



N° 4793 ASSEMBLÉE NATIONALE – N° 285 SÉNAT

© iStock\_ym german

## Summary

- *The transition to the digital age has been accompanied by the production of ever-increasing volumes of data that need to be stored, posing a number of challenges today, including energy, environmental and security issues.*
- *The use of DNA to store data could be a disruptive innovation, capable of addressing a number of these problems.*
- *However, the technology will not be a solution on its own and will doubtless need to be supported by changes in our practices.*

Mr Ludovic Hays, senator

### ■ The data storage challenge

#### • Ever more data...

The computer age, following on from the digital revolution of the 1970s, is sometimes referred to as the information age. The reality is that new technologies have profoundly altered our relationship with information, leading to exponential growth in the amount of data exchanged and stored. Hence, the volumes of data stored in data centres is likely to grow at an ever faster pace as the number of users increases, terminals multiply and new practices develop (internet of things, cloud computing, big data, etc.); the global datasphere, which was 33 zettabytes (ZB,  $10^{21}$  bytes) in 2018, is expected to grow to 175 ZB in 2025,<sup>1</sup> and could even reach 5,000 ZB in 2040.<sup>2</sup>

#### • ... overwhelming existing storage solutions

However, this growth is not without consequences. Underestimated by users because of the invisibility of the facilities used, the environmental impact of digital technology is nonetheless far from negligible: in 2019, this sector accounted for 3.6% of global energy consumption and is expected to account for between 4.8% and 5.9% by 2025.<sup>3</sup> Data centres, which process and store digital data, accounted for 19% of the world's energy consumption in 2017,<sup>4</sup> and require rare metals and earths for their construction, as well as large volumes of water for cooling purposes.

Despite improvements in data centre performance, energy efficiency gains are unlikely to be enough to compensate for the exponential growth in usage, leading to an increase in the carbon footprint of data storage,<sup>5</sup> particularly since current storage technologies are

approaching their theoretical optimum,<sup>6</sup> suggesting that efficiency gains will soon slow down.<sup>5</sup>

In addition, our capacity to produce enough facilities to meet the expected increase in storage demand is now being questioned.<sup>7,8</sup> Consequently, a disruptive innovation in the field of data storage appears to be necessary.

### ■ DNA: an attractive and promising solution

#### • The hard drive used by nature...

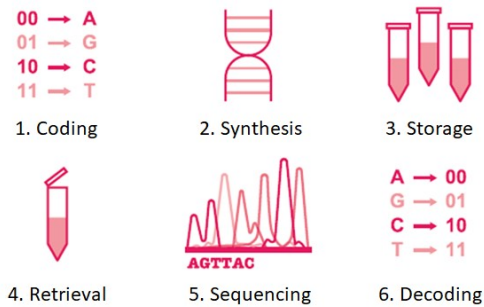
For several billion years, DNA molecules have been used by living beings to store their genetic information. These molecules consist of two antiparallel strands, wrapped around each other to form a double helix structure and composed of an assembly of nucleotides. Each nucleotide includes one of the four nitrogenous bases (adenine (A), cytosine (C), guanine (G), thymine (T)) linked to a deoxyribose sugar, itself linked to a phosphate group. Thanks to these four nitrogenous bases, DNA offers a quaternary numbering system for encoding information.

#### • ... and of the future?

The elucidation of the molecular structure of DNA, awarded the Nobel Prize in Medicine in 1962, played a key role in the development of molecular biology. The second half of the 20th century then saw the development of techniques for both DNA synthesis and sequencing.<sup>9</sup> Hence, since today's technologies allow us to both "write" and "read" a DNA sequence, it is possible to draw inspiration from nature and use DNA as a data carrier.

For this purpose, the binary sequence corresponding to a file to be stored must be encoded into a DNA sequence (a sequence of nucleotides), which can then be synthesised

and stored. To retrieve the file, the DNA molecule must be sequenced and this sequence decoded.



Credit: DNA Data Storage Alliance<sup>11</sup>

DNA could offer a number of advantages as a data storage solution.<sup>2</sup> Firstly, its information-storing density is considerable: only about 50 atoms are needed to store 1 bit of data, whereas magnetic storage requires 1 million.<sup>2,10</sup> The longevity of DNA is another of its strong points: stored in appropriate conditions, a DNA molecule can be kept for tens of thousands of years, whereas data stored on conventional media needs to be recopied every 5 to 10 years to prevent deterioration. Moreover, as the support of genetic information, DNA is not subject to any risk of obsolescence and will remain a universal medium. Finally, it is very easy to duplicate data stored in the form of DNA with the help of mastered enzyme-based technologies.

Hence, among the various alternatives that have been proposed to address the problem of data storage, the use of DNA seems to be one of the most promising and is the first to have mobilised numerous academic and industrial players internationally.<sup>11</sup> However, there are still a number of technological challenges to be overcome before we see DNA really being used as an information carrier.

### ■ Current developments

The development of a storage system using DNA as a carrier involves a broad range of scientific disciplines<sup>12</sup> and therefore requires synergy between scientists, both academic and industrial, working on a variety of topics. Research is therefore currently hinged around collaborative and interdisciplinary projects, such as the US Molecular Information Storage (MIST) consortium,<sup>13</sup> the European OligoArchive<sup>14</sup> programme or the French dnarXiv<sup>15</sup> and MoleculArXiv<sup>16</sup> projects.

#### • Coding method

By defining the sequence that will be synthesised, stored and sequenced, the coding method has a direct influence on these different steps, and hence on the whole process. Consequently, the development of an efficient coding method (minimising the number of nucleotides required) will make it possible to achieve both a higher information density and a more efficient overall process (the synthesis and sequencing steps being both slow and costly).

However, the coding system must also be developed in such a way as to minimise the occurrence of errors during

the synthesis and sequencing steps, and therefore depends directly on the technologies used. To this end, the coding system must respect certain constraints regarding the sequence and proportions of the various nitrogenous bases,<sup>17</sup> and include control nucleotides for error detection and correction. It is also preferable not to use DNA strands longer than 200 nucleotides (for which synthesis errors are more frequent with current technologies)<sup>18</sup> and therefore to segment the file into different fragments. Thus, the coding method must include an indexing and addressing system to allow the file to be reassembled during the decoding stage.<sup>19</sup> Finally, this step also allows the encryption of the stored data to avoid any cybersecurity problems.<sup>20</sup>

#### • Synthesis

The most commonly used technique for synthesising DNA fragments is currently based on the principle of sequential synthesis, where nucleotides are added one by one. To prevent several nucleotides from being added at the same time, the nucleotides carry a protective group, preventing the addition of another nucleotide after it. The addition of a nucleotide to the chain under construction can be done in different ways: by chemical reaction (the synthesis is then called "chemical") or via an enzyme, a terminal deoxynucleotidyl transferase<sup>21</sup> (in this case it is known as "enzymatic" synthesis). After each addition, the protective group is cleaved so that the next nucleotide can be added in turn.

The chemical route, which is the traditional synthesis method, has now been automated and miniaturised; parallelization makes it possible to construct up to one million different DNA fragments comprising 200 nucleotides in 24 hours.<sup>22</sup> Although it is not currently marketed, the enzymatic route, is being developed by several companies (including the French company DNA Script, which leads the field in this area) and is expected to drive genuine progress: less polluting, it speeds up synthesis, reduces the cost and lowers the error rate.

Alternatively, rather than adding nucleotides one by one, short segments of pre-synthesised oligonucleotides can be assembled directly. Thanks to cohesive single-stranded ends, these segments can be linked to each other by the action of an enzyme: a DNA ligase. This method makes it possible to synthesise long DNA strands, while maintaining a low error rate and considerably speeding up synthesis. Using this method, the US company Catalog DNA has developed a machine capable of synthesising the equivalent of 500 KB/s.<sup>2</sup>

It is the synthesis stage that poses the greatest number of challenges that will need to be overcome if we are to hope to achieve widespread DNA data storage: its speed still needs to be substantially improved and the associated costs drastically reduced. While current efforts are mainly focused on massive parallelization of existing techniques, the development of new enzymatic processes could bring

significant progress. In addition, current technologies have been developed with a view to medical applications, and there is a difference in needs between these applications (which require a low error rate, even if it means compromising on issues of scale, cost and speed) and data storage (which can handle more errors, but where issues of scale, cost and speed are paramount). The recent emergence of companies totally dedicated to the use of DNA for storage<sup>23</sup> should make it possible to address these specific needs and develop appropriate solutions.

- **Storage**

To be preserved for long periods of time, DNA molecules need to be kept away from water and oxygen but also protected from light and high temperatures.<sup>24</sup> Conventional techniques to avoid DNA damage rely on low-temperature storage – which is both costly (in terms of space, equipment, energy and maintenance) and leads to a risk of loss in the event of malfunction – and do not take full advantage of the benefits of DNA as a storage medium compared to current technologies.

To address this situation, the French company Imagene has developed stainless steel capsules, containing a glass insert, for the storage of DNA molecules (which are first dried under vacuum and then placed in an inert atmosphere). Degradation extrapolations for this technology predict shelf lives of up to several tens of thousands of years at room temperature. The encapsulation process has also been fully automated and could therefore be easily integrated into a storage system.

The possibility of storing DNA fragments *in vivo* in living cells or organisms has also been studied. However, although some bacteria can survive for millions of years, the amount of information that can be stored in each host is limited, making it impossible to achieve high information densities using this method. In addition, this method of storage raises the issue of the tolerance of the fragments stored by the host organism, since DNA sequences may be toxic to the carrier or the surrounding biotope<sup>25</sup>. For these performance and bioethical reasons, *in vivo* storage has only been studied on an experimental scale, *in vitro* alternatives being preferred.

- **Sequencing**

In recent years, sequencing technologies have made prodigious progress, moving much faster than the pace described by Moore's Law,<sup>26</sup> highlighted in computer science. While it cost between US\$500 million and US\$1 billion to sequence the first human genome (completed in 2001),<sup>27</sup> the current cost is close to US\$700. Nevertheless, improvements – less significant than for synthesis – are still necessary to envisage the real development of DNA data storage

The most widely used technology today, developed by the Illumina company, is directly derived from the traditional method,<sup>28</sup> improved and massively parallelised. The DNA molecule of interest is first fragmented and converted into

single-strand DNA. After grafting and amplification of these fragments onto a solid surface, a DNA polymerase is used to construct the complementary strands in the presence of nucleotides carrying a cleavable fluorochrome (of a different colour for each nitrogenous base), which also serves as a protective group. After each addition, a digital photograph is taken to identify the nature of the added nucleotide. The fluorochrome can then be removed, enabling a new nucleotide to be added. Finally, the sequence is obtained by computer processing from the different images, using the overlaps to reorder the fragments.

An innovation that could become a real asset for DNA data storage and uses a completely different principle has recently been developed by Oxford Nanopore Technologies. By applying an electric field, the DNA fragments of interest – previously converted into single-stranded DNA – are drawn through a membrane containing nanopores. By measuring the ion flow through each nanopore, the nature of the nucleotide passing through it can be determined in real time. Although associated with a higher error rate,<sup>29</sup> this new method has proved to be much faster (it is currently possible to sequence up to 450 nucleotides per second) and eliminates the need for amplification and computer calculations to reconstruct the sequence.

- **Alternatives to the use of DNA**

The use of non-DNA polymers (i.e. polymers using non-nucleotide molecules as monomers) is also being studied for the storage of digital information.<sup>30</sup> The main advantage of this alternative lies in the freedom provided by the choice of monomers. These can be designed in such a way as to achieve a higher information density than in the case of DNA,<sup>31</sup> to give the polymer used greater stability or specific properties, or to facilitate the synthesis or sequencing steps which are currently limiting. While this approach appears to be promising and could eventually surpass the opportunities offered by DNA, it still requires significant research, as the synthesis and sequencing techniques for these polymers are not yet efficient and competitive compared to those for DNA.<sup>32</sup>

## ■ Outlook

- **Moderate impacts**

To become viable and widely used, the storage of information in the form of DNA must be fully automated and include synthesis and sequencing steps with measured costs and times. Although a first proof of concept was provided in 2019 by a team from the University of Washington and Microsoft, with a prototype capable of carrying out the entire process autonomously,<sup>33</sup> major advances still need to be made. It is currently estimated that costs need to be reduced by a factor of one thousand for sequencing and one hundred million for synthesis.<sup>2</sup> Although formidable, these targets should be seen in the light of recent advances in these sectors,

raising hope that significant progress may be possible in the coming years. Furthermore, the associated market is liable to experience significant growth, as estimated by BCC Research.<sup>34</sup>

In addition, the DNA writing and reading steps are expected to remain relatively slow compared to the technologies currently used for digital data storage. Therefore, DNA will be positioned – at least initially – as a storage medium for “cold” data, i.e. data that needs to be kept for long periods of time but does not need to be consulted or modified regularly.<sup>35</sup>

Since most stored data is very rarely consulted beyond a period of 100 days or so, “cold” data accounts for 60% of digital data.<sup>36</sup> However, magnetic tape devices, which are the main medium for storing such data at present, account for only a small proportion of the electricity consumed by data centres.<sup>2,36</sup> The use of DNA for “cold” data storage will therefore primarily bring benefits in terms of longevity and density, rather than energy.

- **Changes in practices required**

While the use of DNA is not expected to provide a data storage solution in terms of energy in the medium term, the opportunities provided by this technology could, conversely, lead to an increase in energy consumption through a “rebound effect” mechanism.<sup>37</sup>

Consequently, the development of data storage in the form of DNA will not dispense with the need for a reflection process on data consumption modes<sup>38</sup> and will have to be accompanied by good practice rules in order to determine which data is worth keeping in the long term (in particular for personal data, for which ethical issues<sup>39</sup> are added to environmental issues).

### ■ Conclusion

DNA data storage is an attractive technology that may be able to offer numerous benefits for data archiving. The development of a dedicated French industry should be encouraged, by promoting dialogue between the various stakeholders who have the necessary expertise.<sup>40</sup> It is very likely that the results obtained within the framework of such an approach could be used for various other applications (health, IT, etc.). It is nonetheless important to remain realistic about the real possibilities offered by this technology and about the environmental issues of data storage. Its development cannot be envisaged without consideration of digital consumption modes, which will require a drive towards greater sobriety.

*The Office's websites:*

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

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



## Persons consulted

---

Mr. Marc Antonini, CNRS research director at the Sophia Antipolis Laboratory of Computing, Signals and Systems

Mr. Raja Appuswamy, senior lecturer in Data Science at Eurecom

Mr. Pascal Barbry, professor in the genomic physiology of eukaryotes at the Institute of Molecular and Cellular Pharmacology

Mr. Dominique Lavenier, CNRS research director at the Institute of computing and random systems research

Mr. Jean-François Lutz, CNRS research director at the Charles Sadron Institute in Strasbourg

Mr. Thomas Ybert, CEO and co-founder of DNA Script

Ms. Sophie Tuffet, managing director, president of the board and co-founder of Imagene

Mr. Erfane Arwani, CEO and co-founder of Biomemory

Mr. François Képès, corresponding member of the French Academy of Agriculture, member of the Academy of Technologies, leader of the “ADN : lire, écrire, stocker l’information” (“DNA: reading, writing and storing information”) working group and co-author of the “Archiver les mégadonnées au-delà de 2040 : la piste de l’ADN” (“Archiving megadata beyond 2040: the DNA option”) report

Mr. Alain Bravo, member and president of the “éthique, société et technologies” (“ethics, society and technologies”) committee at the French Academy of Technologies, co-author of the “Big Data – Questions éthiques” (“Big Data – Ethical questions”) report

Mr. Louis Dubertret, member of the French Academy of Technologies, co-author of the “Big Data – Questions éthiques” (“Big Data – Ethical questions”) report

Ms. Anne Siegel, CNRS research director, deputy scientific director at the CNRS Institute of information sciences and their interactions (INS2I)

Mr. Jean-Marc Rietsch, international expert in dematerialisation, electronic signature and archiving, legal expert at the Paris Court of Appeal

## References

---

<sup>1</sup> European Commission, “European data strategy”, 2020 ([https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_fr#documents](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_fr#documents)).

<sup>2</sup> Académie des technologies, “Archiver les mégadonnées au-delà de 2040 : la piste de l’ADN” (French Academy of Technologies, “Archiving megadata beyond 2040: the DNA option”), 2020 ([http://academie-technologies-prod.s3.amazonaws.com/2020/10/15/10/17/37/9fbaaf5e-d5a6-4baf-921d-815da5d7983f/ADN\\_web.pdf](http://academie-technologies-prod.s3.amazonaws.com/2020/10/15/10/17/37/9fbaaf5e-d5a6-4baf-921d-815da5d7983f/ADN_web.pdf)).

<sup>3</sup> As regard the carbon footprint, it corresponded to 3.5% of global emissions in 2019, with the proportion expected to increase to between 5.5 and 6.9% in 2025. See: The Shift Project, “Impact environnemental du numérique : tendances à 5 ans et gouvernance de la 5G” (“Environmental impact of digital data: 5-year trends and governance of 5G”), 2021 (<https://theshiftproject.org/article/impact-environnemental-du-numerique-5g-nouvelle-etude-du-shift/>).

<sup>4</sup> The Shift Project, “Lean ICT : Pour une sobriété numérique” (“Lean ICT: Towards digital sobriety”), 2018 (<https://theshiftproject.org/article/pour-une-sobriete-numerique-rapport-shift/>). A study conducted in 2020 by the consultancy firms Citizing and KPMG for the Senate’s Sustainable Development and Spatial and Regional Planning Committee estimated that data centres accounted for 14% of France’s carbon footprint in 2019 (see endnote 5).

<sup>5</sup> “Pour une transition numérique écologique” (“For an ecological digital transition”), Information report 555 (2019-2020) by Messrs Guillaume Chevrollier and Jean-Michel Houllégatte, senators, produced on behalf of the Sustainable Development and Spatial and Regional Planning Committee by the Information mission on the environmental footprint of digital technology ([http://www.senat.fr/commission/dvpt\\_durable/mission\\_dinformation\\_sur\\_lempreinte\\_environnementale\\_du\\_numerique.html](http://www.senat.fr/commission/dvpt_durable/mission_dinformation_sur_lempreinte_environnementale_du_numerique.html)).

<sup>6</sup> The size of transistors is now close to the atomic scale, an insurmountable physical barrier. The performance of semiconductors (whose evolution is described by Moore’s law, see endnote 27) is therefore likely to level off in the next few years. See: 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/>).

<sup>7</sup> 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>).

---

<sup>8</sup> 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>).

<sup>9</sup> The sequencing of a DNA molecule consists in determining the order in which the nucleotides making up the molecule are linked together.

<sup>10</sup> This advantage seems particularly remarkable at a time when the amount of data stored is increasing exponentially and the issue of the space taken up by data centres is being raised. See: C. Diguët, F. Lopez, “*L’impact spatial et énergétique des data centers sur les territoires*” (“The spatial and energy impacts of data centres on territories”), ADEME report, 2019 (<https://librairie.ademe.fr/urbanisme-et-batiment/908-impact-spatial-et-energetique-des-data-centers-sur-les-territoires-l.html>). According to François Képès: “Today, data centres occupy a millionth of the Earth’s landmass. At the current rate of growth, one thousandth of the Earth’s landmass will be occupied by data warehouses by 2040”, whereas DNA “could store the whole of humanity’s current datasphere in a volume corresponding to the size of a small van. In 2040, it will be the volume of a truck.” See: 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-mondiale-dans-le-volume-d-une-camionnette-clame-francois-kepes-membre-de-l-academie-des-technologies.62299>). In addition, the density offered by DNA may prove to be particularly useful in terms of cryptography, since it is in fact easy to hide information in a long sequence of nucleotides.

<sup>11</sup> In October 2020, Illumina, Microsoft, Twist Bioscience and Western Digital created a “DNA Data Storage Alliance” designed to unite companies and institutions working in fields related to the use of DNA for data storage (<https://dnastoragealliance.org/>).

<sup>12</sup> This field includes mathematics and signal theory (for the development of an efficient coding system), chemistry and molecular biology (for DNA synthesis), genomics and IT (for the sequencing step), microfluidics and robotics (for the construction of automated devices).

<sup>13</sup> The US Molecular Information Storage (MIST) programme, funded by the Intelligence Advanced Research Projects Activity (IARPA) to the tune of US\$48 million (i.e. around 40 million euros), aims to be able to synthesise the equivalent of a terabyte by 2025 and to sequence the equivalent of 10 terabytes in 24 hours, at a cost of US\$1,000 (<https://www.iarpa.gov/index.php/research-programs/mist>).

<sup>14</sup> The European OligoArchive project is being funded for a period of 3.5 years by the European Union’s H2020 programme to the tune of 3 million euros (<https://oligoarchive.github.io>).

<sup>15</sup> The French dnarXiv project, funded by the French national research institute for digital sciences and technologies (INRIA) and LabEx CominLabs, involves several Brittany-based research laboratories (<https://project.inria.fr/dnarxiv/>).

<sup>16</sup> The MoleculArXiv project, led by the French national centre for scientific research (CNRS) and involving more than 20 laboratories, was awarded funding of 20 million euros over a period of 84 months in September 2021 in the context of exploratory priority research programs and facilities (PEPR). The 5-year objective is to be able to complete the read/write cycle at a rate of 1 bit/s (i.e. 10 GB of data in 24 hours) in order to be able to roll out demonstrators.

<sup>17</sup> The proportion of C and G bases should not exceed the A and T proportion and the successive repetition of the same nucleotide more than 3 times or of a pattern should be avoided. In fact, the simplest encoding system (00 → A; 01 → C; 10 → G; 11 → T) is not necessarily the most appropriate one. See: G. M. Church, Y. Gao, S. Kosuri, *Science* 2012, 337, 1628 (<https://science.sciencemag.org/content/337/6102/1628>).

<sup>18</sup> The longer a fragment is, the higher the risk of finding an error in one of its nucleotides. Currently, the error rate for each nucleotide is at best about 0.1%, so it is preferable not to use fragments of more than 200 nucleotides to keep the error rate acceptable.

<sup>19</sup> A system of this type is also essential to enable random access to stored data.

<sup>20</sup> In addition to the opportunities provided by the digital part of the process, the biological steps could also enhance the security of the storage system. The dnarXiv project (see endnote 16) is working on this aspect in particular. One of the potential options envisaged consists in using the primers as “keys”: without knowledge of the sequence of the primers, the amplification stage is impossible, preventing sequencing and therefore “reading”.

<sup>21</sup> Unlike most DNA polymerases, terminal deoxynucleotidyl transferases do not add nucleotides from a single-stranded template but in a random fashion. In the case of enzymatic DNA synthesis, nucleotides are added one by one (with a chemical protective group) to control the sequence.

<sup>22</sup> Performance achieved by Twist Bioscience, see endnote 3.

<sup>23</sup> There are currently a few companies that have chosen to focus their activities exclusively on this emerging field: Catalog DNA (USA), Helixworks Technologies (Ireland) and Biomemory (France).

<sup>24</sup> To these can be added irradiation, enzymes, microorganisms, ozone and various other air pollutants and xenobiotics. However, these sensitivities can also be seen as assets in the context of the storage of sensitive information, by enabling the instant destruction of information if necessary (notably thanks to the temperature).

<sup>25</sup> Apart from *in vivo* storage, the storage of data in the form of DNA does not seem to pose any ethical problems since the synthesised molecules are not introduced into cells and therefore likely to be interpreted.

<sup>26</sup> 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](https://www.eng.auburn.edu/~agrawvd/COURSE/E7770_Spr07/READ/Gordon_Moore_1975_Speech.pdf)).

- 
- <sup>27</sup> 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>).
- <sup>28</sup> F. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. USA 1977, 74, 5463 (<https://pubmed.ncbi.nlm.nih.gov/271968/>).
- <sup>29</sup> Although it is preferable to minimise the error rate, it is possible to correct some of these errors by using error-correcting codes, such as those used for compact discs, for example (to compensate for microscopic defects in the medium, fingerprints, scratches, etc.).
- <sup>30</sup> It is also possible to use unconventional nucleotides, in particular to enable a higher information density to be achieved.
- <sup>31</sup> Non-DNA polymers do not need to be made up of two strands and the monomers chosen can be smaller than the nucleotides, thus enabling a higher information density. In addition, the number of separate monomers that can be used is not limited, so it is possible to use a numbering system enabling more compact writing than the quaternary system provided by DNA (an octal or hexadecimal system, for example).
- <sup>32</sup> Furthermore, unlike DNA, the use of such polymers does not avoid the phenomenon of obsolescence, which is problematic for the storage of data over long periods.
- <sup>33</sup> C. N. Takahashi, B. H. Nguyen, K. Strauss, L. Ceze, Sci. Rep. 2019, 9, 4998 (<https://www.nature.com/articles/s41598-019-41228-8>).
- <sup>34</sup> BCC Research, “DNA Data Storage: Global Markets and Technologies”, 2020 (<https://www.bccresearch.com/market-research/biotechnology/dna-data-storage-market.html>).
- <sup>35</sup> Calculations by CNRS researchers estimate that a cost of €1 to write or read 1MB of data would enable DNA storage to penetrate the cold data market for valuable documents requiring rare access (e.g. contracts, property titles, laws). At a cost of €1 for writing or reading 1GB, molecular storage would become an attractive solution for data archiving, thanks to the volume reduction and increased durability offered. Finally, at a cost of €1 for writing or reading 1 TB, DNA would become interesting for (physically) transferring very large-volume data, with these data transfers currently limited by server capacities.
- <sup>36</sup> Horison Information Strategies, “Tiered Storage 2020: Building the Optimal Storage Infrastructure”, 2020 (<https://horison.com/publications/tiered-storage-2020>).
- <sup>37</sup> Data that would not otherwise have been stored might be kept because of the use of DNA as a storage medium.
- <sup>38</sup> F. Képès, “*Information et environnement*” (“Information and environment”), 2021 political and parliamentary review (<https://www.revuepolitique.fr/%ef%bb%bf%ef%bb%bfinformation-et-environnement/>).
- <sup>39</sup> Académie des technologies, “*Big Data – Questions éthiques*” (“French Academy of Technologies”, “Big Data – Ethical questions”), 2019 (<https://www.academie-technologies.fr/blog/categories/publications-de-l-academie/posts/big-data-questions-ethiques>).
- <sup>40</sup> The recent funding of the exploratory PEPR MolecuArXiv (see endnote 17) demonstrates a certain commitment to such an approach and should be applauded.