht.

Um sich diese Nanostruktur von DNA besser vorstellen zu können, sieht man in Abbildung 21 und Abbildung 22 dargestellt, wie genau diese unterschiedlichen Formen der DNA eigentlich ausschauen können.

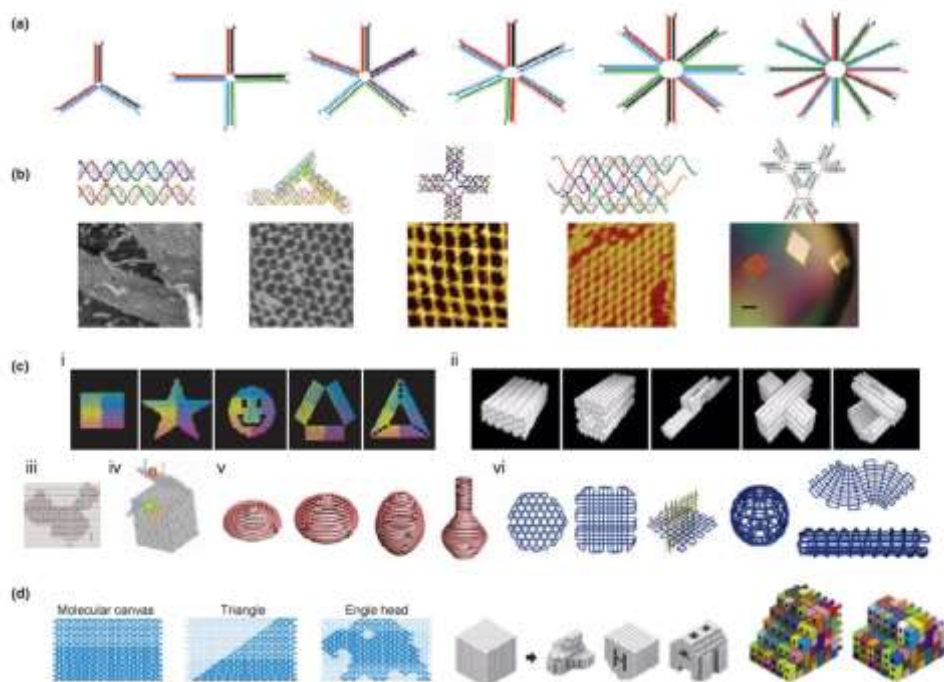


Abbildung 21, Nanostrukturen von DNA. ¹⁶²

Mithilfe der Nanostrukturen kann die DNA in unterschiedliche Formen gebracht werden, indem sie gefaltet wird. Diese Nanostruktur soll dann, wie eine Art Steckbrett verwendet werden, auf dem Proteine angebracht werden können, welche als Marker für das Dasein einer Eins oder einer Null fungieren. Dafür wird ein fester Satz von DNA verwendet. Dadurch sind die Materialkosten günstiger und auch besser skalierbar, weil nicht für jede Änderung der Daten ein neuer Satz an DNA produziert und synthetisiert werden muss.

¹⁶² Jie CHAO, Yunfeng LIN, Huajie LIU, Lianhui WANG, Chunhai FAN, DNA-based plasmonic nanostructures. *Materials Today* (2015) 1-10, 5.

Die Dauer der Speicherung von binären Daten auf den Nanostrukturen dauert ungefähr 90 Minuten, wenn der zu speichernde Datensatz die Zahl der benötigten Strukturen von 100 nicht überschreitet. Jegliche nachträgliche Veränderung der Daten nimmt 2 bis 12 Stunden in Anspruch, je nach Größe der Veränderung. Da es sich bei Daten im Archiv grundsätzlich um unveränderliche Daten handelt, kann man davon ausgehen, dass dieser Service von Archiven nicht in Anspruch genommen werden muss.

Auf einer einzelnen Nanostruktur des NEO-Projektes können bis zu 200 Bits gespeichert werden. Die einzelnen Nanostrukturen werden gemeinsam in einer Lösung in Röhrchen aufbewahrt. Die Zahl der Nanostrukturen innerhalb der Röhrchen sind dadurch beschränkt, wie viele Marker zur Unterscheidung der einzelnen Strukturen angefertigt werden können. Die Röhrchen, welche 50 Microliter umfassen, werden für eine langfristige Lagerung bei -20 bis -70 Grad Celsius eingefroren. Diese Temperaturen gewährleisten ein langes Leben der DNA.

Zum Auslesen der Daten kann die Lösung unter ein Rasterelektronenmikroskop gelegt werden, welches ein Bild generiert, auf der die einzelnen Nanostrukturen zu erkennen sind. Auf diesen erscheinen dann die gesteckten Proteine in Form von hellen Punkten, welche sich ersichtlich vom Hintergrund der Nanostrukturen und der Lösung abheben. Die Technische Universität Graz setzt an diesem Punkt das AI-Modell ein, welches gerade dazu trainiert wird, diese unterschiedlichen Erhebungen zwischen Lösung, Nanostruktur und Proteinen auslesen zu können, und daraus dann den binären Code der Datei wiederherzustellen. Die AI arbeitet also mit den Höheninformationen und kann die Daten aus der Lösung, quasi wie aus einer CD ablesen. Für das Auslesen selbst braucht die AI nicht länger als ein paar Millisekunden. Um schlussendlich aber den richtigen Code wieder zu erlangen, muss noch eine Error Correction über den ausgelesenen Code laufen. Dieser kann redundante Zusatzdaten erkennen und aufzeigen. Diese werden, wie bei anderen Methoden der DNA Datenspeicherung, angefertigt, um etwaigen Schäden vorzubeugen und einen Verlust von Daten zu verhindern, oder um Fehlern bei der Synthetisierung entgegenzuwirken. Das Ziel ist hier, dass diese beiden Schritte zu einem zusammengeführt werden können. Es soll gleichzeitig das Bild mit den Daten ausgewertet und die redundanten Daten entfernt werden, um gleich den richtigen

binären Code zu erlangen. Das würde zu einer dateneffizienteren und qualitativ besseren Datenauslesung führen.

Wie könnte nun die praktische Umsetzung dieses Projektes ausschauen?

Grundsätzlich handelt es sich um eine „umweltschonendere“ Alternative als die zurzeit verwendeten Methoden der Datenspeicherung auf technischen Datenträgern. Aufgrund der ständigen Datenmigration entsteht laufend Elektroschrott. Außerdem basieren diese Datenspeicher auf Siliciumtechnologie. Im Gegensatz dazu basiert DNA Synthese auf ökologisch nachhaltigeren Chemie.

Die Herstellung der DNA und deren strukturelle Zusammensetzung würde laut Dr. Keller bei Fremdfirmen stattfinden, welche spezifische Nanostrukturen für Auftraggeber herstellen können. Ebenso können die Daten anschließend bei Fremdfirmen gelagert werden, welche auch das Auslesen und Modifizieren derselben übernehmen, wenn es denn von Nöten ist. Das würde für Archive und deren Angestellten keine zusätzliche Ausbildung voraussetzen. Ebenso müssten sich Archive nicht um die Anschaffung neuer Gerätschaften für das Auslesen der Daten auf den DNA Nanostrukturen kümmern. Allerdings würden sich auch diese Anschaffungen laut Dr. Keller in einem überschaubaren Rahmen halten, da es sich hauptsächlich um den Ankauf und Verwendung des Mikroskops handelt, mit welchem die Bilder der Nanostrukturen aufgenommen werden müssen. Eine solche Aufnahme dauert ungefähr eine halbe Stunde, bevor es dann von der Software ausgelesen werden kann.

Für kleinere Archive wäre die Inanspruchnahme einer Fremdfirma wahrscheinlich günstiger, ein größeres Archiv könnte zumindest die Lagerung der Daten und auch das Auslesen mit eigenen Mitteln übernehmen.

Das Ziel des Projektes ist es nicht, am Ende eine fertige Speichermethode entwickelt zu haben, welche direkt angewandt werden kann. Allerdings handelt es sich hierbei um einen Beitrag zur Forschung, um der tatsächlichen Anwendung näher zu kommen. Dr. Peharz hat in diesem Zusammenhang von dem "Technology Readiness Levels" gesprochen. Die Wahrscheinlichkeit, dass diese Speichermöglichkeit zur umfassenden Anwendung kommt, sei frühestens in ca. 50 Jahren gegeben, und auch hierfür wird die Wahrscheinlichkeit nicht sonderlich hoch angegeben. Ein großer

Faktor, der bisweilen noch dagegenspricht, ist die Kostenfrage der Herstellung eines solchen Speichers. Auch wenn die Kosten für die Synthetisierung und das Schreiben auf DNA in den vergangenen Jahren zurückgegangen sind, so liegen sie dennoch immer noch vergleichsweise hoch. Das Schreiben eines BITs (Binary Digit) liegt zwischen 12 und 14 Cent, was bedeutet, dass 1 GB (Gigabyte) an Daten zu schreiben noch 1 Million Dollar kostet.

Der Ansatz der Nanotechnologie der DNA Strukturen ist ein sehr interessanter und erscheint auch in Bezug auf die Speicherung von Daten im Archiv sinnvoll. Die Verwendung von Siliciumplatten, auf denen die DNA Strukturen unter einer Beschichtung aufbewahrt werden können, ohne dass es besondere Vorkehrungen im Hinblick auf die Aufbewahrung braucht, stellt eine einfache Möglichkeit der Lagerung von archivischen Informationen auf kleinstem Platz dar.

Ich möchte mich an dieser Stelle auch herzlich bei den Leitenden und Mitwirkenden des NEO-Projektes bedanken, die mir so bereitwillig und schnell die Möglichkeit gaben, einen genaueren Einblick in das Projekt zu erlangen.

6.2.2. Disco – DNA Data Storage

Das Disco Projekt wurde vom *European Innovation Council Pathfinder Challenge* ins Leben gerufen und setzt sich aus einem Team der Tapdance Research Groupe der Maynooth University, Major Groove und Tilibit nanosystems zusammen. Geleitet wird das Projekt von Dr. Damian Woods von der Maynooth University.¹⁶³

Das Projekt an sich wurde im letzten Jahr ins Leben gerufen und steht auch noch im Bereich der Publikationen am Anfang.¹⁶⁴

¹⁶³ About us, in: Disco tech, <https://disco-tech.eu/partners/> [27.07.2024].

¹⁶⁴ Publications, in: Disco-tech, <https://disco-tech.eu/publications/> [27.07.2024].

6.2.3. Microsoft Research von DNA Data Storage

Microsoft Research publiziert seit dem Jahr 2016 regelmäßig diverse Paper zu DNA Datenspeicherung.

Ins Leben gerufen wurde das Projekt im Jahr 2015. Auch dieses Projekt ist mit dem Wissen entstanden, dass es auf Grund der wachsenden digitalen Datenmenge neue Herangehensweisen zur digitalen Speicherung braucht. Besonders der hohe Platzverbrauch heutiger Speicher digitaler Daten steht in Fokus des Projektes und soll mit einem DNA-Speicher gelöst werden. Gemeinsam mit der University of Washington wird seitdem untersucht, wie DNA als Speicher digitaler Daten verwendet werden kann. So sollen biomolekulare Aspekte gemeinsam mit der Computertechnologie kombiniert werden, um eine effektivere Datenspeicherung zu gewährleisten.¹⁶⁵

Eines der Paper beschäftigt sich mit dem Vergleich von DNA Datenspeicherung und traditionellen technischen Speichern. Es wird das DNA Speichermodell mit dem Open System Interconnection Modell (OSI) verglichen. Dieser Vergleich zeigt auf, dass sich die beiden Modelle nicht grundsätzlich viel unterscheiden, genauer gesagt sind die ersten drei Ebenen der beiden Modelle in ihrer Funktion gleich.¹⁶⁶ In Abbildung 23 werden die beiden Modelle nebeneinander bildlich dargestellt.



Abbildung 22, OSI-Modell und DNA-Datenspeicher-Modell.¹⁶⁷

¹⁶⁵ DNA Storage, in: Microsoft Research, <https://www.microsoft.com/en-us/research/project/dna-storage/>, [14.06.2024].

¹⁶⁶ Dave LANDSMAN, Karin STRAUSS, The DNA Data Storage Model. *IEEE Computer* 56/7 (2023) 78-85, 79.

¹⁶⁷ LANDSMAN, STRAUSS, The DNA Data Storage Model, 79.

Mit dem Paper soll gezeigt werden, dass, auch wenn DNA Datenspeicherung in seiner Form grundsätzlich gänzlich anders zu sein scheint als herkömmliche Speicher, die Arbeit und Verarbeitung der Daten in den einzelnen Schritten sehr ähnlich ist. Mit dem Vergleich soll ein besseres Verständnis für DNA Datenspeicherung hergestellt werden, um somit die Arbeit mit dieser neuen Art des Speichers voranzutreiben.¹⁶⁸

Das Projekt beschäftigt sich auch mit diversen anderen Aspekten der DNA Datenspeicherung. Diese umfassen unter anderem die Speicherung der DNA in abbaubaren Organosilica¹⁶⁹ oder der Speicherung von DNA in magnetischen Nanopartikeln¹⁷⁰. Aber sie beschäftigen sich auch damit, wie ein wiederholter Zugriff auf die Daten ermöglicht werden kann. Der Ansatz, welcher untersucht wird, beschäftigt sich mit der Verkapselung von DNA-Datensätzen, welche in thermoresponsiven Mikrokapseln aufbewahrt werden, deren Membranen halbdurchlässig sind. Dadurch werden molekulare Überschneidungen verringert und die Möglichkeit geschaffen, mehrere DNA-Dateien originalgetreu zu vervielfältigen.¹⁷¹

6.2.4. OLIGOARCHIVE – Synthetische DNS als digitale Speicherlösung der nächsten Generation

Das EU Projekt hat sich zwischen 2019 und 2023 mit DNA als Speicherlösung für zukünftige Speicherung digitaler Daten beschäftigt.¹⁷² Der Name des Projektes beinhaltet den Begriff des Oligo, kurz für Oligonukleotid. Dabei handelt es sich um einen kurzen Strang synthetisch hergestellter DNA oder RNA. Diese können

¹⁶⁸ LANDSMAN, STRAUSS, The DNA Data Storage Model, 84.

¹⁶⁹ Julian KOCH, Ann-Christin KERL, Natascha SCHAWALDER et al., Preserving DNA in Biodegradable Organosilica Encapsulates. *American Chemical Society* (2022), 11191-11198.

¹⁷⁰ Weida D. CHEN, A. Xavier KOHLL, Bichlien H. NGUYEN et al., Combining Data Longevity with Storage Capacity. Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. *Advanced Functional Materials* 29/28 (2019), 1-8.

¹⁷¹ Bas W. BÖGELS, Bichlien H. NGUYEN, David WARD et al., DNA storage in thermoresponsive microcapsules for repeated random multiplexed data access. *Nature Nanotechnology* 18 (2013), 912-921, 912f.

¹⁷² Oligoarchive - Intelligent DNA Storage for Archival, in: Cordis Europe, <https://cordis.europa.eu/project/id/863320> [29.07.2024].

einzelsträngig oder doppelsträngig sein. Der Name kommt von dem griechischen Wort Oligo, was so viel bedeutet wie „klein“ oder „wenige“.¹⁷³

Ziele des Projektes waren unter anderem ein verbesserter Sequenzierungsprozess, mit welchem ein Auslesen der digitalen Daten schneller durchgeführt werden sollte als bisher möglich. Außerdem sollte die Methode der Synthetisierung verbessert werden, um die Herstellung der DNA kostengünstiger zu gestalten. Ein weiteres Ziel war die verbesserte Anpassung der Codierungsmethoden. Je nach Datentyp sollen unterschiedliche Codierungen angewandt werden können, um ein möglichst fehlerfreies Ergebnis der Codierung zu erreichen. Eine weitere Verbesserung soll im Bereich des Auslesens und der Datenanalyse erreicht werden. Durch biomolekulare Techniken sollen die Daten direkt in der DNA analysiert werden.¹⁷⁴

Das Projekt konzentriert sich auf die dritte Ebene der Speicherhierarchie, der Archiv-Ebene. Bei der sogenannten Speicherhierarchie handelt es sich um einen Begriff aus der Informatik. Dabei sollen Datenkategorien unterschiedlichen Speichermedien zugewiesen werden, um eine Senkung der Gesamtspeicherkosten zu erreichen. Zusätzlich werden auch die Leistungen und die Verfügbarkeit verbessert. Die Zuweisung zu den Speichermedien folgt einer Hierarchie, welche auf dem geschäftlichen Wert der Daten basiert. Das bedeutet, dass Daten, welche öfter benutzt werden, einen höheren Wert haben. Daneben zählen auch die Leistung und die Medienkosten der einzelnen Daten. Die Speicherebenen werden aus dem Englischen als Tiers benannt.¹⁷⁵ Daten der Archivebene sind jene Daten, welche selten in Gebrauch sind. Dadurch werden diese Daten auch auf Medien gespeichert, welche keine hohen Ansprüchen im Bereich der Leistung erbringen müssen. Diese Daten sind hauptsächlich auf Bändern gespeichert, denn sie sind vergleichsweise günstiger und länger haltbar als SSD (State Storage Device) oder HDD (Hard Disk Drives). Mit dem Oligoarchive sollen diese Bänder durch DNA als Speichermedium ersetzt werden.¹⁷⁶

¹⁷³ Jeff SCHMERKER, What is an Oligo?, in: Integrated DNA Technologies, <https://eu.idtdna.com/pages/community/blog/post/what-is-an-oligo#:~:text=What%20are%20oligo%20drugs?%20Oligo%20drugs,%20or%20oligonucleotide%20therapeutics,%20can> [06.08.2024].

¹⁷⁴ Forschungsergebnis, in: Oligoarchive, <https://cordis.europa.eu/project/id/863320/results/de>, [29.07.2024].

¹⁷⁵ Tiered Storage, in: Computer Weekly, <https://www.computerweekly.com/de/definition/Tiered-Storage-Abgestuftes-Storage#:~:text=Tiered%20Storage%20stellt%20Speicherebenen%20zur%20Verf%C3%BCgung,%20auf%20denen%20Daten%20einer> [05.08.2024].

¹⁷⁶ Raja APPUSWAMY, Kevin LEBROGAND, Pascal BARBRY et al., OligoArchive. Using DNA in the DBMS storage hierarchy, 1-7, 2.

Vor dem Schritt der Codierung der digitalen Daten werden diese aus der Datenbank extrahiert. Unter einer Datenbank wird die Organisationsstruktur verstanden, mit welcher digitale Daten in Computersystemen gespeichert werden. Diese Datenbanken werden durch ein DBMS, ein Datenbankmanagementsystem organisiert und kontrolliert.¹⁷⁷ Durch die Komprimierung der Daten, welche nach der Extrahierung aus der Datenbank stattfindet, wird es möglich, mehrere Records in einem Oligo zu speichern. Im nächsten Schritt werden die digitalen Daten mit Hilfe des Huffman Codes codiert. Durch die Verwendung des Huffman Codes sollen Homopolymere und eine hohe Anzahl von G und C-Basen vermieden werden. Bisherige Projekte mussten für das Wiederauffinden der einzelnen Daten dazugehörige Metadaten oder „addressing information“¹⁷⁸ auf die einzelnen Oligos kodieren. Das Projekt Oligoarchive hat hier eine andere Herangehensweise gesucht, das sogenannte „schema-aware encoding“¹⁷⁹. Genauer gesagt werden hier die Daten so kodiert, dass jeder Record auf einem Oligo abgebildet wird. Dadurch, dass jede Tabelle einer relationalen Datenbank eigene Primärschlüssel besitzt, können diese auch eindeutig durch diese identifiziert werden. Ein Primärschlüssel ist eine Spalte innerhalb einer relationalen Datenbank und ist dafür zuständig, die einzelnen Records dieser Datenbank einmalig darzustellen. Als verständliches Beispiel kann die Studienkennzahl der Studierenden einer Universität genommen werden. Die Kennzahl würde als Primärschlüssel dienen, mit der jeder einzelne Student/ jede einzelne Studentin unverkennbar identifiziert werden kann. Durch den Primärschlüssel einer Datenbank können die einzelnen Records derselben schnell und einfach gefunden werden, auch wenn diese untereinander eventuelle Überschneidungen der Daten aufzeigen würden.¹⁸⁰ Durch diese Methode ist es nicht mehr nötig, die zusätzlichen Informationen zum Auffinden der einzelnen Daten ebenfalls zu kodieren. Für Records, welche für einzelne Oligos zu groß sind, wird ein anderes Schema der Kodierung angewandt.¹⁸¹ Anfänglich werden die einzelnen Datenbanken, welche auf DNA gespeichert werden sollen, analysiert und bekommen

¹⁷⁷ What is a database?, in: Oracle, <https://www.oracle.com/database/what-is-database/#:~:text=A%20database%20is%20an%20organized%20collection%20of%20structured%20information,%20or> [06.08.2024].

¹⁷⁸ APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 3.

¹⁷⁹ APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 3.

¹⁸⁰ Tim STATLER, What is a Primary Key?, in: Computer Science Central, [https://compscicentral.com/what-is-a-database-primary-key/#:~:text=A%20primary%20key%20is%20a%20column%20\(s\)%20within%20a%20relational](https://compscicentral.com/what-is-a-database-primary-key/#:~:text=A%20primary%20key%20is%20a%20column%20(s)%20within%20a%20relational) [06.08.2024].

¹⁸¹ APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 3.

jeweils eine ID (Identifikator) zugeteilt, welche auf jedes Oligo mit Daten aus der Datenbank kodiert wird, um eine Zuordnung ermöglichen zu können. Daten, welche zu groß sind, als dass sie auf einem Oligo gespeichert werden können, werden vertikal fragmentiert und die einzelnen Spaltensets werden auf den Oligos gespeichert. Um die Daten bei der Decodierung wieder zusammensetzen zu können, werden den einzelnen Spaltensets die Primärschlüssel vorangesetzt.¹⁸² Auch Oligoarchive arbeitet mit Error Correction, wie schon in vorherigen Kapiteln erklärt und bei anderen Projekten vorgestellt, um einem Datenverlust entgegenzuwirken.¹⁸³

Die Decodierung des Oligoarchives beginnt mit der Gruppierung der Oligos, welche die gleiche ID für die einzelnen Datenbanken aufweist. Danach wird die Decodierung wie schon im Kapitel 5.2.5. erklärt, durchgeführt.¹⁸⁴

6.3. Arbeitskonzept zur Speicherung von GIS-Daten auf DNA

Der letzte Abschnitt dieses Kapitels soll ein theoretisches Konzept beinhalten, welches als ein Workflow für Archive konzipiert ist, die sich mit der Implementierung eines DNA Datenspeichers auseinandersetzen wollen. Für die anschaulichere Darstellung des Workflows werden die Daten der GIS (Geo Informationssystem) Steiermark als praktisches Beispiel verwendet, um bezüglich der Kosten und der Implementierung eines solchen Speichersystems ein greifbareres Bild zu gestalten.

Das theoretische Konzept soll schematisch darstellen, wie es mit dem heutigen Stand der Forschung möglich wäre, DNA Datenspeicherung in einem Archiv anzuwenden. Es soll vor allem ein Fokus daraufgelegt werden, was ein Archiv für die Ausführung und Anwendung dieser Speichermethode braucht, wie sich der Arbeitsumfang für das Archiv gestalten würde und ob und welche zusätzlichen Anschaffungen das Archiv tätigen muss.

¹⁸² APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 3.

¹⁸³ APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 4.

¹⁸⁴ APPUSWAMY, LEBROGAND, BARBRY et al., OligoArchive, 4.

Wie eingangs schon erwähnt, sollen hierfür die GIS Daten verwendet werden. Diese Daten werden von Landesdienststellen, von Bundeseinrichtungen, Gemeinden und Privatpersonen generiert und umfassen neben anderen Daten Geodaten, Orthofotos, Geländedarstellungen und Kataster.¹⁸⁵

Die Daten werden digital in einer Datenbank gespeichert. Auf diese Daten kann jederzeit zugegriffen werden, um mit diesen zu arbeiten. Die zu bearbeitenden Daten umfassen rund 750 Terabyte. Dabei handelt es sich unter anderem um Geo-GIS Daten, Exceldateien, Orthophots und Laserdaten. Die Daten werden nach der Lieferung als Originaldaten gespeichert, aber es werden auch Ergebnisse der Berechnungen mit ebendiesen Daten aufgehoben. Hin und wieder können auch aufwendigere Zwischenergebnisse einzelner Berechnungen gespeichert werden, um eine Vereinfachung der Datenbearbeitung zu ermöglichen. Neben dem digitalen Speicher in der Datenbank wird ein Back-Up für die Daten angefertigt, welches auf Magnetbänder gespeichert wird. Bei diesem Back-Up handelt es sich hauptsächlich um die Originaldaten, die wichtigen und aufwendigeren Zwischenergebnisse und die Endergebnisse der Berechnungen.

Dieses Arbeitskonzept soll sich auf eben diese Back-Up-Speicherungen konzentrieren. Da auf die Daten in der Datenbank regelmäßig zur Bearbeitung zugegriffen wird, erscheint es sinnvoller, ein Konzept zur Speicherung der Back-up-Daten auf DNA zu erarbeiten. Einerseits müssen diese Daten des Back-Ups einer regelmäßigen Datenmigration unterzogen werden, um einem Datenverlust vorzubeugen. Dabei entsteht zum jetzigen Zeitpunkt unter anderem auch regelmäßiger Elektroschrott. Andererseits befinden sich die Daten in der Datenbank in regelmäßiger Bearbeitung, wodurch ein regelmäßiger Zugriff für mehrere Personen auf die DNA ermöglicht werden müsste. Dies ist unter heutigen Bedingungen noch nicht problemlos möglich. Bei einer Speicherung der Back-Up-Daten auf DNA würde man einer regelmäßigen Migration der zu speichernden Daten entgehen. Dadurch würde man auch die immer wieder auftretenden Kosten der Migration einsparen und gleichzeitig den laufend entstehenden Elektroschrott verringern oder vermeiden. Zusätzlich würde es zu einer drastischen Platzreduktion durch die wesentlich kleineren Datenträger kommen.

¹⁸⁵ GIS Steiermark, in: Landesentwicklung Steiermark, <https://www.landesentwicklung.steiermark.at/cms/ziel/141976122/DE/>, [05.08.2024].

Da die Vorteile der DNA Datenspeicherung vor allem im Bereich der Speicherung großer digitaler Datenmengen liegen, konzentriert sich das Konzept auf eben solche Daten. Ein weiterer Vorteil für die Verwendung der GIS-Daten liegt darin, dass für den Erhalt zum jetzigen Stand eine regelmäßige Migration der Daten vorgenommen werden muss, um keine Daten zu verlieren. Durch das Speichern der Daten auf DNA wird, durch das Wegfallen der Migration, ein regelmäßiger Kostenfaktor eingespart. Weiters handelt es sich bei den GIS-Daten um einen Back-Up Speicher, auf den in Ausnahmefällen für die Wiederherstellung von Daten zugegriffen werden kann. Da auf DNA gespeicherte Daten keine regelmäßigen Kontrollen benötigen, wenn sie unter den richtigen Bedingungen gelagert werden, kann dieser Back-Up Speicher ohne viel Aufwand und zusätzliche Kosten betrieben werden.

GIS wird grundsätzlich als „Informationssystem für raumbezogene Daten bezeichnet“¹⁸⁶. Darunter fallen vier Kompetenzbereiche, welche im GIS möglich sein sollen. Einerseits die Erfassung und Verwaltung der Daten und andererseits die Analyse und die Ausgabe derselben Daten. Innerhalb des Systems können auch neue Daten und Informationen geschaffen werden, diese Bearbeitung wird in Schichten oder Layern dargestellt. Diese können miteinander kombiniert werden, um bestehende Zusammenhänge sichtbar zu machen oder neue Daten herzustellen.¹⁸⁷

Der Workflow muss sich im ersten Schritt damit beschäftigen, welche Daten für die Speicherung in Frage kommen und welche Ziele mit dieser Speicherung erreicht werden sollen. In weiterer Folge muss auch die Frage geklärt werden, was das Archiv, beziehungsweise die aufzubewahrende Stelle, für die Umsetzung der Speicherung der ausgewählten Daten braucht. Darunter fällt die Herstellung der DNA, die Codierung und Speicherung der Daten an sich, die Lagerung der DNA-Samples und in weiterer Folge auch die Decodierung und das Auslesen der Daten.

Die erste Frage, die sich bei diesem Schritt stellt, ist die Frage, welche Daten für dieses Projekt auf DNA gespeichert werden sollen. Um welchen Umfang handelt es sich hierbei und von welcher Beschaffenheit sind diese Daten.

Die zweite Frage sollte sich damit befassen, was durch den DNA-Speicher erreicht werden soll. Bei der Speicherung von GIS-Daten auf DNA soll es sich um einen Back-Up Speicher handeln, wie in der Einführung oben schon erwähnt. Das

¹⁸⁶ KERSCHNER, Aspekte der Archivierung digitaler Geodaten und kartographischer Darstellungsformen, 25.

¹⁸⁷ KERSCHNER, Aspekte der Archivierung digitaler Geodaten und kartographischer Darstellungsformen, 26.

bedeutet, dass es keinen täglichen Zugriff auf die Daten bedarf und dass es auch zu keiner regelmäßigen Bearbeitung dieser Daten kommt. Zur Bearbeitung und Weiterverarbeitung der Daten werden diese aus dem aktiven Speicher gezogen. Nur eventuell bedeutende Zwischenergebnisse und Endergebnisse werden wieder im Back-Up Speicher und somit auch auf DNA gespeichert und aufbewahrt.

Der DNA-Speicher soll also die Speicherung auf Magnetbänder ersetzen und hier den schon in der Arbeit erwähnten Nachteilen dieser Speichermethode entgegenwirken. Die Kosten werden im Bereich der Lagerung verringert, da auch der Lagerplatz der Daten verkleinert wird, aber auch in Bezug der Datenmigration. Es wird mehr Platz für zukünftige Daten ermöglicht und die Erhaltung der Daten vergünstigt.

Nach der Etablierung des Umfanges der Daten, deren Beschaffenheit und der Ziele und des Nutzens für den DNA-Speicher, stellt sich die Frage, was das Archiv oder die aufzubewahrende Stelle benötigt, um eine Speicherung auf DNA zu ermöglichen oder durchführen zu können. Bei dem momentanen Forschungsstand bietet sich die Zusammenarbeit mit einer Fremdfirma an, welche sich um die Codierung der digitalen Daten und Synthetisierung der DNA-Samples kümmert und dem Archiv im Endeffekt das Endprodukt der gespeicherten Daten auf DNA zur Lagerung und Aufbewahrung übergibt. Beziehungsweise kann die Fremdfirma auch die auf DNA gespeicherten Daten zur Aufbewahrung in einem Speicher vor Ort behalten und bei Bedarf dem Archiv oder der aufzubewahrenden Stelle direkt die digitalen Daten zukommen lassen. Dabei ergibt sich für das Archiv zwar ein finanzieller Aufwand für die Entschädigung der Fremdfirma. Allerdings würde sich die Codierung und Herstellung der DNA als wesentlich aufwendiger in zeitlicher und in finanzieller Hinsicht für das Archiv gestalten.

In einem zweiten Schritt muss das theoretische Konzept ausgearbeitet werden, basierend auf den theoretischen Überlegungen des ersten Schrittes. In diesem werden die einzelnen Informationen der Überlegungen in die zu tätigenen Arbeitsschritte eingearbeitet. In einem ersten Schritt muss eine passende Fremdfirma als Kooperationsfirma gefunden werden, welche deren Produkte und Angebote auf die Bedürfnisse des Archives abstimmt. Ebenso sollte es der Firma möglich sein, große Datenmenge auf DNA speichern zu können und diese in einen Lagerungszustand zu bringen, welcher für das Archiv eine vereinfachte

Aufbewahrung ermöglicht, wenn die Archivbetreiber es bevorzugen, die Daten in ihrem eigenen Depot aufzubewahren. Ist dem der Fall, so müssen hier Überlegungen angestellt werden, wie das Archiv die DNA-Samples räumlich aufbewahren kann. Da die klimatischen Bedingungen für die meisten Ausführungen der DNA-Samples wesentlich andere sind als für herkömmliches Archivgut, muss mit großer Wahrscheinlichkeit ein Raum dafür geschaffen werden, in dem die klimatischen Bedingungen von ca. minus 18 Grad hergestellt werden können. Allerdings ist es auch möglich, die DNA auf Siliciumplatten zu lagern, welche keine besonderen klimatischen Voraussetzungen brauchen. Diese können einfach in den Depotregalen gelagert werden. Zusätzlich braucht das Archiv eine Einführung der Lagerung und Handhabung des neuen Speicherträgers für seine Mitarbeiterinnen und Mitarbeiter. Da es sich um eine neue Technologie handelt, sollte ein kurzer Überblick darüber gegeben werden, wie die Speicherung funktioniert, wie die Handhabung der DNA-Samples bevorzugt ablaufen sollte und welche Besonderheiten die neuen Träger aufweisen. Wenn das Archiv seine DNA-Samples in seinem eigenen Depot lagern will, stellt sich weiterführend die Frage, ob das Archiv seine Daten komplett selber auslesen möchte, oder ob es sich hier wieder in eine Zusammenarbeit mit der Fremdfirma begeben möchte. Für die eigene Auslese der Daten braucht das Archiv weitere Schulungen für die Mitarbeiterinnen und Mitarbeiter. Ebenso müssen hierfür Geräte für diesen Prozess angeschafft werden. Sollte sich das Archiv dazu entscheiden, die Daten bei der Fremdfirma zu lagern, so muss sich das Archiv für eine Firma entscheiden, welche diese Möglichkeit auch anbietet.

Der dritte Schritt des Workflows umfasst die praktische Umsetzung der theoretischen Überlegungen des Arbeitskonzeptes. Dafür braucht es eine Kooperation mit einer Fremdfirma, welche die Codierung der digitalen Daten und die Synthetisierung der DNA für das Archiv oder die aufzubewahrende Stelle übernimmt. Die Fremdfirma benötigt die genauen Rahmenbedingungen der Daten, unter anderem den genauen Umfang der Daten und die Art der Speicherformate derselben Daten. Nach der Übermittlung der Daten an die Fremdfirma, kümmert sich diese um die Codierung, Synthetisierung, Sequenzierung und schlussendlich auch um die Decodierung der Daten. Für das Archiv ergeben sich somit in diesem Bereich keine zusätzlichen Anschaffungen. Die Archivarinnen und Archive brauchen in diesem Bereich keine zusätzlichen Ausbildungen oder fachspezifisches Wissen im Bereich der DNA-Synthetisierung oder Codierung digitaler Daten. Dennoch ist eine Schulung bezüglich

eines allgemeinen Überblicks über die Funktionsweise der einzelnen Arbeitsschritte und der Aufbewahrung und Handhabung eine sinnvolle Überlegung für das Archiv, um einen vereinfachten und verbesserten Umgang mit dem neuen Speicherformat zu ermöglichen. Das Archiv und dessen Angestellte brauchen allerdings Zusatzinformationen für die Aufbewahrung und die richtige Lagerung der DNA-Samples, falls das Archiv die Daten bei sich vor Ort aufbewahren muss oder will. Die Art der Lagerung hängt unter anderem auch mit der Art der Aufbewahrung der DNA-Samples zusammen. In Kapitel 5 wurden unterschiedliche Methoden der Aufbewahrung von DNA-Samples angeführt. Einerseits können die DNA-Samples in Siliziumdioxid gelagert werden, oder in sogenannten Silica-Nanobeads. Diese können bei minus 18 Grad Celsius gelagert werden. Hierfür braucht das Archiv Räumlichkeiten in denen diese Gegebenheiten ermöglicht werden können. Die Kosten, welche für das Archiv bezüglich der Herstellung und Speicherung der Daten auf DNA entstehen, sind ebenfalls von der spezifischen Fremdfirma abhängig. Wenn man mit den Kosten von 1 Million Dollar pro GB rechnet, so würden sich für die GIS-Daten, welche einen Umfang von 750 TB haben, Kosten von 750 Milliarden Dollar für die Speicherung des gesamten Speichers auf DNA ergeben. Auch wenn es sich hierbei um eine einmalige Ausgabe handelt, so muss dennoch die Überlegung angestellt werden, ob die digitalen Daten schrittweise auf DNA und in welcher Reihenfolge sie gespeichert werden sollen.

Bei einer Lagerung in Flüssigkeit, bevorzugterweise in einem TE-Puffer, können die DNA-Samples bei unterschiedlichen Temperaturen gelagert werden. Einerseits können die Puffer bei Raumtemperaturen aufbewahrt werden¹⁸⁸, andererseits können die Lösungen auch bei bis zu minus 20 Grad gelagert werden¹⁸⁹. Die DNA-Samples können auch in festen, physischen Behältern übergeben werden, beispielsweise in Glas, Edelstahl oder alkalischen Salzen. Die Temperatur der Lagerung hängt schlussendlich von der jeweiligem Fremdfirma ab und wie diese dem Archiv die DNA-Samples für die Lagerung übergibt.

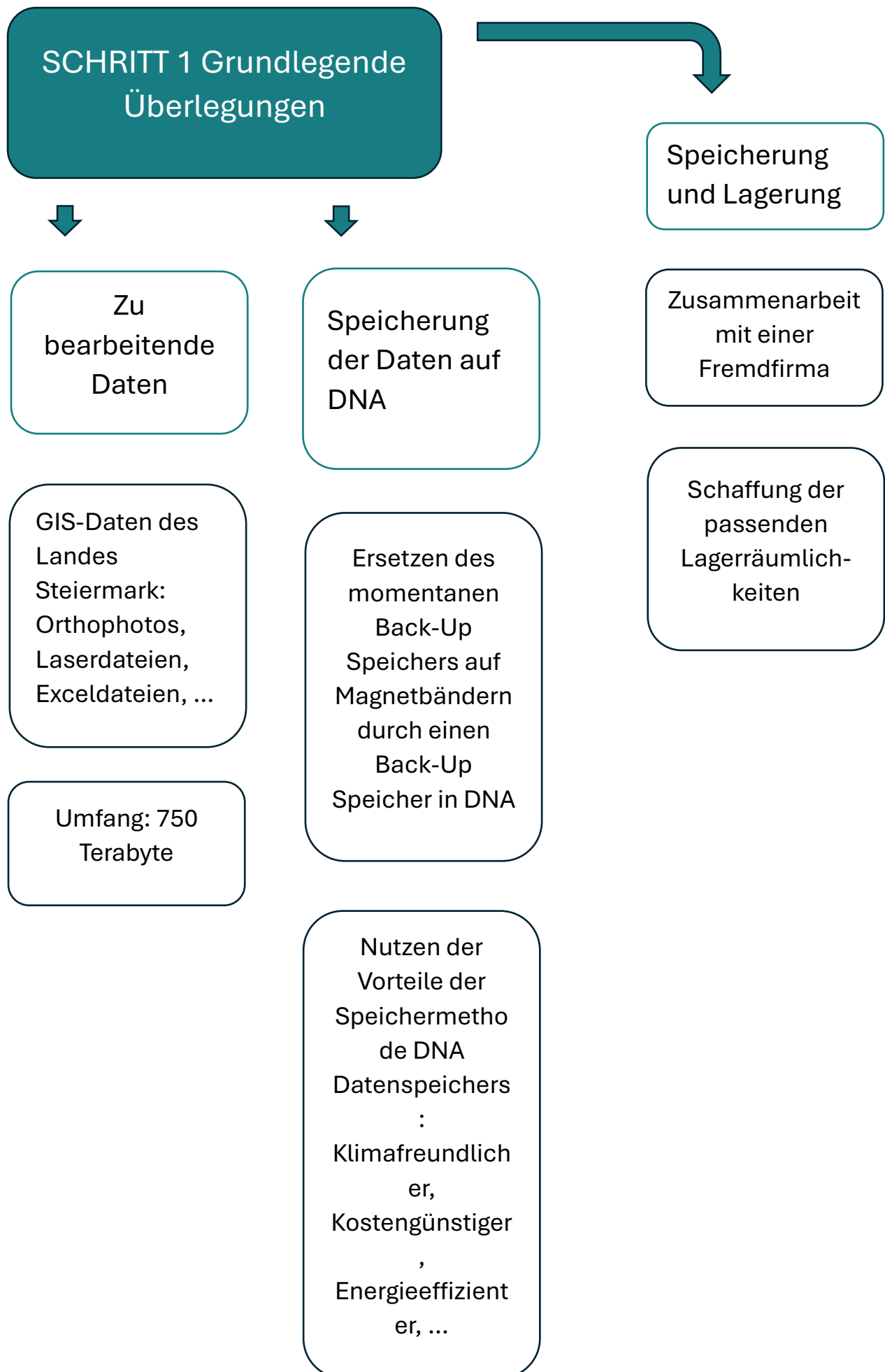
¹⁸⁸ TE-Puffer, in: SERVA,

https://www.serva.de/deDE/ProductDetails/949_544_42554_TE_Puffer_100x_16_544.html#:~:text=Tris-EDTA-Puffer%20ist%20ein%20h%C3%A4ufig%20verwendeter%20Molekularbiologiepuffer%20zum%20Resuspendieren%20und/oder%20Verd%C3%BCnnen [18.08.2024].

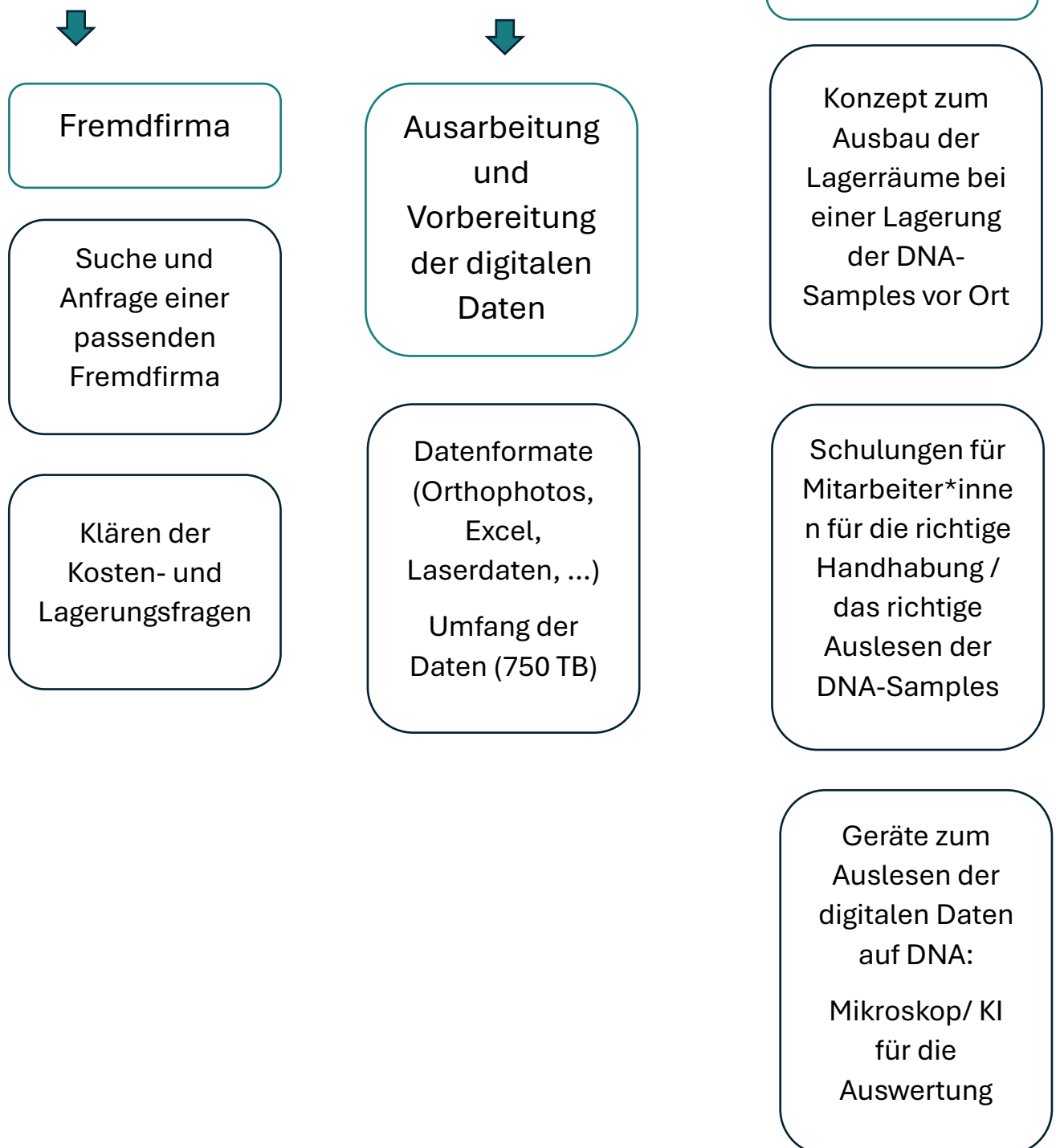
¹⁸⁹ DNA- & RNA-Aufreinigung, Oligonukleotidhandhabung und -stabilität, in: Merck,

<https://www.sigmaaldrich.com/AT/de/technical-documents/technical-article/genomics/dna-and-rna-purification/sample-purification-and-quality-assessment> [18.08.2024].

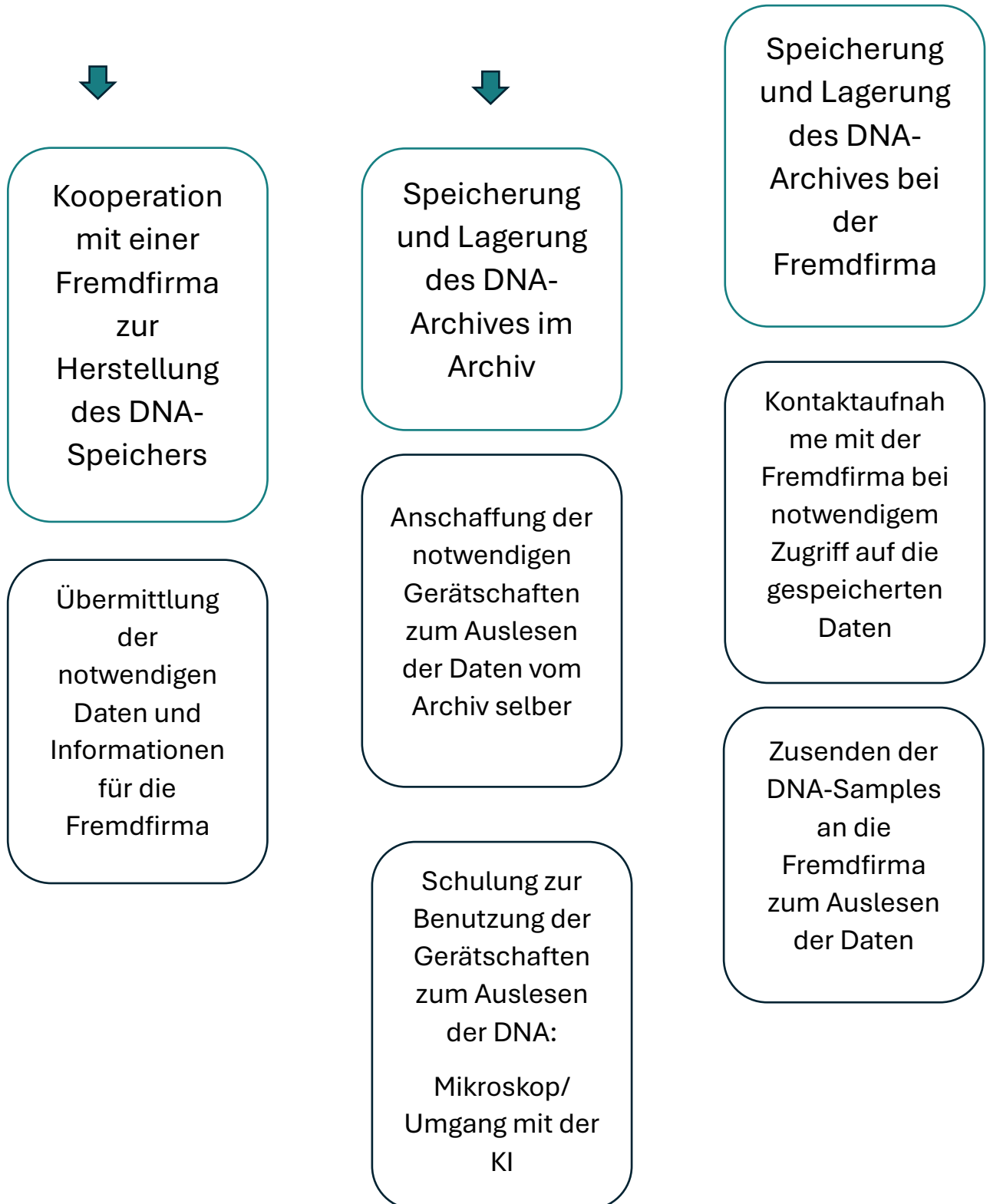
Zum jetzigen Zeitpunkt sind die meisten Firmen oder Projekte noch nicht weit genug fortgeschritten, um aktiv größere Datenmengen auf DNA zu speichern. Daher handelt es sich bei dieser Ausarbeitung und den Überlegungen um ein theoretisches Konzept, mit dem eventuell in Zukunft eine praktische Umsetzung möglich sein wird.



SCHRITT 2 Theoretische Ausarbeitung und Vorbereitungen für die praktische Umsetzung



SCHRITT 3 Praktische Umsetzungen



7. Fazit

Die Masse an digitalen Daten befindet sich in einem stetigen Wachstum. Die Daten werden jährlich mehr und in weiterer Folge nimmt damit auch die Anzahl der zu archivierenden digitalen Daten zu, nicht nur im Bereich der Behörden und Verwaltung auf staatlicher oder privater Ebene, sondern auch im Bereich von sozialen Netzwerken, YouTube und Webseiten, welche in Zukunft möglicherweise archiviert werden müssen.

Der momentane Stand der digitalen Archivierung basiert auf der Verwendung von Speicherträgern, welche in regelmäßigen Abständen einer Datenmigration unterzogen werden, damit die darauf gespeicherten Daten nicht an Bit-Rot oder Ähnlichem verloren gehen. Zusätzlich wird der Platz aufgrund der zunehmenden Datenmenge immer geringer und die Kosten für die digitale Archivierung steigen im Laufe der Lagerung und der Migrationen immer weiter an.

Daher werden immer wieder neue Methoden und Ansätze der digitalen Archivierung ausgearbeitet und getestet. Diese neuen Ansätze werden unter dem Begriff der Optical Storages, oder auch der Write Once, Read Forever-Methoden, zusammengefasst. Darunter fällt nicht nur die DNA-Datenspeicherung, sondern unter anderem auch die Speicherung auf Papier, bei dem die digitalen Daten in geringer Größe auf A4-Seiten archiviert werden, das Eingravieren auf Keramikplatten und der Lagerung derselben in einem Salzbergwerk oder das Projekt Microsoft Silica, welches Daten auf Silica-Platten eingraviert. All diese Methoden haben ähnliche Vorteile. Sie werden nur einmal geschrieben und benötigen danach keine weitere Datenmigration mehr, um die zu archivierenden Daten zu erhalten. Die Speicherung und Aufbewahrung der Daten verbraucht im Vergleich zu digitalen Speichern weniger Strom. Dadurch können diese Speichermethoden als klimafreundlicher eingestuft werden und benötigen im Gegensatz zu den herkömmlichen digitalen Speichern keine regelmäßige Datenmigration.

Der Vorteil, der bei einer DNA Datenspeicherung ganz besonders hervorzuheben ist, ist die Tatsache, dass bei dieser Methode auf besonders wenig Platz eine große Menge an Daten gespeichert werden kann. Zusätzlich dazu besitzt DNA eine große Haltbarkeit, vor allem wenn sie unter den richtigen oder passenden Bedingungen

gelagert und aufgehoben wird. Dies macht DNA zu einem besonders geeigneten Speicher für eine immer größer werdende Datenmenge digitaler Daten.

Für DNA Datenspeicherung werden ausschließlich synthetisch hergestellte DNA-Samples verwendet, welche den binären Code der digitalen Daten darstellen können. Mit Hilfe von Codierungsmethoden kann der binäre Code der digitalen Daten in die vier Basen der DNA umgewandelt werden. Diese werden dann passend synthetisiert und stellen jeweils einen spezifischen Datensatz dar. Für das Auslesen der Daten muss die gleiche Codierungsmethode in umgedrehter Anwendung verwendet werden, damit aus den Basen wieder der Binäre Code hergestellt werden kann. Aufbewahrt werden die DNA-Samples entweder in einer chemischen oder in einer physischen Kapselung. Bei den chemischen Lösungen handelt es sich hauptsächlich um sogenannte TE-Puffer, in welchen die DNA in Röhrchen aufbewahrt wird. Die physische Kapselung kann unterschiedliche Formen annehmen, beispielsweise Silica Nanobeads oder in Glas, Edelstahl oder alkalischen Salzen. Die Lagerung der unterschiedlichen Aufbewahrungsformen fällt unterschiedlich aus, auch wenn es sich dabei nur um minimale Unterschiede in den Temperaturen handelt.

Neben den Vorteilen der Methode gibt es allerdings auch Nachteile. Einerseits ist der gesamte Prozess der Synthetisierung und Sequenzierung der Daten vergleichsweise teuer. Mit dem Stand aus dem Jahr 2017 liegen die Kosten der Herstellung und der Speicherung von digitalen Daten auf DNA bei 3.500 Dollar pro Megabyte. Laut Mitarbeitern des NEO Projektes kostet die Herstellung von DNA immer noch 1 Million Dollar pro Gigabyte. Andererseits entstehen bei der Synthetisierung immer wieder Fehler, wodurch eine gewisse Menge an DNA zusätzlich hergestellt werden muss, damit keine der Daten bei dem Prozess verloren gehen. Auch im Bereich des Zugriffs auf die DNA im Speicher gibt es unterschiedliche Ansätze, aber auch Probleme, welche zu einem vermehrten Kostenfaktor führen können.

Allerdings beschäftigt sich die Forschung im Bereich der theoretischen Ausarbeitungen, aber auch im praktischen Ansatz laufend damit, neue Ansätze zur Verbesserung der DNA Datenspeicherung zu finden. Es wird an dem Auslesen der Daten aus dem DNA Pool der zu speichernden Daten gearbeitet. Hier wird vor allem am Bereich des zufälligen Zugriffs und anderen Möglichkeiten derselben gearbeitet. In der Arbeit werden vier unterschiedliche theoretische Ansätze vorgestellt. Jeder dieser Ansätze konzentriert sich auf unterschiedliche Aspekte der DNA

Datenspeicherung. Das Team von Bornholt konzentriert sich vorrangig auf die Verwendung von PCR-Primern für das Auslesen spezifischer Datensätze. Goldman und sein Team testeten den In Vitro Ansatz und konnten theoretisch darlegen, dass eine Speicherung von Daten bis in den Petabyte-Bereich mit ihrer Methode möglich ist. Das Team von Stefano arbeitet mit der Theorie zur Verwendung eines Chips zur Speicherung der DNA, welcher schlussendlich auch mit Handys oder Laptops kompatibel sein soll.

Die Ansätze der praktischen Ausarbeitungen der DNA Datenspeicherung konzentrieren sich auf unterschiedliche Aspekte. Microsoft versucht mit seinem Projekt einerseits ein verbessertes Verständnis für DNA Datenspeicherung zu schaffen und zieht Vergleiche zwischen diesem Speichermodell und herkömmlichen digitalen Speichern und deren Aufbau. Andererseits fokussiert sich ein Teil der Forschungsarbeit auf die verbesserte Verkapselung von DNA und damit einhergehend auf einen erleichterten und schnelleren Zugriff auf die Daten. Das NEO-Projekt konzentriert sich auf die Arbeit mit Nanotechnologie, bei dem die DNA in einer Origamifalttechnik aufbewahrt wird, auf der einzelne Enzyme helfen, den binären Code darzustellen. Die Daten werden schließlich aus einem Mikroskopbild von einer KI ausgelesen und wieder zu einem binären Code umgewandelt. Die geschriebenen Daten können mit diesem Ansatz schneller und leichter verändert werden, weswegen sich das Projekt vor allem auf Daten konzentriert, welche noch im laufenden Prozess der Bearbeitung verändert werden können und müssen. Das Projekt OligoArchive beschäftigt sich mit einem verbesserten Mechanismus der Auslese der Daten. Außerdem konzentriert sich das Projekt mit der dritten Ebene der digitalen Speicherhierarchie, der Archivebene.

Das Konzept der DNA Datenspeicherung ist für den Bereich der Archive noch wenig bis kaum ausgearbeitet. Der Großteil der Projekte beschäftigt sich mit der Speicherung von kleinen Datenmengen auf DNA zum Austesten der unterschiedlichen Ansätze und theoretischen Überlegungen. Die größeren Projekte, bei denen mehrere Universitäten und Firmen beteiligt sind, beschäftigen sich ebenfalls mit kleineren Datenmengen oder konzentrieren sich vermehrt auf Firmen und deren Datenspeicherung als auf Archive und deren digitale Datenspeicherung.

Daher beschäftigt sich das letzte Kapitel meiner Arbeit vor dem Fazit mit einem theoretischen Workflow, beziehungsweise einem theoretischen Arbeitskonzept, bei

dem anhand der in der Arbeit erläuterten Informationen dargelegt werden soll, welche Schritte ein Archiv setzen muss, damit eine Speicherung digitaler Daten auf DNA durchgeführt werden kann. Für ein besseres Verständnis wurde dieses Konzept auf Basis der GIS Daten der Steiermark ausgearbeitet. Hierbei handelt es sich um große Datenmengen, welche in einem Back-Up auf Magnetbändern aufbewahrt werden. Diese müssen in regelmäßigen Abständen einer Datenmigration unterzogen werden, um einem Verlust von Daten vorzubeugen. Daher bieten sich solche Datensätze perfekt an, um als Back-Up in einem DNA Speicher aufbewahrt zu werden. Die Speicherung der Daten erfolgt in dem Konzept mit einer Fremdfirma, welche sich um die Synthetisierung der DNA und die Codierung des binären Codes befasst. Die Lagerung der Daten kann entweder ebenfalls durch die Fremdfirma durchgeführt werden oder aber vom Archiv übernommen werden. Sollte das letztere der Fall sein, so muss das Archiv die passenden Bedingungen für eine Lagerung der DNA-Samples schaffen. Das Auslesen der Daten erfolgt ebenfalls wieder über die Fremdfirma.

Das Archiv braucht für dieses Szenario kaum neue Anschaffungen oder zusätzliche Ausbildungen tätigen. Zum jetzigen Zeitpunkt sind allerdings die technischen Voraussetzungen, solche Datenmengen preiswert auf DNA zu speichern noch nicht gegeben. Für das Archiv würde sich auf längere Zeit eine Kostenersparnis ergeben, ebenso wie eine sofortige Platzersparnis.

Abschließend lässt sich sagen, dass ein Archiv mit DNA gespeicherten Daten noch in der Zukunft liegt. Es werden laufend neue Projekte gestartet, welche sich mit der praktischen Umsetzung eines solchen Speichermediums beschäftigen. Die Entwicklung der Methode wird vorangetrieben, um einen Weg zu finden, mit dem DNA Datenspeicherung effektiv und kostengünstig in einem breiteren Feld angewandt werden kann. Auch wenn es für Archive nicht die vorrangige Lösung der digitalen Datenspeicherung sein wird, so ist es doch interessant und wichtig, die Weiterentwicklung der digitalen Datenspeicherung und der DNA Datenspeicherung weiterhin zu verfolgen, um auf etwaige Veränderungen eingehen zu können.

8. Literaturverzeichnis

Daniel ABAWI, Oliver GREINER, E-Mails. Archivierung-Not und Tugend.
Wirtschaftsinformatik und Management (Internet) 3/1 (Wiesbaden 2011) 56-63.

Reinhard ALTENHÖNER, Markus BRANTL, Klaus CEYNOWA, Digitale
Langzeitarchivierung in Deutschland. Projekte und Perspektiven. *Zeitschrift für
Bibliothekswesen und Bibliographien* 58/ 3-4 (2011) 184-196.

Philipp ANTOWIAK, Jory LIETARD, Mohammad DARESTANI et al., Low cost DNA
data storage using photolithographic synthesis and advanced information
reconstruction and error correction. *Nature Communications* (2020) 1-10.

Raja APPUSWAMY, Kevin LEBROGAND, Pascal BARBRY et al., OglioArchive.
Using DNA in the DBMS storage hierarchy, 1-7.

Über Uns, in: Archium, <https://www.archium.org/digital-definition.html> [19.08.2024].

Was ist Analogisierung?, in: Archium,
<https://www.archium.org/analogisierung.html#sec-3ac7> [19.08.2024].

Archivdienstleistungen, in: Archium, <https://www.archium.org/archivdienstleister.html>
[19.08.2024].

Krisensicheres Backup, in: Archium, [https://www.archium.org/mikrofilm-
microsheets.html](https://www.archium.org/mikrofilm-microsheets.html) [19.08.2024].

Microsheets und Mikrofilm, in: Archium, [https://www.archium.org/mikrofilm-
microsheets.html](https://www.archium.org/mikrofilm-microsheets.html) [19.08.2024].

Archivdatenbank, in: Archium, <https://www.archium.org/archivdatenbank.html>
[19.08.2024].

Archivierung auf Papier, in: Archium, <https://archium.org/mikrofilm-microsheets.html###>, [11.12.2024].

ASCII-Code und ASCII-Tabellen, in: heise online, <https://www.heise.de/tipps-tricks/ASCII-Code-und-ASCII-Tabellen-4558152.html#:~:text=Wir%20zeigen%20Ihnen%20im%20folgenden,%20was%20es%20mit%20diesem%20Code> [26.01.2024].

James BANAL, Tyson SHEPHERD, Joseph BERLEANT et al., Random access DNA memory using Boolean search in an archival file storage system. *Nature Materials* 20 (2021) 1272-1280.

Leander, BEIL, Gigantische Datenmengen auf DNA, in: ARD Alpha, <https://www.ardalpha.de/wissen/speichermedien-der-zukunft-dna-speicher-daten-100.html> [22.11.2024].

Tobias BEINERT, Armin STRAUBE, Aktuelle Herausforderungen der digitale Langzeitarchivierung, in: Was bleibt?. Nachhaltigkeit der Kultur in der digitalen Welt, hg von Paul KLIMPEL - Jürgen KEIPER (Berlin 2013) 27-46.

Bas W. BÖGELS, Bichlien H. NGUYEN, David WARD et al., DNA storage in thermoresponsive microcapsules for repeated random multiplexed data access. *Nature Nanotechnology* 18 (2013) 912-921.

James BORNHOLT, Randolph LOPEZ, Douglas M. CARMEAN et al., A DNA-Based Archival Storage System. *ASPLOS '16/ 2-6* (Atlante 2016) 1-13.

Lina BOUNTOURI, Archives in the Digitale Age. Standards, Policies and Tools, (Cambridge 2017).

Polymerase Chain Reaction, in: Britannica, [https://www.britannica.com/science/polymerase-chain-reaction#:~:text=polymerase%20chain%20reaction%20\(%20PCR\),%20a%20technique%20used%20to%20make](https://www.britannica.com/science/polymerase-chain-reaction#:~:text=polymerase%20chain%20reaction%20(%20PCR),%20a%20technique%20used%20to%20make) [08.08.2024].

Thomas BUKO, Nella TUCZKO, Takao ISHIKAWA, DNA Data Storage. *Bio Tech* 12, (2023) 1-17.

Orthophoto, in: Bundesamt für Eich- und Vermessungswesen, <https://www.bev.gv.at/Services/Produkte/Luftbildprodukte/Orthophoto-Farbe.html> [19.08.2024].

Elisa DE LLNAO, Haichao MIAO, Yasaman AHMADI et al., Adenita. Interactice 3D Modelling and visualization of DNA nanostructures. *Nucleic Acids Research* (2020) 1-7.

Jie CHAO, Yunfeng LIN, Huajie LIU, Lianhui WANG, Chunhai FAN, DNA-based plasmonic nanostructures. *Materials Today* (2015) 1-10.

Weida D. CHEN, A. Xavier KOHLL, Bichlien H. NGUYEN et al., Combining Data Longevity with Storage Capacity. Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. *Advanced Functional Materials* 29/28 (2019) 1-8.

George M. CHURCH, Yuan GAO, Sriram KOSURI, Next-Generation Digital Information Storage in DNA. *Science* (2012) 1-2.

Tiered Storage, in: Computer Weekly, <https://www.computerweekly.com/de/definition/Tiered-Storage-Abgestuftes-Storage#:~:text=Tiered%20Storage%20stellt%20Speicherebenen%20zur%20Verf%C3%BCgung,%20auf%20denen%20Daten%20einer> [05.08.2024].

Forschungsergebnis, in: Oligoarchive, <https://cordis.europa.eu/project/id/863320/results/de>, [29.07.2024].

Oligoarchive - Intelligent DNA Storage for Archival, in: Cordis Europe, <https://cordis.europa.eu/project/id/863320> [29.07.2024].

Elisa DE LLNAO, Haichao MIAO, Yasaman AHMADI et al., Adenita. Interactice 3D Modelling and visualization of DNA nanostructures. *Nucleic Acids Research* (2020) 1-7.

About us, in: Disco tech, <https://disco-tech.eu/partners/> [27.07.2024].

Publications, in: Disco-tech, <https://disco-tech.eu/publications/> [27.07.2024].

DNA Data Storage Alliance, Preserving our digital legacy. An Introduction to DNA Data Storage (2021) 1-37.

Andrea DORICCHI, Casey PLATNICH, Andreas GIMPEL et al., Emerging Approaches to DNA Data Storage. Challenges and Prospects. *ACS Nano* (2022) 17552-17571.

Yiming DONG, Fajia SUN, Zhi PING et al., DNA storage. Research landscape and future prospects. *National Science Review* 7 (2020) 1092-1107.

Edward EDELSON, Francis Crick and James Watson and the building blocks of life (New York 1998).

Katharina ERNST, Standards und Normen im Bereich der Langzeitarchivierung. *Archivar. Zeitschrift für Archivwesen* 74/2 (2021) 62-70.

Chisom EZEKANNAGHA, Anke BECKER, Dominik HEIDER et al., Design considerations for advancing data storage with synthetic DNA for long-term archiving. *Materials Today Bio* (2022).

Nicole FEGGER, Digitaler CO₂-Fußabdruck. Digitalisierung und Nachhaltigkeit, in: Overlook, [https://overlook.one/blog/digitaler-CO₂-Fu%C3%9Fabdruck](https://overlook.one/blog/digitaler-CO2-Fu%C3%9Fabdruck) [08.06.2024].

Ryan GABRYS, Hans Mao KIAH, Olgica MILENKOVIC, Asymmetric Lee Distance Codes for DNA-Based Storage. *IEEE Transactions of Information Theory* 6/8 (2017) 4982-4995.

Christy GEARHEART, Eric ROUCHKA, Benjamin ARAZI, DNA-Based Active Logic Design and its Implications. *Journal of Emerging Trends in Computing and Information Science* 3/5 (2021) 756-766.

Nick GOLDMAN, Paul BETRONE, Siyuan CHEN et al., Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. *Nature* (2013), 77-80.

Reinhard HECKEL, An archive written in DNA. *Nature Biotechnology* 36 (2018) 236-237.

Thomas HEINIS, Jamie ALNASIR, Survey of Information Encoding Techniques for DNA. *ResearchGate* (2022) 1-28.

Burkhard HEISE, Dagmar ULLRICH, Digitale Langzeitarchivierung. *PIK* 27/2 (2004) 103-109.

Huffman-Codierung, in: Informatik Verstehen, <https://www.informatik-verstehen.de/lexikon/huffman-codierung/#:~:text=Huffman-Codierung.%20Die%20Huffman-Codierung%20ein%20Algorithmus%20zur%20verlustfreien%20Komprimierung%20bin%C3%A4rer%20Daten>, [26.01.2024].

JP, NÖ Landesarchiv wird zum Digitalen Gedächtnis des Landes. Digitalisierung des Archivs ist Pionierprojekt in Österreich, in: Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare, <https://voeb-b.at/noe-landesarchiv-wird-zum-digitalen-gedaechtnis-des-landes/#:~:text=Durch%20das%20digitale%20Archiv%20k%C3%B6nnen%20so%20elektronische%20Akten%20aus%20dem> [07.06.2024].

Jonas KERSCHNER, Aspekte der Archivierung digitaler Geodaten und kartographischer Darstellungsformen (Wien 2013).

Raphael KIM, Larissa PSCHETZ, Conor LINEHAN et al., Archives in DNA. Workshop Exploring Implications of an Emerging Bio-Digital Technology through Design Fiction. *Academic Mindtrek* (2021) 102-105.

Paul KLIMPEL, Jürgen KEIPER (hg.), Was bleibt? Nachhaltigkeit der Kultur in der digitalen Welt (Berlin 2013).

Julian KOCH, Ann-Christin KERL, Natascha SCHAWALDER et al., Preserving DNA in Biodegradable Organosilica Encapsulates. *American Chemical Society* (2022), 11191-11198.

KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/jpeg2000.html> [07.06.2024].

KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/odf.html> [07.06.2024].

KOST, Koordinationsstelle für dauerhafte Archivierung elektronischer Unterlagen, <https://kost-ceco.ch/cms/txt.html> [07.06.2024].

Landesentwicklung, Zahlen/Fakten/Karten, in: GIS-Steiermark, <https://www.landesentwicklung.steiermark.at/cms/ziel/141976122/DE/> [05.08.2024].

Dave LANDSMAN, Karin STRAUSS, The DNA Data Storage Model. *IEEE Computer* 56/7 (2023) 78-85.

Christopher LEE, Open Archival Information System (OAIS) Reference Model. *Encyclopaedia of Library and Information Sciences* (2010), 4020-4030.

DNA- & RNA-Aufreinigung, Oligonukleotidhandhabung und -stabilität, in: Merck, <https://www.sigmaaldrich.com/AT/de/technical-documents/technical-article/genomics/dna-and-rna-purification/sample-purification-and-quality-assessment> [18.08.2024].

A Thousand of Books for a Million Years, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/1000books/#:~:text=A%20Thousand%20Books%20for%20a%20Million%20Years.%20Das%20Memory%20of> [24.01.2024].

Home, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/> [22.05.2024].

Wer soll MOM finden?, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/who-is-supposed-to-find-mom/>, [22.05.2024].

Wie bewahrt man Informationen dauerhaft für 1 Million Jahre auf?, in: Memory of Mankind, <https://www.memory-of-mankind.com/de/how-is-information-kept-legible-for-1-million-years/>, [22.05.2024].

Project Silica, in: Microsoft, <https://www.microsoft.com/en-us/research/project/project-silica/> [24.01.2024].

Project Silica Overview, in: Microsoft, <https://www.microsoft.com/en-us/research/project/project-silica/>, [24.05.2024].

Research, in: Microsoft, <https://www.microsoft.com/en-us/research/video/research-talk-storing-data-for-millennia/>, [24.05.2024].

Paul MUIR, Shantao LI, Shaoke LOU et al., The real cost of sequencing. Scaling computation to keep pace with data generation. *Genome Biology* (2016) 4-9.

Steve NADIS, Tough microbe's DNA could be a lasting archive. *Microbiology* 376/6480 (2020) 840-841.

Heike NEUROTH, Achim OSSWALD, Regine SCHEFFEL, Stefan STRATHMANN, Mathias JEHN (Hg.), Nestor Handbuch. Eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Kapitel 4. Das Referenzmodell OAIS-Open Archival Information System, (2010).

What is a database?, in: Oracle, <https://www.oracle.com/database/what-is-database/#:~:text=A%20database%20is%20an%20organized%20collection%20of%20structured%20information,%20or> [06.08.2024].

Digitale Archivierung im Österreichischen Staatsarchiv, in: Webseite des Österreichischen Staatsarchiv, <https://www.oesta.gv.at/ueber-uns/digitales-archiv-oesterreich.html#:~:text=Die%20Verantwortung%20f%C3%BCr%20die%20Umsetzung%20der%20digitalen%20Archivierung%20obliegt%20gem%C3%A4%C3%9F> [07.06.2024].

Zhi PING, Shihong CHEN, Sha Joe ZHU, Carbon-based archiving. Current progress and future prospects of DNA-based data storage. *GigaScience* (2019), 1-10.

Potomac Institute für Policy Studies, The Future of DNA Data Storage (Arlington 2018).

Storage of DNA, in: Qiagen, <https://www.qiagen.com/us/knowledge-and-support/knowledge-hub/bench-guide/dna/handling-dna/storage-of-dna#:~:text=Storage%20of%20DNA.%20Purified%20DNA%20should%20be%20stored%20at%20%E2%80%9320%C2%B0C> [18.04.2024].

S.V. RAMANAMURTHY, P.L. KIRANMAYI, S.D. SARAYU et al., DNA Data Storage. *Journal for Research* 3/11 (2018) 32-37.

Thomas SCHÄRLI, Kollektives Gedächtnis, Geschichtswissenschaft und digitale Archive, in: Digitales Gedächtnis. Archivierung und die Arbeit der Historiker der Zukunft, hg. von Patrick JUCKER-KUPPER - Christoph KOLLER - Gerold RITTER (Geschichte und Informatik 13/14, Zürich 2004) 57-82.

Jeff SCHMERKER, What is an Oligo?, in: Integrated DNA Technologies, <https://eu.idtdna.com/pages/community/blog/post/what-is-an-oligo#:~:text=What%20are%20oligo%20drugs?%20Oligo%20drugs,%20or%20oligonucleotide%20therapeutics,%20can> [06.08.2024].

Markus SCHMIDGALL, Zum Stand der digitalen Archivierung in den österreichischen Bundesländern. *Scrinium* 73, Zeitschrift des Verbandes der Österreichischen Archivarinnen und Archivare (2019), 86-89.

TE-Puffer, in: SERVA,
https://www.serva.de/deDE/ProductDetails/949_544_42554_TE_Puffer_100x_16_544.html#:~:text=Tris-EDTA-Puffer%20ist%20ein%20h%C3%A4ufig%20verwendeter%20Molekularbiologiepuffer%20zum%20Resuspendieren%20und/oder%20Verd%C3%BCnnen [18.08.2024].

Richard SOLOMON, Melitte BUCHMANN, Clark JOHNSON et al., Write Once, Read Forever (WORF). Low-energy storage in perpetuity of high-density, multi-state data. *Final Program and Proceedings* 5 (2014) 118-122.

Richard J. SOLOMON, Eric ROSENTHAL, Pedro G.C. DE OLIVEIRA et al., WORF (Write Once, Read Forever). Next Generation Archival Big Data Storage, *IEEE Aerospace Conference* (2021) 1-7.

Tim STATLER, What is a Primary Key?, in: Computer Science Central,
[https://compscicentral.com/what-is-a-database-primary-key/#:~:text=A%20primary%20key%20is%20a%20column%20\(s\)%20within%20a%20relational](https://compscicentral.com/what-is-a-database-primary-key/#:~:text=A%20primary%20key%20is%20a%20column%20(s)%20within%20a%20relational) [06.08.2024].

George B. STEFANO, Fuzhou WANG, Richard M. KREAM, DNA MemoChip. Long-Term and High-Capacity Information Storage and Select Retrieval. *Medical Science Monitor* (2018) 1185-1187.

Wolfgang STIELER, Petabyte pro Gramm, in: Heise Online,
<https://www.heise.de/hintergrund/Petabyte-pro-Gramm-3664504.html#:~:text=Dr.%20Wolfgang%20Stieler.%20DNA-Molek%C3%BCle%20sind%20gewisserma%C3%9Fen%20Festplatten%20%E2%80%93%20nur%20eben> [26.01.2024].

Martin STÜRZLINGER; Antrittsvorlesung FH Potsdam, 18.12.2023.

Yiming w, Fajia SUN, Zhi PING et al., DNA storage. Research landscape and future prospects. *National Science Review* 7 (2020) 1092-1107.

Sealed in Glas, in: Unlocked Microsoft, <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.,> [24.05.2024].

Mit DNA zur nachhaltigen Datenspeicherung, in: Universität Paderborn, <https://www.uni-paderborn.de/nachricht/127311>, [13.03.2024].

Sealed in Glas, in: Unlocked Microsoft, <https://unlocked.microsoft.com/sealed-in-glass/#:~:text=Symbiosis%20in%20the%20cloud&text=Data%20is%20stored%20in%20glass,any%20of%20the%20storage%20units.,> [24.05.2024].

WORF, in: Creative Technology, <http://www.creative-technology.net/CTECH/WORF.html>, [24.01.2024].

S.M. Hossein Tabatabaei YAZID, Han Mao KIAH, Eva Ruiz GARCIA, DNA-Based Storage. *Trends and Methods* (2015) 1-20.

9. Abbildungsverzeichnis

Abbildung 1, OAIS-Referenzmodell.	16
Abbildung 2, Archivierung auf Papier.	39
Abbildung 3, Ein Beispiel von WORM Medien, welche aus mehrfach lesbaren Informationen bestehen.	40
Abbildung 4, Ceramic data carrier.	40
Abbildung 5, Schritt 1 und 2 des Projekt Silicas.	41
Abbildung 6, Schritt 3 und 4 von Project Silica.	41
Abbildung 7, Darstellung des Ablaufes der Schaffung eines DNA Data Storage.	42
Abbildung 8, Ablauf des DNA Data Storages.	45
Abbildung 9, Ablauf des DNA Data Storages.	45
Abbildung 10, Binary Transcoding.	48
Abbildung 11, ASCII Codierung der Nachricht.	49
Abbildung 12, Huffman Codierung.	50
Abbildung 13, Ablauf der chemischen Synthetisierung von DNA.	52
Abbildung 14, Synthese mit Ligation.	53
Abbildung 15, Sequencing by Synthesis.	56
Abbildung 16, Nano Sequencing.	57
Abbildung 17, Speicherhierarchie nach Bornholt et al.	63
Abbildung 18, Darstellung des Speichersystems von Bornholt et al.	64
Abbildung 19, Darstellung des Hinzufügens des PCR-Primers als Key.	65
Abbildung 20, Darstellung des Auslesens und der Verstärkung des Primers.	65
Abbildung 21, Nanostrukturen von DNA.	70

10 Abkürzungsverzeichnis

A-Base - Adenin

AI – Artificial Intelligence

ASCII – American Standard Code for Information Interchange

BIT – Binary Digit

C-Base - Cytosin

DBMS - Datenbankmanagementsystem

DNA/ DNS – deoxyribonucleic acid/ Desoxyribonukleinsäure

EDTA - Ethylendiamintetraessigsäure

ELAK – Elektronischer Akt

G-Base – Guanin

GB - Gigabyte

GIS – Geographic Information System

HDD – Hard Drive Devices

ID – Identifikator

ISO - International Organization for Standardization

JPEG – Joint Photographic Experts Group

MP3 – Moving Picture Experts Group

NGS- Next Generation Sequencing

OAIS – Open Archival Informationssystem

ODF – Open Document Format

ODP – Open Document Präsentation

ODT – Offene Textdatei

OSI – Open System Interconnection

PCR – Polymerase Chain Reaction

PDF/A – Portable Document Format/ Archive

SBS – Single Base Substitution

SSD – State Storage Devices

T-Base – Thymin

TB - Terabyte

TE-Puffer – TRIS-EDTA-Puffer

TRIS- Trishydroxymethylamionmethan

TXT - Text

WORF – Write Once, Read Forever – Speichermethoden

XML – Extensible Markup Language

ZIP – Zipper

11. Anhang

FRAGEBOGEN NEO-PROJEKT

1. Was war Ihre Motivation, bei diesem Projekt mitzumachen? Was möchten Sie damit erreichen?
2. Welche Ziele haben Sie sich für das Forschungsprojekt gesetzt?
3. Wie baut sich Ihr Forschungsteil auf?
 - 3.1. Welche Arbeitsschritte umfasst Ihr Bereich?
4. Welche technischen Voraussetzungen braucht es für Ihre Arbeit?
 - 4.1. Wie lange dauert der Prozess der Speicherung von Daten auf DNA?
 - 4.2. Wie genau funktioniert Ihre Methode der Speicherung von Daten auf DNA?
5. Welche Kosten fallen bei einem DNA Data Storage an?
 - 5.1. Welche Kosten entstehen bei der Herstellung der DNA?
 - 5.2. Welche Kosten entstehen während der Lagerung der DNA?
 - 5.3. Welche Kosten entstehen beim Auslesen der auf DNA gespeicherten Informationen?
6. Welche Vorteile hat die Speicherung auf DNA gegenüber herkömmlichen Speichermethoden?
7. Handelt es sich bei der DNA Speicherung um eine klimafreundlichere Alternative zu bisherigen Speichern?
8. Wie könnte die praktische Anwendung Ihres Speichermediums ausschauen?
 - 8.1. Kann man die Haltbarkeit eines DNA Data Storages zeitlich abschätzen?
 - 8.2. Wie lange dauert ein Speichervorgang?

9. Für welche Bereiche können Sie sich die Anwendung eines solchen Speichermediums vorstellen? (IT, Universitäten, Privatwirtschaft, Silicon Valley, ...?)
10. Wie schätzen Sie die praktische Anwendung des DNA Data Storage im Archiv ein?
- 10.1. Welche Voraussetzungen braucht ein Archiv, um mit dieser Speichermethode zu arbeiten?
11. Können Sie sich das Endprodukt Ihrer Forschung/ Ihre Forschungsergebnisse als Consumerprodukt vorstellen?