DATA STORAGE: THE DNA REVOLUTION

Speaker: Marc ANTONINI, Director of research at the CNRS - I3S Laboratory at the Côte d'Azur University | France

With **Casimiro VIZZINI,** Medical Doctor and Institutional Relations Officer at the One Sustainable for All Foundation | France

DNA data storage harnesses the potential of nature to stock digital information

DNA data storage is a revolutionary technology in both the fields of computer science and biology. This technology is not aimed at collecting genetic information but at leveraging the density and robustness of DNA to store digital data within it. The conversion of digital data from a binary format to a quaternary numbering system is the first step of the data encoding process. The quaternary format corresponds to the nitrogenous bases of DNA which are adenine, guanine, cytosine, and thymine. Once converted, the data is stored in DNA molecules. Those are then placed in small capsules protected from water, light, and oxygen. For the data to be retrieved, the DNA must be rehydrated, amplified with a PCR method, and sequenced. After being sequenced, the information in quaternary format is reformatted back into binary code for analysis. This technology allows biology to serve computer science and thus involves multidisciplinary interventions.

DNA data storage offers numerous benefits for data archiving and product traceability

Research projects on the topic of DNA data storage are ongoing in the United States and the European Union. France could potentially establish a complete sovereign industry within the national territory. This technology can be used to store "cold" data, i.e., archival data that does not require frequent or rapid access, and which often needs to be retained for legal reasons. With specific accreditations, molecular marking can also be used for product traceability (medicines, equipment, or paints).

DNA data storage will surpass the limitations of digital storage methods

This technology offers new opportunities for long-term information storage. Indeed, DNA and the digital data it contains can be stored, in a suitable environment, for hundreds or even thousands of years without being altered. In comparison, the lifespan of a hard drive is 5 to 10 years, and that of magnetic tape is 20 years. The DNA mini capsule is very compact and can store the information equivalent to at least one data centre. The entirety of the world's data could theoretically be stored in a space about the size of a shoebox. Hence, this storage method is also more environmentally friendly. Since DNA is stable and robust, there is no need to migrate data from one medium to another to avoid the risk of information loss.

Significant technological challenges are to be overcome to generalize the use of this technology

The main obstacles to the widespread adoption of this tool are related to its automation and its scalability. DNA synthesis is still very costly and considerably slower than that of a data centre. Storing more digital data, especially as its volume is exponentially increasing, requires an acceleration of the synthesis process. The MoleculArXiv research project aims at increasing the speed of synthetic DNA synthesis by a factor of 100 (equivalent to 1 megabit per second) within 3 years. Another goal is to catch up with the writing speed of hard drives by 2030. Improving data sequencing, so that information can be retrieved, integrated, and legible, is also a significant challenge. The issue of random access to any stored data must also be solved: how can one access a specific file without sequencing the entire content of the capsule?

As of today, the capsule has no access to the internet. However, in 5 to 10 years, significant security issues will need to be addressed. Indeed, automatic DNA storage systems connected to the internet will be used. Securing data will require cryptography for binary or a quaternary data signals. Internal private cloud with limited access for members of the organization that owns the data, could also be needed.

Finally, there are few ethical issues regarding this technology. Since the DNA is synthetic, only binary-encoded information can be stored in it, unlike genomic sequences or viruses.