



© iStock_ym german

■ Die Sackgasse der Datenspeicherung

• Stetig wachsende Datenmengen...

Das Computerzeitalter, das der digitalen Revolution der 1970er Jahre folgt, wird manchmal ebenfalls als Informationszeitalter bezeichnet. Tatsächlich haben die neuen Technologien unseren Umgang mit der Information tiefgehend geändert und zu einem exponentiellen Wachstum ausgetauschter und gespeicherter Datenmengen geführt. Die steigende Anzahl der Nutzer, die Vervielfachung der Endgeräte und die Entwicklung neuer Praktiken („Internet der Dinge“, „Cloud“- und „Big-Data“-Ansätze usw.) dürfte so zu einer raschen Zunahme der in den *data centers*¹ aufbewahrten Datenmengen führen; die globale Datenmenge wird von 33 Zettabyte (ZB, 10²¹ Bytes) im Jahr 2018 voraussichtlich auf 175 ZB im Jahr 2025 ansteigen² und könnte bis 2040 sogar 5000 ZB erreichen³.

• ... und demgegenüber veraltete Speichermedien

Dieses Wachstum bleibt jedoch nicht ohne Folgen. Die Auswirkungen der Digitalisierung auf die Umwelt werden von den Nutzern aufgrund der Tatsache, dass die genutzte Infrastruktur nicht sichtbar ist, unterschätzt, sind jedoch alles andere als anekdotisch: 2019 entfiel auf diese Sparte 3,6 % des weltweiten Energieverbrauchs und soll 2025 voraussichtlich einen Anteil von zwischen 4,8 und 5,9 % erreichen⁴. Die *data centers*, in denen digitale Daten verarbeitet und gespeichert werden, machten 2017 19% des Energieverbrauchs der Digitalbranche aus⁵. Für deren Herstellung sind Metalle und Rohstoffe der seltenen Erden sowie große Mengen an Kühlwasser notwendig.

Zusammenfassung

- Der Übergang in das digitale Zeitalter wird von einer stetig steigenden Produktion an Datenmengen begleitet, die anschließend gespeichert werden müssen. Dies führt zu einer neuen Problematik, insbesondere in Bezug auf Energie, Umwelt und Sicherheit.
- Der Einsatz von DNA-Datenspeichern könnte bahnbrechend sein und mehrere dieser Probleme auf einen Schlag lösen.
- Für sich allein genommen ist diese Technologie jedoch keine Lösung und müsste mit Änderung in unseren Praktiken einhergehen.

Ludovic Haye, Senator

Trotz der leistungsfähigeren *data centers* wird deren Energieeffizienzsteigerung wahrscheinlich nicht ausreichen, um die exponentiell zunehmende Nutzung auszugleichen, die sich in einer Zunahme des CO₂-Fußabdrucks aufgrund der Datenspeicherung niederschlägt.⁶ Dies gilt umso mehr, als sich die derzeitigen Speichertechnologien ihrem theoretischen Optimum nähern⁷, was auf eine baldige Verlangsamung der Effizienzsteigerungen hindeutet⁶.

Außerdem stellt sich die Frage, ob die für die Herstellung von Geräten notwendigen Fähigkeiten in ausreichender Menge zur Verfügung stehen, um der zu erwartenden zunehmenden Nachfrage an Speichern gerecht zu werden.^{8,9} Eine bahnbrechende Innovation im Bereich der Datenspeicherung scheint daher notwendig.

■ DNA - eine verlockende und vielversprechende Lösung

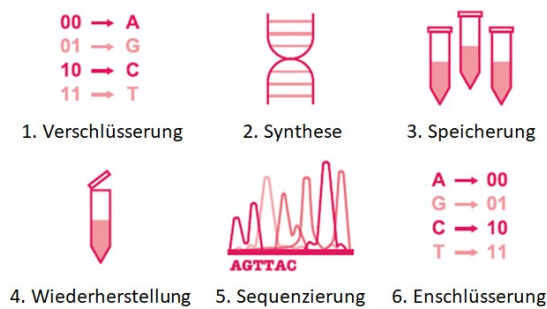
• Die Festplatte der Natur...

Seit mehreren Milliarden Jahren werden in den DNA-Molekülen der Lebewesen ihre genetischen Informationen gespeichert. Diese Moleküle bestehen aus zwei antiparallelen Strängen, die zur Bildung einer Doppelhelixstruktur umeinander gewickelt sind und aus einer Ansammlung von Nukleotiden bestehen. Jedes Nukleotid enthält eine der vier Stickstoffbasen (Adenin (A), Cytosin (C), Guanin (G), Thymin (T)), die an einen Desoxyribosezucker gebunden ist, der wiederum an eine Phosphatgruppe gebunden ist. Da die DNA über vier stickstoffhaltige Basen verfügt, bietet sie ein quaternäres Zahlensystem für die Informationskodierung.

• ... und für die Zukunft?

Die Entschlüsselung der molekularen DNA-Struktur, die 1962 mit dem Nobelpreis für Medizin ausgezeichnet wurde, spielte eine Schlüsselrolle bei der Entwicklung der Molekularbiologie. In der zweiten Hälfte des 20. Jahrhunderts wurden Techniken zur DNA-Synthese und DNA-Sequenzierung¹⁰ entwickelt. Mit den heutigen Technologien können DNA-Sequenzen sowohl „geschrieben“ als auch „gelesen“ werden. Man könnte sich also die Natur zum Vorbild nehmen und die DNA als Datenspeicher einsetzen.

Dazu muss die Binärsequenz einer zu speichernden Datei in eine DNA-Sequenz (eine Folge von Nukleotiden) übersetzt werden, die dann synthetisiert und gespeichert werden kann. Um die Datei wiederherzustellen, muss das DNA-Molekül sequenziert und die Sequenz anschließend entschlüsselt werden.



Credits: DNA Data Storage Alliance¹²

DNA als Datenspeichermedium könnte viele Vorteile bieten. Zuallererst eine beträchtliche Informationsdichte: Um 1 Bit zu speichern sind etwa 50 Atome nötig, während bei einer magnetischen Datenspeicherung 1 Million benötigt werden^{3,11}. Eine weitere Stärke der DNA ist ihre Alterungsbeständigkeit: Wird sie unter den richtigen Bedingungen gelagert, kann ein DNA-Molekül mehrere tausende Jahre aufbewahrt werden, während auf herkömmlichen Datenträgern gespeicherte Daten alle 5 bis 10 Jahre kopiert werden müssen, um deren Verfall zu vermeiden. Außerdem ist die DNA als Träger der genetischen Information nicht dem Risiko einer Überalterung ausgesetzt und wird universelles Speichermedium bleiben. Als abschließendes Argument sei noch zu erwähnen, dass DNA-Datenspeicher und deren Daten mithilfe bereits beherrschter enzymbasierter Technologien einfach zu vervielfältigen sind.

So scheint unter den verschiedenen Alternativen, die zur Bewältigung der Problematik der Datenspeicherung vorgeschlagen wurden, die Verwendung von DNA eine der vielversprechendsten zu sein und ist die erste, die international zahlreiche akademische und industrielle Akteure mobilisiert hat¹². Dennoch gibt es noch verschiedene technologische Hürden zu nehmen, bevor DNA als Informationsträger genutzt werden kann.

■ Aktuelle Entwicklungen

Für die Entwicklung eines DNA-Datenspeichers ist die Zusammenarbeit zahlreicher wissenschaftlicher Disziplinen notwendig.¹³ Daher sind Synergien zwischen

Wissenschaftlern, sowohl aus der akademischen Welt als auch aus der Industrie notwendig, die an verschiedenen Themen arbeiten können. Derzeit basiert die Forschung auf kollaborativen und interdisziplinären Projekten, wie das US-amerikanische Konsortium *Molecular Information Storage* (MIST)¹⁴ das europäische Programm *OligoArchive*¹⁵ oder die französischen Projekte *dnarXiv*¹⁶ und *MolecularXiv*¹⁷.

• Kodierungsverfahren

Das Kodierungsverfahren, mit dem die Daten synthetisiert, gespeichert und anschließend sequenziert werden, wirkt sich direkt auf die einzelnen Schritte und folglich auf den gesamten Prozess aus. Durch die Entwicklung eines effizienten Kodierungsverfahrens (das die Anzahl der benötigten Nukleotide minimiert) kann nicht nur die Informationsdichte erhöht, sondern auch der Gesamtprozess optimiert werden (die Synthese- und Sequenzierungsschritte sind sowohl langsam als auch kostspielig).

Das Kodierungssystem muss jedoch auch so entwickelt werden, dass Fehler während der Synthese- und Sequenzierungsschritte bestmöglich vermieden werden und hängt daher direkt von den eingesetzten Technologien ab. Deshalb muss das Kodierungssystem gewisse Beschränkungen hinsichtlich der Reihenfolge und Anteile der verschiedenen Stickstoffbasen¹⁸ berücksichtigen und es müssen Kontrollnukleotide zur Fehlererkennung und -korrektur eingebaut werden. Außerdem sollten vorzüglich keine DNA-Stränge mit mehr als 200 Nukleotiden verwendet (bei denen mit den derzeitigen Technologien häufiger Synthesefehler auftreten¹⁹) und Dateien in verschiedene Fragmente unterteilt werden. Das Kodierungsverfahren muss außerdem über ein Indexierungs- und Adressierungssystem verfügen, mit dem die Datei während dem Verschlüsselungsschritt wieder zusammengesetzt werden kann²⁰. Außerdem ermöglicht dieser Schritt ebenfalls eine Verschlüsselung der gespeicherten Daten, um jegliche Probleme im Zusammenhang mit Cybersicherheit zu vermeiden²¹.

• Zusammenfassung

Die derzeit mehrheitlich eingesetzte Technologie bei der Synthese von DNA-Fragmenten basiert auf dem Prinzip der sequentiellen Synthese, bei der die Nukleotide nacheinander hinzugefügt werden. Um zu verhindern, dass mehrere Nukleotide gleichzeitig angehängt werden, trägt jedes verwendete Nukleotid eine Schutzgruppe, die verhindert, dass ein anderes angehängt wird. Nukleotide können auf verschiedene Weisen an die zu konstruierende Kette angehängt werden: mittels chemischer Reaktion (in diesem Fall eine sog. „chemische“ Synthese) oder mittels einer terminalen Desoxynukleotidyltransferase²², einem Enzym (sog. „enzymatische“ Synthese). Wurde ein Nukleotid angehängt, wird die Schutzgruppe abgespalten, um das nächste Nukleotid hinzuzufügen.

Die chemische Methode, die die historisch eingesetzte Methode schlechthin darstellt, wurde automatisiert und miniaturisiert. Die Parallelisierung ermöglicht die Herstellung innerhalb von 24 Stunden von bis zu einer Million verschiedener DNA-Fragmente mit 200 Nukleotiden²³. Die enzymatische Synthese, die derzeit noch nicht auf dem Markt ist und von mehreren Unternehmen entwickelt wird (u. a. vom dem in diesem Bereich führenden französischen Unternehmen DNA Script) dürfte einen echten Fortschritt darstellen: sie ist umweltfreundlicher und ermöglicht einen höheren Durchsatz bei geringeren Kosten und einer niedrigeren Fehlerquote.

Eine weitere Möglichkeit besteht darin, anstatt jedes Nukleotid einzeln hinzuzufügen, kurze Abschnitte präsynthetisierter Oligonukleotide zusammensetzen. In diesem Fall können mithilfe kohäsiver Einzelstrangenden und der Wirkung eines spezifischen Enzyms, der DNA-Ligase, diese Abschnitte miteinander verbunden werden. Diese Methode ermöglicht die Synthese langer DNA-Stränge mit einer geringen Fehlerquote bei einer beträchtlich höheren Synthesegeschwindigkeit. Das US-amerikanische Unternehmen Catalog DNA hat ein Gerät entwickelt, das diese Methode einsetzt und mit dem das Äquivalent von 500 KB/s synthetisiert werden kann³.

Der kritischste Schritt ist die Synthese, bei der, will man die DNA-Datenspeicherung allgemein verbreiten, die meisten Hürden zu nehmen sind: Einerseits muss der Durchsatz erheblich verbessert und andererseits müssen die inhärenten Kosten drastisch gesenkt werden. Derzeit konzentriert man sich vor allem auf eine massive Parallelisierung bereits existierender Techniken, jedoch könnte man durch die Entwicklung neuer enzymatischer Prozesse große Fortschritte erreichen. Außerdem wurden die derzeitigen Technologien vor allem für medizinische Anwendungen entwickelt, deren Anforderungen (mit einer geringstmöglichen Fehlerquote und Kompromissen bei Größenordnung, Kosten und Durchsatz) stark von denen der Datenspeicherung abweichen (bei der eine höhere Fehlerquote akzeptabel ist, Größenordnung, Kosten und Durchsatz jedoch vorrangig sind). Das Aufkommen von auf einzig auf die DNA-Datenspeicherung spezialisierten Unternehmen²⁴ sollte diesen spezifischen Anforderungen gerecht werden und zur Entwicklung entsprechend geeigneter Lösungen führen.

• Speicherung

Um über lange Zeiträume hinweg aufbewahrt werden zu können, müssen DNA-Moleküle vor Wasser und Sauerstoff, aber auch vor Licht und vor hohen Temperaturen geschützt gelagert werden²⁵. Bei herkömmlichen Techniken zur Vermeidung von DNA-Schäden werden Moleküle bei niedrigen Temperaturen gelagert. Diese Methode ist einerseits teuer (Platz, Ausrüstung, Energie und Wartung) und birgt andererseits das Risiko von Verlusten bei Fehlfunktionen. Die Folge ist,

dass man, im Vergleich zu derzeit eingesetzten Speichertechnologien, nicht von den Vorteilen eines DNA-Datenspeichers voll profitieren kann.

Um Abhilfe zu schaffen hat das französische Unternehmen Imagene Edelstahlkapseln mit einem Glaseinsatz zur Aufbewahrung von DNA-Molekülen (die zuvor vakuumgetrocknet und anschließend in eine Inertatmosphäre gebracht werden) entwickelt. Hochrechnungen zur Alterungsbeständigkeit bei dieser Technologie haben gezeigt, dass bei Raumtemperatur mit einer Haltbarkeit von bis zu mehreren tausend Jahren gerechnet werden kann. Der Einkapselungsprozess wurde zudem vollständig automatisiert und könnte daher leicht in ein Speichersystem integriert werden.

Ebenfalls untersucht wurde die Möglichkeit, DNA-Fragmente *in vivo* in lebenden Zellen bzw. Organismen zu konservieren. Selbst wenn manche Bakterien mehrere Millionen Jahre überleben können, ist die Datenmenge, die in einem Wirt gespeichert werden kann, begrenzt, so dass mit dieser Methode keine hohe Informationsdichte erreicht werden kann. Außerdem wirft diese Art der Speicherung die Frage nach der Verträglichkeit der injizierten Fragmente durch den Wirtsorganismus auf, da sich DNA-Sequenzen als toxisch für den Träger oder das umgebende Biotop erweisen können²⁶. Aus den eben genannten Gründen der Leistung und Bioethik wurde eine *In-Vivo*-Speicherung nur im experimentellen Rahmen untersucht und alternative *In-Vitro*-Methoden vorgezogen.

• Sequenzierung

In den letzten Jahren wurden bei den Sequenzierungstechnologien enorme Fortschritte gemacht, und zwar weitaus schneller, als vom Mooreschen Gesetz²⁷ in der Informatik beschrieben wird. Während die Sequenzierung des ersten menschlichen Genoms (die 2001 abgeschlossen wurde)²⁸ zwischen 500 Millionen und 1 Milliarde US-Dollar gekostet hatte, liegen die Kosten heute bei knapp 700 US-Dollar. Dennoch sind weitere Verbesserungen notwendig – jedoch nicht so große wie bei der Synthese –, bevor DNA als echter Datenspeicher in Betracht gezogen werden kann.

Die derzeit verbreitetste, von der Firma Illumina entwickelte Technologie, leitet sich direkt von der historisch²⁹ eingesetzten, verbesserten und massiv parallelisierten Methode ab. Das DNA-Zielmolekül wird zunächst fragmentiert und in einen DNA-Einzelstrang umgewandelt. Nach Anheftung und Amplifikation der Fragmente an eine planare Oberfläche wird eine DNA-Polymerase verwendet, um komplementäre Stränge in Anwesenheit von Nukleotiden zu erhalten. Jedes Nukleotid verfügt über einen abtrennbaren Fluor-Farbstoff (jede Stickstoffbase hat eine andere Farbe), der ebenfalls als Schutzgruppe dient. Nach jeder Zugabe wird ein digitales Bild erstellt, mit dem die Art des hinzugefügten Nukleotids identifiziert werden kann. Anschließend wird

der Fluor-Farbstoff abgespalten und ein neues Nukleotid kann hinzugefügt werden. Die Sequenz wird letztendlich auf Grundlage der verschiedenen Bilder durch Computerverarbeitung gewonnen, wobei die Überlappungen für eine Neuordnung der Fragmente genutzt werden.

Kürzlich entwickelte die Firma Oxford Nanopore Technologies eine innovative Technik, die auf einem ganz anderen Prinzip basiert und die ein echter Gewinn für die DNA-Datenspeicherung darstellen könnte. Mittels Einsatz eines elektrischen Feldes werden die eingangs in DNA-Einzelstränge umgewandelten DNA-Zielfragmente durch eine Membran mit Nanoporen transportiert. Durchtritt ein Nukleotid eine Nanopore wird der entstehende Ionenfluss gemessen und so der Nukleotid-Typ bestimmt. Diese neue Methode ist zwar mit einer höheren Fehlerrate³⁰ verbunden, verfügt jedoch über eine sehr kurze Laufzeit (derzeit können bis zu 450 Nukleotide pro Sekunde sequenziert werden) und macht die Schritte der Amplifikation und der Computerberechnung zur Rekonstruktion der Sequenz überflüssig.

- **Alternativen zur Verwendung von DNA**

Auch der Einsatz von unabhängigen DNA-Polymerasen (Polymerasen, bei der als Monomere andere Moleküle als Nukleotide verwendet werden) wird als Alternative für die Speicherung digitaler Informationen untersucht³¹. Hauptvorteil dieser Alternative liegt in der Freiheit, die sich durch die Wahl der Monomere ergibt. Denn diese können gestaltet werden, um eine höhere Informationsdichte als bei DNA-Datenspeichern³² zu erreichen, um der verwendeten Polymerase eine bessere Stabilität oder eine bestimmte Eigenschaften zu verleihen, oder um die derzeit einschränkenden Synthese- oder Sequenzierungsschritte zu vereinfachen. Auch wenn es sich um einen vielversprechenden Ansatz handelt, mit dem man die von der DNA gebotenen Möglichkeiten noch übertreffen könnte, so ist noch viel Forschungsarbeit notwendig, da die Synthese- und Sequenzierungstechniken von Polymerasen derzeit noch nicht so leistungs- und konkurrenzfähig wie die für DNA sind.³³

■ Perspektiven

- **Mäßige Auswirkungen**

Damit DNA-Datenspeicher einsatzfähig werden und eine breite Anwendung finden, muss diese Technik vollständig automatisiert ablaufen mit Synthese- und Sequenzierungsschritten in einem akzeptablen Kosten- und Zeitrahmen. 2019 wurde ein erster Proof of Concept von einem Team der Washington University und Microsoft vorgestellt, mit einem Prototyp, der den gesamten Prozess selbstständig durchführen kann³⁴, dennoch sind noch wichtige Fortschritte zu erzielen. Derzeit schätzt man, dass die Kosten für die Sequenzierung um das tausendfache und die der Synthese noch um das hundert-millionenfache reduziert werden müssten³. Obwohl diese

Ziele beachtlich sind, müssen sie den jüngsten Entwicklungen in diesen Bereichen gegenübergestellt werden, die auf erhebliche Fortschritte in den nächsten Jahren hoffen lassen. Der damit verbundene Markt wird nach Schätzungen von BCC Research³⁵ ein erhebliches Wachstum verzeichnen.

Darüber hinaus ist abzusehen, dass die Vorgänge des Schreibens und Lesens der DNA im Vergleich zu den Technologien, die derzeit für die Speicherung digitaler Daten verwendet werden, relativ langsam bleiben. Tatsächlich wird DNA – zumindest anfänglich – als Speichermedium für „kalte“ Daten positioniert werden, d. h. Daten, die über lange Zeiträume aufbewahrt werden müssen, aber nicht regelmäßig abgerufen oder verändert werden müssen.³⁶

Wenn man berücksichtigt, dass die meisten aufbewahrten Daten nach 100 Tagen nur sehr selten abgerufen werden, machen die „kalten“ Daten 60% der digitalen Daten aus³⁷. Magnetbandgeräte, die derzeit das Hauptmedium für die Aufbewahrung dieser Daten sind, machen jedoch nur einen geringen Teil des Stromverbrauchs von Rechenzentren aus^{3,37}. Die Verwendung von DNA für die Speicherung „kalter“ Daten wird daher vor allem Vorteile in Bezug auf Langlebigkeit und Dichte bringen und weniger in Bezug auf die Energieeffizienz.

- **Notwendigkeit praktischer Entwicklungen**

Während die Verwendung von DNA – zumindest mittelfristig – keine energetische Lösung für die Datenspeicherung bieten dürfte, könnten die durch diese Technologie gebotenen Möglichkeiten *umgekehrt* durch einen „Rebound-Effekt“-Mechanismus zu einem Anstieg des Energieverbrauchs führen³⁸.

Die Entwicklung der DNA-Datenspeicherung wird daher nicht von Überlegungen über die Art und Weise des Datenverbrauchs entbinden³⁹ und muss insbesondere mit Regeln für eine gute Nutzung einhergehen, um zu bestimmen, welche Daten langfristig aufbewahrt werden sollten (insbesondere personenbezogene Daten, bei denen Umweltprobleme durch ethische Probleme⁴⁰ ergänzt werden).

■ Schlussfolgerung

Die Speicherung von Informationen in Form von DNA ist eine attraktive Technologie, die viele Vorteile für die Datenarchivierung bieten kann. Um die Entwicklung der französischen Branche zu fördern, muss der Dialog zwischen den verschiedenen Akteuren mit dem notwendigen Fachwissen unterstützt werden⁴¹. Die im Rahmen eines solchen Ansatzes erzielten Ergebnisse könnten höchstwahrscheinlich für verschiedene andere Anwendungen (im Gesundheitswesen, in der Informatik ...) genutzt werden. Dennoch sollte man in Hinblick auf die Möglichkeiten, die diese Technologie tatsächlich bieten wird und die Umweltherausforderung realistisch bleiben. Der Ausbau dieser Technologie muss an Überlegungen

zum digitalen Konsumverhalten gekoppelt werden, die wiederum ein hohes Maß an Sachlichkeit erfordern.

Website des Büros:

<http://www.assemblee-nationale.fr/commissions/opicst-index.asp>

<http://www.senat.fr/opicst>

Befragte Personen

Marc Antonini, CNRS-Forschungsleiter am Labor für Informatik, Signale und Systeme in Sophia Antipolis

Raja Appuswamy, Dozent für Data Science am Eurecom

Pascal Barbry, Professor für Genomstruktur der Eukaryoten am frz. Institut für molekulare und zelluläre Pharmakologie

Dominique Lavenier, CNRS-Forschungsdirektor am frz. Forschungsinstitut für Informatik und stochastische Systeme

Jean-François Lutz, CNRS-Forschungsdirektor am Institut Charles Sadron in Straßburg

Thomas Ybert, Vorstandsvorsitzender und Mitbegründer der Gesellschaft DNA Script

Sophie Tuffet, Generaldirektorin, Vorstandsvorsitzende und Mitbegründerin der Gesellschaft Imagene

Erfane Arwani, Vorstandsvorsitzender und Mitbegründer der Gesellschaft Biomemory

François Képès, korrespondierendes Mitglied der „Académie d'Agriculture de France“ (frz. Akademie für Landwirtschaft), Mitglied der „Académie des Technologies“ (frz. Akademie für Technologie), Leiter der Arbeitsgruppe „ADN: lire, écrire, stocker l'information“ (DNA: Daten lesen, schreiben und speichern) sowie Mitverfasser des Berichts „Archiver les mégadonnées au-delà de 2040: la piste de l'ADN“ (Archivierung von Megadaten nach 2040: der DNA auf der Spur)

Alain Bravo, Mitglied und Vorsitzender des Ausschusses „éthique, société et technologies“ (Ethik, Gesellschaft und Technologien) der „Académie des Technologies“ (frz. Akademie für Technologie), Mitverfasser des Berichts „Big Data – Questions éthiques“ (Big Data - Ethische Fragen)

Louis Dubertret, Mitglied der „Académie des Technologies“ (frz. Akademie für Technologie), Mitverfasser des Berichts „Big Data – Questions éthiques“ (Big Data - Ethische Fragen)

Anne Siegel, Forschungsdirektorin am CNRS, stellvertretende wissenschaftliche Direktorin am „Institut des sciences de l'information et de leurs interactions (INS2I)“ (Institut für Informationswissenschaften und ihre Wechselwirkungen) des CNRS

Jean-Marc Rietsch, internationaler Experte für Dematerialisierung, elektronische Signatur und Archivierung, Gerichtssachverständiger beim Berufungsgericht Paris

Références

¹ Die *data centers* (auf Deutsch: Datenzentren) entsprechen Infrastrukturen, in denen Daten gespeichert und verteilt werden.

² Europäische Kommission, „Europäische Datenstrategie“, 2020 (https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de).

³ Académie des technologies (frz. Akademie für Technologie), „Archiver les mégadonnées au-delà de 2040: la piste de l'ADN“ (Archivierung von Megadaten nach 2040: der DNA auf der Spur), 2020 (http://academie-technologies-prod.s3.amazonaws.com/2020/10/15/10/17/37/9fbaaf5e-d5a6-4baf-921d-815da5d7983f/ADN_web.pdf).

⁴ In Bezug auf den CO₂-Fußabdruck entspricht dies 3,5 % der weltweiten Emissionen im Jahr 2019 und einem Anteil zwischen 5,5 und 6,9 % der Emissionen im Jahr 2025. VoiaDr: The Shift Project, „Impact environnemental du numérique : tendances à 5 ans et gouvernance de la 5G“ (Umweltauswirkungen der Digitalisierung: 5-Jahres-Trend und 5G-Management), 2021 (<https://theshiftproject.org/article/impact-environnemental-du-numerique-5g-nouvelle-etude-du-shift/>).

⁵ The Shift Project, „Lean ICT : Pour une sobriété numérique“ (Lean ICT: Für eine digitale Nüchternheit), 2018 (<https://theshiftproject.org/article/pour-une-sobriete-numerique-rapport-shift/>). Eine 2020 von den Gesellschaften Citizing und KPMG für den Ausschuss für Raumordnung und nachhaltige Entwicklung des frz. Senats durchgeführte Studie schätzte ihrerseits, dass die *data centers* im Jahr 2019 14 % des digitalen CO₂-Fußabdrucks in Frankreich ausmachen könnten (siehe Fußnote Nr.6).

⁶ „Pour une transition numérique écologique“, Informationsbericht Nr. 555 (2019-2020) der Senatoren Guillaume Chevrollier und Jean-Michel Houllégatte, erstellt durch die Informationsstelle für Auswirkungen der Digitalisierung auf die Umwelt im Namen des Ausschusses für Raumordnung und nachhaltige Entwicklung (http://www.senat.fr/commission/dvpt_durable/mission_dinformation_sur_lempreinte_environnementale_du_numerique.html).

⁷ Transistoren haben heutzutage schon fast Atomgröße erreicht, einer unüberwindbaren physikalischen Barriere. Die Leistung von Halbleitern (deren Entwicklung durch das mooresche Gesetz beschrieben wird, siehe Endnote 27) dürfte daher in den nächsten Jahren abflachen. Siehe: a) M. M. Waldrop, Nature 2016, 530, 144 (<https://www.nature.com/news/the-chips-are-down-for-moore-s-law-1.19338>); b) A. Shehabi, S. J. Smith, N. Horner, I. Azevedo, R. Brown, J. Koomey, E. Masanet, D. Sartor, M. Herrlin, W. Lintner, „United States Data Center Energy Usage Report“, Lawrence Berkeley National Laboratory, Berkeley, California, 2016 (<https://www.osti.gov/servlets/purl/1372902/>).

⁸ DNA Data Storage Alliance, „Preserving our digital legacy: an introduction to DNA data storage“, 2021 (<https://dnastoragealliance.org/dev/wp-content/uploads/2021/06/DNA-Data-Storage-Alliance-An-Introduction-to-DNA-Data-Storage.pdf>).

⁹ a) V. Zhirnov, R. M. Zadegan, G. S. Sandhu, G. M. Church, W. L. Hughes, *Nat. Mater.* 2016, 15, 366; b) Mark Whitby (Senior Vice President, Branded Products Group at Seagate Technology, EMEA) to TechRadar, 2015 (<https://www.techradar.com/news/computing-components/storage/the-data-capacity-gap-why-the-world-is-running-out-of-data-storage-1284024>).

¹⁰ Bei der Sequenzierung eines DNA-Moleküls wird die Abfolge der Nukleotide, aus denen das Molekül besteht, bestimmt.

¹¹ Dieser Vorteil erscheint besonders bemerkenswert in einer Zeit, in der die Menge der gespeicherten Daten exponentiell ansteigt und sich die Frage nach der Rolle von *data centers* stellt. Siehe: C. Diguët, F. Lopez, „L'impact spatial et énergétique des data centers sur les territoires“ (*Auswirkungen des Raum- und Energiedarfs von data centers auf die Territorien*), Rapport ADEME, 2019 (<https://bibliothèque.ademe.fr/urbanisme-et-batiment/908-impact-spatial-et-energetique-des-data-centers-sur-les-territoires-l.html>). Gemäß François Képès: „Heute belegen Rechenzentren ein Millionstel der unter Wasser liegenden Fläche. Bei der derzeitigen Entwicklungsgeschwindigkeit wird im Jahr 2040 ein Tausendstel der unter Wasser liegenden Landfläche von Datenspeichern belegt sein“, während es mithilfe der DNA „möglich sein könnte, die gesamte aktuelle Datensphäre der Menschheit auf einem Datenspeicher der Größe eines kleinen Lieferwagens zu lagern. Im Jahr 2040 wird er die Größe eines Lastwagens erreichen“. Siehe: a) A. Schwyter, *Challenges* 2021 (<https://www.challenges.fr/high-tech/fini-les-data-centers-place-aux-donnees-stockees-sur-l-adn-744391>); b) A. Couto, *Industrie & Technologies* 2020 (<https://www.industrie-techno.com/article/l-adn-pourrait-nous-permettre-de-conserver-l-ensemble-des-donnees-mondiales-dans-le-volume-d-une-camionnette-clame-francois-kepes-membre-de-l-academie-des-technologies.62299>). Außerdem kann die Dichte der DNA in der Kryptografie besonders nützlich sein, da es leicht möglich ist, Informationen in einer langen Sequenz von Nukleotiden zu verbergen.

¹² Im Oktober 2020 gründeten die Unternehmen Illumina, Microsoft, Twist Bioscience und Western Digital unter anderem die „DNA Data Storage Alliance“ mit dem Ziel, Unternehmen und Institutionen zusammenzuführen, die im Bereich der Verwendung von DNA zur Datenspeicherung arbeiten (<https://dnastoragealliance.org/>).

¹³ Das US-amerikanische Programm *Molecular Information Storage* (MIST), das von der *Intelligence Advanced Research Projects Activity* (IARPA) mit 48 Millionen US-Dollar (ca. 40 Millionen Euro) finanziert wird, hat sich zum Ziel gesetzt, ab 2025 ein Terabyte zu synthetisieren und 10 Terabyte innerhalb von 24 Stunden zu einem Preis von 1.000 US-Dollar zu sequenzieren (<https://www.iarpa.gov/index.php/research-programs/mist>).

¹⁴ Das europäische Projekt *OligoArchive* wird für eine Dauer von 3,5 Jahren durch das H2020-Programm der Europäischen Union mit 3 Millionen Euro gefördert (<https://oligoarchive.github.io>).

¹⁵ Das französische Projekt *dnarXiv*, das von INRIA und dem LabEx *CominLabs* finanziert wird, umfasst verschiedene bretonische Forschungslabors (<https://project.inria.fr/dnarxiv/>).

¹⁶ Das vom CNRS getragene Projekt *MolecularArXiv*, an dem mehr als 20 Laboratorien beteiligt sind, erhielt im September 2021 eine Finanzierung in Höhe von 20 Millionen Euro für 84 Monate im Rahmen der Programme und Ausstattungen für Explorationsforschung (PEPR). Das Ziel in fünf Jahren ist es, den Lese-/Schreibzyklus mit einer Geschwindigkeit von 1 Bit/s (d. h. 10 GB Daten in 24 Stunden) durchführen zu können, um anschließend Demonstratoren einsetzen zu können.

¹⁷ In diesem Bereich kommen insbesondere Mathematik und Signaltheorie (für die Entwicklung eines effizienten Kodierungssystems), Chemie und Molekularbiologie (für die DNA-Synthese), Genomik und Informatik (für den Sequenzierungsschritt), Mikrofluidik und Robotik (für den Bau automatisierter Geräte) zum Einsatz.

¹⁸ Der Anteil der Basen C und G darf den Anteil der Basen A und T nicht übersteigen und die aufeinanderfolgende mehr als dreimalige Wiederholung desselben Nukleotids oder eines Musters muss vermieden werden. Tatsächlich ist die einfachste Codierung (00 → A; 01 → C; 10 → G; 11 → T) nicht zwangsläufig die anpassungsfähigste. Siehe: G. M. Church, Y. Gao, S. Kosuri, *Science* 2012, 337, 1628 (<https://science.sciencemag.org/content/337/6102/1628>).

¹⁹ Je länger ein Fragment ist, desto größer ist das Risiko, einen Fehler in einem seiner Nukleotide zu finden. Da die Fehlerrate für jedes Nukleotid derzeit bestenfalls bei etwa 0,1% liegt, ist es besser, keine Fragmente mit mehr als 200 Nukleotiden zu verwenden, um eine akzeptable Fehlerrate zu erhalten.

²⁰ Ein solches System ist auch von entscheidender Bedeutung, um einen zufälligen Zugriff auf gespeicherte Daten zu ermöglichen.

²¹ Zusätzlich zu den Möglichkeiten, die der digitale Teil des Prozesses bietet, könnten die biologischen Schritte auch die Sicherheit des Speichersystems erhöhen. Das *dnarXiv*-Projekt (siehe Fußnote 16) arbeitet unter anderem an diesem Aspekt. Eine Möglichkeit, die in Betracht gezogen wird, ist die Verwendung der Primer als „Schlüssel“: Ohne Kenntnis der Primersequenz ist der Amplifikationsschritt unmöglich, was die Sequenzierung und damit das „Lesen“ verhindert.

²² Im Gegensatz zu den meisten DNA-Polymerasen fügen die terminalen Desoxynukleotidyltransferasen die Nukleotide nicht von einer einzelsträngigen Matrize, sondern zufällig hinzu. Bei der enzymatischen DNA-Synthese werden die Nukleotide einzeln (mit einer chemischen Schutzgruppe) hinzugefügt, um die Sequenz zu steuern.

²³ Erreicht durch Twist Bioscience, siehe Fußnote 3.

²⁴ Es gibt derzeit einige Unternehmen, die sich dafür entschieden haben, ihre Aktivitäten ausschließlich in diesem aufstrebenden Bereich anzusiedeln: *Catalog DNA* (USA), *Helixworks Technologies* (Irland) und *Biomemory* (Frankreich).

²⁵ Darüber hinaus können Bestrahlungen, Enzyme, Mikroorganismen, Ozon sowie verschiedene andere Luftschadstoffe und Xenobiotika hinzugefügt werden. Diese Empfindlichkeiten können jedoch auch als Vorteil bei der Speicherung sensibler Informationen gesehen werden, da sie eine sofortige Zerstörung der Informationen bei Bedarf ermöglichen (insbesondere durch die Temperatur).

²⁶ Abgesehen von der Speicherung *in vivo* scheint die Speicherung von Daten in Form von DNA keine ethischen Probleme aufzuwerfen, da die synthetisierten Moleküle nicht in Zellen eingebracht werden und daher interpretiert werden können.

-
- ²⁷ a) G. E. Moore, „Cramming More Components Onto Integrated Circuits“, Electronics 1965 (<https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf>); b) G. E. Moore, „Progress in Digital Integrated Electronics“, IEEE Text Speech, 1975 (https://www.eng.auburn.edu/~agrawvd/COURSE/E7770_Spr07/READ/Gordon_Moore_1975_Speech.pdf).
- ²⁸ National Human Genome Research Institute, „The Cost of Sequencing a Human Genome“ 2020 (<https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>).
- ²⁹ F. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. USA 1977, 74, 5463 (<https://pubmed.ncbi.nlm.nih.gov/271968/>).
- ³⁰ Obwohl es wünschenswert ist, die Fehlerrate zu minimieren, können Fehler teilweise durch Fehlerkorrekturverfahren korrigiert werden, wie sie z. B. bei Compact Discs verwendet werden (um mikroskopische Defekte des Mediums, Fingerabdrücke, Kratzer ... auszugleichen).
- ³¹ Es ist auch möglich, unkonventionelle Nukleotide zu verwenden, wodurch insbesondere eine höhere Informationsdichte erreicht werden kann.
- ³² Nicht-DNA-Polymere müssen nicht aus zwei Strängen bestehen und die gewählten Monomere können kleiner als Nukleotide sein, wodurch eine höhere Informationsdichte erreicht werden kann. Außerdem ist die Anzahl der verschiedenen Monomere, die verwendet werden können, nicht begrenzt, so dass es möglich ist, ein Zahlensystem zu verwenden, das eine kompaktere Schreibweise als das von der DNA bereitgestellte Quaternärsystem ermöglicht (z. B. ein Oktal- oder Hexadezimalsystem).
- ³³ Außerdem kann man sich bei der Verwendung solcher Polymere im Gegensatz zu DNA nicht vom Phänomen der Veralterung befreien, das für die Aufbewahrung von Daten über lange Zeiträume problematisch ist.
- ³⁴ C. N. Takahashi, B. H. Nguyen, K. Strauss, L. Ceze, Sci. Rep. 2019, 9, 4998 (<https://www.nature.com/articles/s41598-019-41228-8>).
- ³⁵ BCC Research, „DNA Data Storage: Global Markets and Technologies“, 2020 (<https://www.bccresearch.com/market-research/biotechnology/dna-data-storage-market.html>).
- ³⁶ Berechnungen von Forschern des CNRS gehen davon aus, dass bei Kosten von 1 € für das Schreiben oder Lesen von 1 MB Daten die DNA-Datenspeicherung den Markt kalter Daten für wertvolle Dokumente, die nur selten abgerufen werden (z. B. Verträge, Eigentumsurkunden, Gesetze), erobern könnte. Bei einem Preis von 1 € für das Schreiben oder Lesen von 1 GB würde der molekulare Speicher dank der angebotenen Volumenreduzierung und erhöhten Haltbarkeit zu einer attraktiven Lösung für die Datenarchivierung werden. Und bei einem Preis von 1 € für das Schreiben oder Lesen von 1 TB würde die DNA interessant werden, um sehr große Daten (physisch) zu übertragen, denn diese Übertragungen werden heutzutage durch die Kapazitäten von Servern begrenzt.
- ³⁷ Horison Information Strategies, „Tiered Storage 2020: Building the Optimal Storage Infrastructure“, 2020 (<https://horison.com/publications/tiered-storage-2020>).
- ³⁸ Daten, die sonst nicht aufbewahrt worden wären, könnten aufgrund der Verwendung von DNA als Speichermedium gespeichert werden.
- ³⁹ F. Képès, „Information et environnement“ (Information und Umwelt), Revue politique et parlementaire 2021 (<https://www.revuepolitique.fr/%ef%bb%bf%ef%bb%bfinformation-et-environnement/>).
- ⁴⁰ Académie des technologies (frz. Akademie für Technologie), „Big Data – Questions éthiques“, 2019 (<https://www.academie-technologies.fr/blog/categories/publications-de-l-academie/posts/big-data-questions-ethiques>).
- ⁴¹ Die kürzlich erfolgte Finanzierung des MolecuArXiv durch die PEPR (siehe Fußnote 17) zeigt ein gewisses Engagement in dieser Richtung und ist zu begrüßen.