

École Polytechnique, Palaiseau
Amphithéâtre BECQUEREL

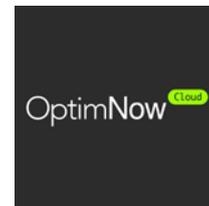
**Data Centers, IA, Cloud,
High Performance Computing.
Quel impact environnemental ?
Comment le mesurer et le contenir ?**

Jeudi 13 avril 2023



Coordination scientifique:

Christophe CALVIN (CEA) Jean LATIERE (OPTIMNOW)



Renseignements, programme...

[https://www.association-aristote.fr/evenements/
data-centers-ia-cloud-high-performance-computing/](https://www.association-aristote.fr/evenements/data-centers-ia-cloud-high-performance-computing/)

ARISTOTE

À la croisée des révolutions numériques



Compte Rendu du séminaire du 13 avril 2023

**« Data Centers, IA, Cloud, High Performance
Computing :
Quel impact environnemental ?
Comment le mesurer et le contenir ? »**

École Polytechnique
Amphithéâtre Becquerel

Organisé par Aristote

Table des matières

<i>1/ Introduction au numérique responsable alias Sustainable IT</i>	3
<i>2/ Retour apaisé sur l'énergie du numérique et sa médiatisation.....</i>	4
<i>3/ Vers une électronique soutenable : enjeux et perspectives.....</i>	5
<i>4/ Evaluer les impacts environnementaux du numérique, Open source vs opacité.</i>	6
<i>5/ Vers une prise en compte des enjeux énergétiques du numérique pour la génomique ..</i>	7
<i>6/ Impact environnemental des Large Language Models</i>	8
<i>7/ Amélioration de l'impact environnemental des grands modèles AI de langage (LLM). </i>	9
<i>8/ L'estimation de l'impact carbone et énergétique lors d'une visite sur un site web : focus sur l'impact côté utilisateur</i>	10
<i>9/ Évaluation du bilan carbone d'une machine de calcul nationale</i>	11
<i>10/ Stockage des données numériques sur ADN synthétique : avancées technologiques et challenges</i>	11
<i>11/ Vers un avenir numérique durable : le potentiel du stockage biosourcé sur ADN ..</i>	12
<i>12/ L'électronique de spin (spintronique) : une solution d'avenir pour un numérique frugale.....</i>	13
<i>13/ Feuille de route numérique responsable de Malakoff Humanis</i>	14
<i>14/ Cloud et sustainability: The path to 24/7 carbon-free energy</i>	15
<i>15/ Jean Zay : exemple d'un supercalculateur particulièrement efficient et écoresponsable</i>	16
<i>16/ ECLAIRION : Calculer plus vite pour un monde meilleur</i>	17

1/ Introduction au numérique responsable alias Sustainable IT

Emmanuel Laroche, sustainable IT, Airbus, et membre de l'Institut du numérique responsable.

Vidéo ici :

Emmanuel Laroche commence par recadrer l'impact du numérique sur le climat. Selon lui, différentes études convergent pour affirmer que le numérique représente 4% des gaz à effet de serre, dès lors qu'on raisonne en analyse de cycle de vie, et qu'on prend en compte l'impact des mines pour extraire les matières premières nécessaires à la fabrication des équipements. Mais le plus préoccupant, c'est la progression, avec une croissance de 6 à 8% par an. « Sans compter que le fameux chiffre de 4% est basé sur une étude de 2019, avec des données de 2017-2018, et doit donc aujourd'hui représenter davantage » prévient-il.

Selon lui, l'écoconception ne suffira pas, et il fait aller plus loin. Il détaille alors, les différents niveaux d'impact : en amont, lors de l'extraction, souvent pas pris en compte, lors de l'utilisation (marginale)

et enfin en fin de vie, où il relève la problématique du recyclage, une grande question, encore, à l'heure actuelle. Emmanuel Laroche a mis en place, chez Airbus, un programme de mesure et de réduction de l'impact écologique du numérique. Le Cloud, avec 21 data centers pour l'entreprise, représente 50% de l'impact carbone numérique de l'entreprise en scope 2 (sans les conséquences indirectes). Il indique avoir modélisé l'impact de l'utilisation des appareils numériques par extrapolation. Le but à moyen et long terme et de devenir « Sustainable by design ».

« Tout commence par la compétence et la formation : les sciences environnementales sont de vraies sciences, c'est primordial », indique-t-il. Mais il est aussi possible d'améliorer de nombreux paramètres d'utilisation d'applications de calculs internes, ou encore de mieux gérer ses data centers pour remplacer les vieillissants mal optimisés. Il présente alors les différents programmes qui ont été mis en place au sein de Airbus, à tous les niveaux.



2/ Retour apaisé sur l'énergie du numérique et sa médiatisation

Pierre Beyssac, fondateur d'Eriomem



[Vidéo ici :](#)

Pierre Beyssac veut surtout explorer la question de la médiatisation du numérique et interrogé ainsi sa perception dans la société pour mieux pouvoir in fine réduire son impact. Selon lui, les médias ne représentent pas la réalité de la question, en déformant les ordres de grandeurs. « Il a une couverture médiatique beaucoup plus élevée à mon sens que son impact relatif », estime-t-il. Ce qui aboutit à une mauvaise prise en compte politique du sujet, car ils écoutent ceux qui parlent le plus fort. Et les ingénieurs n'ont pas l'habitude de se mettre en avant politiquement. Il prend pour exemple certains sujets : « Les médias montent en sauce des sujets particuliers de manière erronée », estime-t-il. Comme le sujet particulier du streaming, qui a commencé à devenir catastrophiste lors d'un rapport du Shift Project en 2018 qui expliquait que « regarder 10 minutes de vidéo sur smartphone équivalait à 5 minutes d'utilisation d'un four à 2000 Watts. » Or, tous les instruments électroniques émettent de la chaleur mais l'ordre de grandeur paraît aberrant, malgré l'utilisation de routeur etc. Le Shift a d'ailleurs fait un Erratum, en affirmant se tromper d'un facteur 50, mais nettement moins médiatisé.

Pour lui, le souci vient de l'utilisation d'une méthode attributive. « Ils ont calculé que la vidéo représentait

10% des usages, donc 10% de l'impact du CO2 du numérique », sans calcul scientifique démontrant de linéarité entre le volume d'usage et le CO2 généré. L'inverse par exemple, n'est pas vrai. Si vous regardez moins de vidéo, vous ne coupez pas votre Wifi pour autant. Et la consommation de courant de votre box quand vous regardez de la vidéo ou non est sensiblement la même. Et c'est pareil au niveau de la chaîne, ces infrastructures sont essentiellement constituées de coûts fixes. « Le problème est qu'on en arrive à demander de la sobriété pour des raisons absurdes, qui n'auront pas d'impact », conclut-il. Car l'aboutissement de ce raisonnement consiste à remettre en cause la neutralité du Net !

Pierre Beyssac revient alors sur les différents chiffres réels de consommation d'appareils électronique. Il aboutit alors à la difficulté scientifique de modéliser convenablement l'impact d'un usage numérique. Car il y a beaucoup d'effets qui se conjuguent : les déchets, les coûts cachés, les effets de réseaux... Il cite alors le statisticien George E. P. Box qui estimait que « Tous les modèles sont faux, mais certains sont utiles ». Son point est donc de dire qu'il faut bien prendre en compte certains modèles, mais avant tout éviter ceux qui nous emmènent dans de mauvaises directions politiques. C'est ainsi qu'il détaille différents modèles : attributif vs conséquentiel, leurs avantages et inconvénients respectifs, afin de bien savoir de quoi l'on parle lorsqu'on mentionne les conséquences climatiques du numérique. Il termine en présentant quelques conseils pour le traitement médiatique de ce sujet : attention aux ordres de grandeur, être sensibilisé scientifiquement à l'impact du numérique, prendre en compte la fabrication,

se méfier des preuves par l'autorité, distinguer les scénarios des recommandations et enfin ne jamais oublier que le but doit être de favoriser la discussion et la rencontre.

3/ Vers une électronique soutenable : enjeux et perspectives

Thomas Ernst, directeur scientifique au CEA- Leti

[Vidéo ici :](#)



Thomas Ernst se concentre, lui, sur l'impact des Hardware. Après avoir donné des chiffres généraux sur l'usage de l'électronique entre 2010 et 2020, il invite à sortir à du franco-francisme, et rappelle que les solutions électroniques sont souvent déployées à échelle mondiale. Ainsi, il ne faut pas oublier que le numérique a aussi des effets positifs, sur les pays en voie de développement notamment. Il liste alors trois impacts positifs des usages numériques : l'éducation, la communication et l'accès au développement. En revanche, on a en parallèle trois impacts négatifs : celui de la fabrication, qui n'est souvent pas pris en compte dans la perception du numérique de par sa dimension mondiale (son impact carbone se situe bien souvent hors de la France), l'impact sur les écosystèmes, notamment via l'extraction minière, et la dépendance géostratégique qui peut donner lieu à des conflits. Il fait alors un lien entre la loi de Moore et la loi de Koomey, qui mesure elle, les calculs réalisés par Joule utilisés. En liant les deux, il aboutit alors au fait que le coût énergétique se situe davantage sur le transfert de données que sur le calcul en lui-même. Il conclut en donnant des pistes d'amélioration : selon Thomas Ernst, il faut changer le business modèle du secteur, en créant des entreprises davantage orientées sur le service que sur la production. Et il invite à renforcer la recherche sur le recyclage des matières premières afin de limiter les procédés d'extraction minière.

4/ Evaluer les impacts environnementaux du numérique, Open source vs opacité.

David Ekchajzer, fondateur de Hubblo et membre de Boavizta

Vidéo ici :

David Ekchajzer plaide pour une vision apaisée du numérique. Il présente le collectif Boavizta, qui est un groupe de réflexion autour des enjeux écologiques du numérique. Il met en avant le mode d'évaluation par Analyse en Cycle de Vie (ADV), qui est définie par un ISO et est multicritère, multi-périmètre, et multi-étape. Le collectif Boavizta milite pour une mesure en commun des impacts, et donc plaide pour mettre tous les acteurs autour de la table. Il souligne que l'évaluation est utilisée pour créer des lois. Elle est donc politique, c'est pourquoi il faut que le public soit capable d'auditer toutes les méthodes et tous les critères. Il doit se l'approprier. C'est donc la transparence qui importe ici. Or, aujourd'hui, ce n'est pas encore le cas. D'autre part, les évaluations



actuelles sont souvent de mauvaise qualité, et l'approche collective permet la mise en place d'une dynamique, et d'un processus d'amélioration continue. La non-transparence crée des problèmes, à l'instar des déclarations environnementales des fabricants d'écrans : les chiffres et les hypothèses retenues ne sont pas détaillées (il en fait la démonstration). La méthode de reverse engineering pour tenter de comprendre n'ont pas abouti. Les équipes arrivent finalement à des différences de mesures situées dans un rapport de 1 à 10 entre la plus petite évaluation et la plus grande, sans aucune explication acceptable. Il ajoute ensuite que l'évaluation est nécessaire car elle crée une dynamique dans les organisations collectives, et permet à chacun de s'intéresser au sujet en le rendant concret.

David Ekchajzer présente ensuite un exemple sur les équipements utilisateurs. Il remet alors en cause la méthode PAIA, souvent présentée par l'argument d'autorité « Développée par le MIT ». Cette méthode n'est finalement que très peu détaillée, et donne une grande largesse pour les fabricants d'utiliser les critères qu'il souhaite, ce qu'il déplore : « on est donc dans un immense far West sur le procédé d'évaluation en réalité », affirme-t-il. De la même manière, il revient sur l'usage des datacenters. Boavizta a mis en place une API permettant, à partir de données Open-Source, et de scrapping de données – une API qui mêle donc différentes méthodes - de pouvoir établir une évaluation de l'impact d'un data center, prenant en compte différents modes d'utilisation (mesure physique par Wattmètre, mesure logicielle, modélisation). Il prend soin de présenter les limites des modèles que le collectif a pris en compte. Il étend alors ses réflexions : selon lui, l'évaluation en ACV se rapproche davantage de la comptabilité environnementale que de la mesure physique et il faut bien distinguer les deux. Trop de paramètre et d'échelle rentre en ligne de compte. On peut ainsi avoir plusieurs approches : quelle est l'impact marginal d'un Gigaoctet sur un data center (ce sera de la mesure physique pure) et une approche conséquentielle qui consiste à mesurer l'impact de la généralisation de son service sur tous les data centers. Cette approche tient davantage des raisonnements sociaux en prenant en compte les effets rebonds, les effets d'empilement, les transformations sociales, économiques etc. Ainsi jusqu'à quel point peut-on combiner les approches et les modèles pour des effets politiques de l'évaluation ?

5/ Vers une prise en compte des enjeux énergétiques du numérique pour la génomique

Christophe Battail, CEA/DRF Nvidia Denergium

[Vidéo ici :](#)



Christophe Battail présente les problématiques liées à l'intervention croissante du numérique dans la génomique. Les travaux sur la génétique ont entrepris de séquencer les informations génétiques du vivant, que ce soit pour les bactéries, les plantes, ou les êtres humains, afin de comprendre les liens entre le développement de certaines maladies et le génome des individus. Cela nécessite d'analyser les informations génétiques de cohortes d'individus, et donc une quantité considérable de données numérisées (près de 3 milliards de caractères par individu). Il compare ainsi les quantités de données qui concerne la génomique et celles qui concernent l'astronomie. D'ici 2030, sauf ruptures dans ces sciences ou les technologies afférentes, la génomique devrait même dépasser l'astronomie en termes de quantité de données recensées. Il rappelle que le but n'est pas d'aboutir à une médecine personnalisée, un vœu pieux, mais à une médecine de précision, c'est-à-dire un traitement adapté à un groupe d'individus qui se rapprochent d'un point de vue génétique. D'un point de vue économique, on est passé d'un coût de 100 000 dollars le séquençage par individu en 2003 en production de données (année du premier séquençage total) à près de 1000 dollars.

Ces informations posent différents problèmes car elles peuvent être très sensibles en termes de vie privée. Le chercheur donne alors différents cas de figure : des enfants qui ne voudraient pas être au courant de maladie de leurs parents, ou ces données pourraient se faire hacker pour orienter les recherches médicales. Il détaille ainsi les différents points qui constituent la topologie des données génomiques : volume, vitesse, variété, véracité et valeur. Le nombre d'outils logiciels sont en constantes augmentation, avec parfois autant de bons comme du mauvais : surtout pour une utilisation dans un cadre clinique, qui demande des certifications que certains n'ont pas encore. Il revient alors sur le processus des projets de recherche dans le domaine, dont un des points sensibles est l'harmonisation des données. D'un point de vue des hardwares utilisés, étant donné le profil multidisciplinaire des chercheurs et chercheuses qui sont généticiens, informaticiens et mathématiciens, ils ne peuvent coder sur des systèmes trop complexes, et travaillent donc essentiellement sur des CPU. Les GPU commencent à arriver, cependant, du fait d'une sensibilisation à la génomique qui conduit des entreprises comme Nvidia à faciliter l'accessibilité de leurs outils. Concernant le sujet de l'impact énergétique des algorithmes et des travaux en génomique il pointe le paradoxe du secteur : la génomique cherche à mieux traiter les patients, mais peut dans ce cas contribuer au réchauffement climatique qui lui, aura un impact négatif sur la santé des gens. Le sujet est encore très récent : la première étude sur le domaine date de 2022. Il affirme alors que l'évaluation en termes de CO2 n'est pas encore « naturelle », et intuitive, et que trouver des comparaisons concrètes aide à prendre la mesure de l'impact réel sur

le climat. Le problème est complexe puisque, par l'exemple de l'a parallélisation, il note que si paralléliser les calculs sur différents CPU peut réduire le temps de calculs, il peut dans certains cas, augmenter fortement le nombre de cœurs CPU utilisés, et donc avoir in fine un impact plus négatif sur le climat qu'avec un temps plus long sur moins de cœurs. D'où la nécessité de traiter au cas par cas. Il détaille de la même manière l'impact sur la mémoire utilisée ou encore sur la localisation des calculateurs, selon le type de production électrique.

6/ Impact environnemental des Large Language Models

Julien Simon, HuggingFace

[Vidéo ici](#)

Julien Simon propose de zoomer sur l'impact écologique du machine learning.

L'émergence des modèles « transformers », en 2017, par Google, ont changé la face du machine learning, et ont battu tous les benchmarks actuels. Ces modèles travaillent sur différents types de données, et deviennent multimodales (texte, image, vidéo...) et donc ils se généralisent : dans les banques, dans le marketing, dans l'industrie... Mais le coût d'entraînement financier et environnementale ont fait plus de bruit. En 2019-2020, deux articles scientifiques ont été publiés posant la question sur la table « Mais les chiffres étaient un peu biaisés », affirme Julien Simon. Le hardware, les paramètres n'étaient pas les plus adaptés. Il y a quelques mois sont apparus les modèles « fondation », qui eux travaillent sur des dizaines de milliards de paramètres (1000 fois plus que les transformers). À partir de ces modèles, sont déclinés d'autres modèles qui eux s'occuperont de tâches particulières. Tous les Gafa en ont, mais aussi HuggingFace, qui a son modèle Bloom. Face aux manques de détails sur la réalité des choses, des moyennes ont été faites pour tenter d'évaluer l'impact de l'entraînement d'un modèle fondation comme GPT-3. En 2020, on parlait d'un mois d'entraînement sur 1000 GPU : soit 85 tonnes de CO2. « Mais pour aboutir à un modèle qui fonctionne, il faut en entraîner des dizaines voire des centaines », ajoute le spécialiste. « Globalement, un modèle de type GPT-3, il faut compter 100 GPU / an », détaillet-il. Une unité informelle qui donne un ordre de grandeur. Avec 170 000 modèles, HuggingFace s'est posé la question de son impact carbone, afin d'informer au mieux ses clients.

Le spécialiste revient alors sur les détails du modèle Fondation, Bloom, et de son impact carbone : en termes de consommation des GPU, le modèle arrive à 24 tonnes, et à 50 tonnes si on prend en compte toute l'infrastructure (réseau, stockage etc.) « Mais l'éléphant dans la pièce, ce n'est pas l'entraînement. Mais 70 à 80% du coût, ce n'est pas l'entraînement, c'est l'inférence », pointe Julien Simon. Chaque fois qu'on interroge le modèle, une fois déployé en entreprise, on fait tourner les GPU, et cela dépasse rapidement le budget d'entraînement. Une entreprise, peut abriter des centaines de modèles. Julien Simon invite alors à s'interroger et avoir conscience du modèle. Sur le choix du modèle, l'inférence, l'infrastructure, et la grille énergétique. Il faut intégrer la comptabilité du CO2 dans les projets et

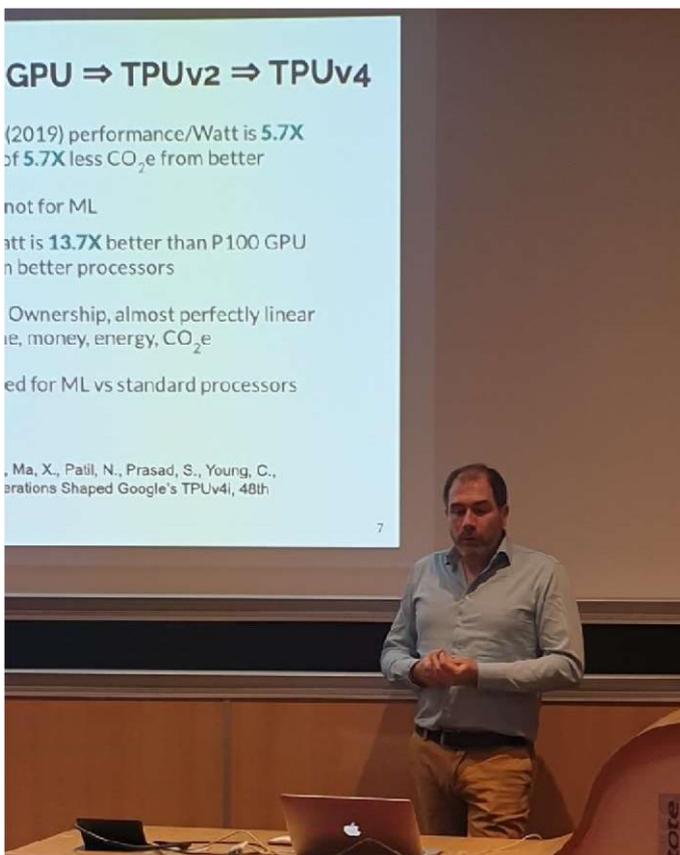


commencer à mesurer (exemple chiffre du coût d'entraînement du modèle affiché...) Il invite aussi dans le bon choix du modèle, à ne pas réinventer la poudre, et à prendre ce dont on a besoin dans les bibliothèques existantes (pour les entreprises). Idem, commencer par le plus petit, avant d'aller plus loin, sur le gros modèle. « Les petits modèles sont bien utilisés. Ils peuvent être optimisés à la marge (fine tuning), mais pas besoin d'aller chercher le plus gros pour une tâche précise » argue-t-il. Il invite enfin, en termes d'infrastructure, à utiliser le Cloud des géants, qui permet d'optimiser et de faire des effets d'échelles à la fois économiques, et à la fois écologiques. Idem, il souligne que les CPU, pour certaines tâches, en inférence, sont bien plus adaptés que les GPU et permettent de nombreuses économies. Ce que l'on a tendance à oublier dans la course à la performance.

7/ Amélioration de l'impact environnemental des grands modèles AI de langage (LLM)

Vincent Poncet, Google Cloud

[Vidéo ici](#)



Vincent Poncet revient sur les avancées de Google en termes d'amélioration de l'impact environnementale des modèles LLM d'intelligence artificielle. Quatre paramètres doivent être pris en compte : le type de modèle, les machines, l'infrastructure et la localisation géographique. L'impact peut être réduite d'un facteur 100 en fonction de tous ces paramètres. Il revient ensuite sur le fait que pour différentes évolutions des modèles transformés, les performances ont permis de gagner un facteur 4,2 en termes de temps de calculs, et donc de CO2 associés, pour un même résultat. Ensuite, sur un même modèle, en modifiant les processeurs, passant de GPU à des TPUv2 et des TPUv4 (des processeurs développés par Google), en quatre ans, les gains ont été d'un facteur 13,7. Il rappelle que le TPU chez Google, n'est né que pour faire des opérations de matrice. Elles ont été développées en 2015, pour le développement du Google Assistant, afin de répondre au problème d'inférence. Le projet ne pouvait pas être étendu à grande échelle sur des GPU

dans les data centers de Google. Il rappelle qu'en entraînement, il y a un cumul d'erreur, qui fait qu'on a besoin de puces précises, mais moins pour l'inférence. D'où la création des TPUv2. Il revient alors sur les caractéristiques techniques développées dans une démarche d'optimisation (format de données, gestion du coût par bits, parallélisation) sur les différents TPU. Il rappelle que ces innovations ne sont développées que pour un usage précis : le calcul matriciel. Elles sont donc incomparables avec un supercalculateur scientifique, mais pour un seul type d'opération, elles ont des puissances de calculs équivalentes à ces super-ordinateurs. Cette spécialisation permet des gains énergétiques importants.

Il évoque ensuite un modèle de machine learning développé uniquement pour optimiser le coût des modèles de machine learning. Il détaille ensuite les travaux effectués autour des data centers, dont la principale source de coûts CO2 se situe sur le refroidissement. Si la moyenne du marché est à 1.6 en termes de Power Usage Effectiveness (pour 1 Watt dans l'équipement informatique, combien de Watts nécessite le bâtiment), ceux de Google se situent à 1.1. Sans négliger l'équation du mix énergétique, dont le taux de carbone par kilowatt dépend énormément des régions. Il revient sur le fait que Google est un grand acheteur d'électricité bas carbone, en achetant en avance la production à venir. La société affirme favoriser ainsi le développement des énergies bas carbone. Idem, il développe l'optimisation qui a été effectuée via les énergies bas carbone en prenant en compte l'intermittence de Google. Vincent Poncet affirme alors qu'en prenant en compte tous les critères d'optimisation, entre 2017 et 2022, l'impact écologique a été réduite d'un facteur 1000. Il précise ensuite qu'entre tous les paramètres, il faut faire attention aux extrapolations, notamment en ce qui concerne le nombre de paramètres, car il n'y a pas de relation linéaire entre le nombre de paramètres et la consommation énergétique.

8/ L'estimation de l'impact carbone et énergétique lors d'une visite sur un site web : focus sur l'impact côté utilisateur

Guillaume Lochon, GreenMetrics

Vidéo ici :

Greenmetrics est une start-up spécialisée dans la mesure et l'optimisation de l'impact carbone du numérique. C'est une équipe de sept personnes doublées d'un comité scientifique pour encadrer les pratiques. Elle propose trois outils : la réduction de l'impact d'un parc numérique, celle des sites internet et celle de la publicité en ligne. Guillaume Lochon, data scientist pour l'entreprise présent alors en détail le modèle destiné à la mesure de l'impact d'un site internet. Si différents modèles existent déjà, Greenmetrics a tenté de pousser le curseur plus loin pour améliorer son modèle. « Une visite sur un site par un utilisateur possède trois sources d'impact différentes : celle des data centers qui hébergent les données, celle de la transmission, filaire ou sans fil, et celle, plus délicate à mesurer, de l'impact du terminal de l'utilisateur. » Les deux premières sources bénéficient déjà de données précises par les constructeurs que l'entreprise a récupérées. C'est donc sur la dernière source, principale zone d'ombre, que l'entreprise a planché. Entre l'impact de la batterie, du microprocesseur, et la variation du temps de consommation du site, les paramètres se multiplient. « Soit il faut mesurer sur chaque terminal son utilisation précise, mais les données sont confidentielles, soit il faut établir un modèle », ajoute le scientifique. Greenmetrics a alors réalisé un algorithme, qui permet de calculer l'impact en fonction des principaux paramètres de consommation d'un site web, puis en calculant, pour chaque type d'appareil, l'équivalent



CO2 total émis en fonction du temps que l'on reste connecté. Le point est donc de mesurer également l'impact de l'utilisation du site internet sur le terminal et le temps passé de l'utilisateur et de l'intégrer à la mesure totale.

En fonction du pays d'origine et du mix énergétique employé, l'impact carbone change énormément (électricité nucléaire, durable ou charbonnée). Guillaume Lochon présente alors les résultats : dans l'exemple d'un site web d'e-commerce textile, la consultation du terminal représente 70 % de l'impact carbone totale du site web.

9/ Évaluation du bilan carbone d'une machine de calcul nationale

Gabriel Hautreux, CINES

[Vidéo ici :](#)

Gabriel Hautreux présente le travail autour de l'évaluation du bilan carbone du CINES (le Centre Informatique National de l'Enseignement Supérieur), effectué en 2021, et notamment la partie qui concerne le super ordinateur (les autres concernaient l'archivage et l'hébergement). Ce travail a été réalisé en partenariat avec GRICAD – UGA. Deux ordinateurs seront présentés : l'actuel, Occigen, et le ordinateur en projet, Adastra.

Après avoir détaillé les caractéristiques du ordinateur Occigen (3366 nœuds de calcul, 85 824 cœurs CPU, 34 racks, 5Po de stockage scratch) le spécialiste donne toute de suite le résultat du bilan : une heure.coeur équivaut à 2,64 grammes de CO2 (pour un mix énergétique français). Le ordinateur représente donc à lui tout seul 70% des émissions. Il détaille ensuite, dans le modèle, les paramètres les plus émissifs : « La partie émission des nœuds de calcul et infrastructure sont liés au ordinateur, elle est donc la plus émettrice pour l'usage, avec près de 90 % d'imputation aux émissions. » La partie nœuds et équipement des racks sont le plus gourmands : 28% pour les nœuds et 50% pour le châssis, les racks, et le power supply. Il faut ensuite ajouter 16% des émissions pour l'infrastructure d'alimentation et le refroidissement du ordinateur.

Il revient alors sur la durée de vie : plus on fait durer l'équipement, plus on amortit les émissions dues à la fabrication, donc l'impact carbone diminue (mais l'amortissement diminue au cours du temps). Mais les équipements plus récents sont mieux conçus et moins gourmands en énergie : in fine, les courbes finissent par se croiser. Il n'a pas encore de réponse exacte à la durée optimale d'utilisation de l'ancien ordinateur.

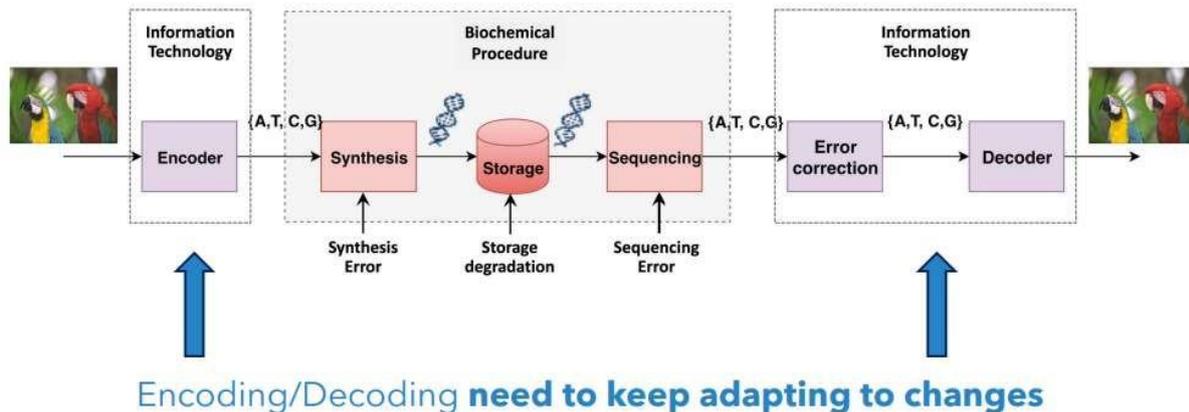
Il présente ensuite le ordinateur Adastra, qui lui sera hybride, avec des CPU et des GPU. Après avoir fait le même bilan, il compare les deux : Adastra, à performances équivalentes, a réduit d'un facteur 6 la densification de l'infrastructure, et d'un facteur 3 la consommation d'énergie.

10/ Stockage des données numériques sur ADN synthétique : avancées technologiques et challenges

Marc Antonini, chercheur au CNRS

[Vidéo ici :](#)

Marc Antonini présente un projet qui a eu lieu au niveau européen intitulé OligoArchive, dont le but est de pouvoir stocker de l'information dans de l'ADN. Il commence par présenter les ordres de grandeur de stockage des données. Aujourd'hui, on considère que de façon théorique, on peut stocker sur un gramme d'ADN, 250 pentabytes de données (soit 250 millions de gigabytes) l'équivalent d'un data center, donc, dans une capsule qu'il présente dans la pièce. Il présente alors les avantages du stockage dans l'ADN : compact, durable, et écologique. En revanche : la vitesse d'écriture est encore très lente, et très cher. Il présente alors le détail du procédé de synthèse des données sur de l'ADN.



D'un point de vue technique, tout est faisable : coder, stocker et décoder. Le principe consiste aussi à réparer les erreurs que ce soit au niveau du séquençage, ou du décodage. Il faudra donc parfois réparer le code. Cette discipline dont l'idée est née dans les années 50, est réellement née en 2012, a progressé nettement en 2015, avec la gestion des erreurs. C'est ce qu'on appelle le transcodage. Il présente alors les défis de la matière : la gestion des erreurs, la synthèse de polymères qui permettront d'aller au-delà du code quaternaire (ACTG), et la structuration de la donnée, pour pouvoir aller chercher un fichier précis dans la capsule. Il faudra aussi contrôler le compromis coût/qualité du processus, l'application à tous les types de données ou les contraintes biochimiques. Il insiste sur l'aspect pluridisciplinaire du domaine : biologiste, biochimique, informaticien, mathématicien, généticien... Il faut donc fédérer toute cette communauté.

11/ Vers un avenir numérique durable : le potentiel du stockage biosourcé sur ADN

Erfane Arwani, Biomemory

[Vidéo ici](#)

Dans la poursuite de la présentation précédente, Erfane Arwani cherche à démontrer l'avantage structurel de l'archivage de données sur ADN. Il présente donc différents ordres de grandeur en présentant de l'ADN dans une capsule (une capsule représente environ quatre hexa octets). Aujourd'hui la masse de données numériques existante est autour de 100 zettabytes dans le monde. Le but des recherches dans le stockage par ADN est d'arriver à atteindre le zêta dans une seule capsule. Cette avancée serait une grande réponse à la question de tous les DSI qui consiste à devoir arbitrer les données

qu'ils doivent conserver. Le chercheur cite alors Microsoft qui dit que si nous sommes capables de conserver aujourd'hui 30% des données, nous ne pourrions en garder que 3% d'ici 2030 à ce rythme.

D'autre part, les data centers sont composés de disques durs et de carte SSD, des composants électroniques qu'il faut changer. Et la plupart sont détruits. Il présente donc les deux types de stockage qui existent aujourd'hui : les stockages électroniques (peu durables) ou non électroniques qui deviennent peu durables ni possibles face à la quantité de données. Il parle alors du besoin de révolution dans le stockage. Il dresse le bilan des solutions, et notamment le stockage moléculaire, qui est peu avancé à l'heure actuelle (à Strasbourg). Et le stockage sur ADN avec les oligos, la solution industrielle acceptée à l'heure actuelle. Il en présente alors les caractéristiques et avantages.

Il revient alors sur les enjeux autour de la biologie moléculaire qui consiste à trouver dans le vivant les mécanismes qui pourraient servir au stockage de données. Entre l'ADN, le plasmide (présente dans la bactérie *Escherichia coli*), il décrit alors les pistes de réflexion qui ont amené au disque dur ADN.

Il arrive alors sur les conséquences de tout ce processus, et notamment la problématique du coût écologique et du coût financier. Cela l'amène à détailler le processus précis de réalisation des encodages sur les plasmides de stockage, et il présente alors le projet de réalisation « d'encre ADN », qu'il ne peut détailler pour des questions de propriétés intellectuelles, qui se réalise sans jamais utiliser de pétrole, avec donc un coût écologique amoindri. Le but consiste à créer des bases de nucléotides à partir du sucre, très simplement et de les assembler dans un milieu aqueux. In fine, on obtient un double brin d'ADN qui peut aller jusqu'à 30 000 bases.

Il conclut enfin en présentant les questions liées à la latence des hardware moléculaires, qui ne pourront jamais descendre en dessous de la seconde.

12/ L'électronique de spin (spintronique) : une solution d'avenir pour un numérique frugale

Kevin Garello, CEA

[Vidéo ici](#)

Kevin Garello est à distance du fait des grèves. Il présente les travaux de Spintec, une unité mixte de recherche via vidéo conférence, concernant la « Spintronique », dérivé de Spin – élec – tronique, qui est une technique consistant à utiliser dans un courant, non seulement sa charge mais aussi son moment magnétique : le spin. « Cela permet d'avoir non seulement un courant de charge, mais aussi un courant qui va se polariser avec une certaine orientation de spin », explique-t-il. Cela permettra plus tard, en le mettant en interaction avec un matériau ferromagnétique, d'avoir la capacité de manipuler ce matériau. Ainsi en injectant ces courants dans des matériaux possédant des couches magnétiques et des couches non-magnétiques, on observe des variations de résistances très fortes, qu'on appelle magnétorésistance, qui peut atteindre de très fortes valeurs (ou très faibles). Ce phénomène a été découvert par Albert Fert et Peter Grünberg, qui ont reçu le Prix Nobel en 2007. En effet, il a eu beaucoup d'implication, car cela permet ainsi dans un même matériau, suivant que l'on injecte un courant, d'obtenir des variations de résistances très fortes ou très faibles selon les couches et donc d'obtenir des 0 ou des 1 et de réaliser des capteurs sensibles (le spinival). Aujourd'hui, de nombreux systèmes embarqués des composants basés sur ce phénomène, pour des capteurs de positions, par exemple, dans des téléphones, des voitures, ou pour le médical. Une deuxième révolution a eu lieu, la Magnetic Random Access Memories, MRAM's. En injectant des courants dans des éléments appelés des Jonctions Tunnels Magnétiques (MTJ), on peut changer la valeur d'un champ magnétique, et les propriétés d'aimantation d'un matériau,

sans avoir à créer de champ de magnétique, en transférant un couple. « On converti une charge en spin via un matériau ferromagnétique ». Ainsi, on peut transmettre une information à un matériau et créer une mémoire. Kevin Garello donne alors les détails sur la réalisation des matériaux qui permettent cette capacité, et que l'industrie maîtrise aujourd'hui très bien.

Outre le fait que cette technologie a d'autres applications, au-delà des capteurs ou des mémoires, il revient en détail sur l'utilisation de ces MRAM's, et des avantages qu'elles apportent pour répondre aux problèmes posés par la miniaturisation des mémoires (coût de fabrication, courant de fuite des composants, fiabilité du fait d'une trop grande quantité de composants par surface...)

Ces technologies NVM (Non volatile Memories), plus dense, permettent de revoir également l'architecture du système, notamment les aspects de communication entre le cache et la mémoire principales (memory wall). Elles permettent également de développer de nouveaux concepts de calculs. Il revient alors en détail sur les types de mémoire non volatiles, et leurs spécificités techniques.

Enfin, il précise que l'utilisation des MRAM's pourrait réduire les communications entre le cache et la mémoire standard, ce qui diminue le coût en énergie et en temps d'utilisation des composants, mais répond aussi à la tendance actuelle qui consiste à effectuer des calculs localement (in memory computing, near memory computing...) tous reliés à l'intelligence artificielle.

C'est en cela que les MRAM's et les technologies de mémoire non volatile qui utilise les technologies spintronique répondent à différents besoins adaptés au numérique de demain.

13/ Feuille de route numérique responsable de Malakoff Humanis

Marie-Anne Clerc, directrice des systèmes d'information d'Humanis-Malakoff

[Vidéo ici](#)

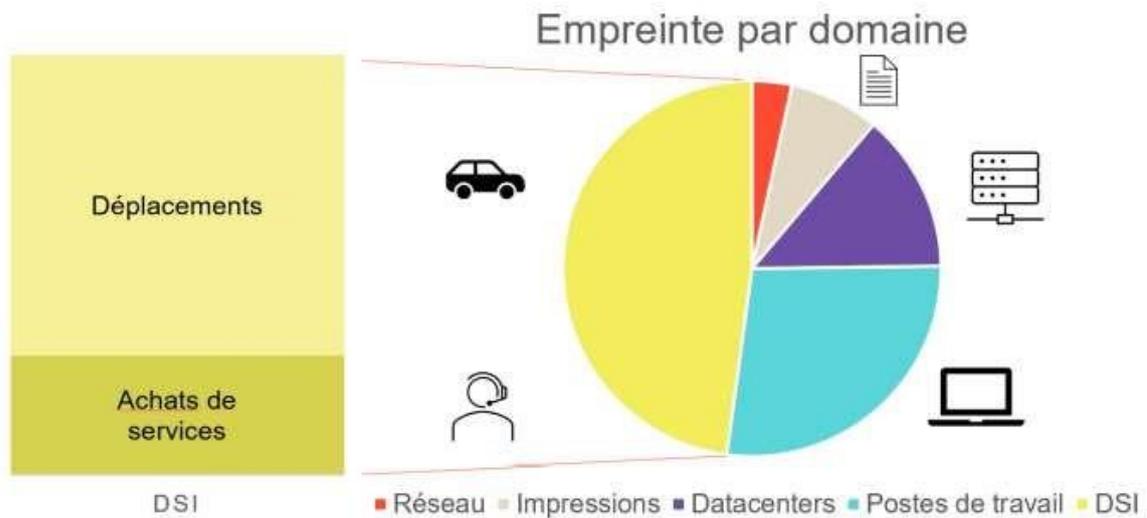
Marie-Anne Clerc revient sur les projets d'Humanis Malakoff en termes de numérique responsable, en tant que groupe d'assureurs, l'IT est au cœur du business. Elle commence par préciser que toutes les ambitions en matière de Green IT au sein du groupe ne sont pas forcément portées par le management, mais davantage issue de convictions personnelles.

Elle insiste sur le fait que le diagnostic est primordial pour évangéliser sur le sujet, car il ne parle pas encore forcément à tout le monde. L'entreprise a travaillé sur une analyse en cycle de vie, avec le Collectif Green IT, « Si beaucoup d'ESN disent savoir-faire, beaucoup en réalité ne savent pas le faire convenablement », estime-t-elle. Le logiciel utilisé s'appelle Swift, elle rappelle que toute analyse comporte évidemment des biais et des raccourcis, mais qu'il faut le savoir. In fine, l'empreinte environnementale de la DSI est mesurée et déclinée en fonction des 4 principaux indicateurs d'impacts : les gaz à effet de serre, l'eau, l'énergie, les ressources.

Elle donne ensuite les résultats : l'impact climatique des collaborateurs de la DSI, notamment par les déplacements, est de 512kg eq CO2. Même si on fait peser le poids de chaque collaborateur sur la DSI, ce qui est imparfait, cela permet de comparer entre DSI. Elle relève que le principal facteur d'impact environnemental est le déplacement des collaborateurs (post covid, donc télétravail inclus). « Le déplacement des collaborateurs et la mise en place de la mobilité douce reste le principal sujet aujourd'hui dans la plupart des entreprises », estime-t-elle.

L'impact du poste de travail également, avec un grand nombre de double écran (2,5 écrans par personne) est aussi très élevée.

Elle fait le détail ensuite de son empreinte par domaine au sein de la DSI :



En rouge, le réseau, puis l'impression (favorisé par la demande des clients, à ce titre, elle précise que 25% des Français n'ont pas d'ordinateurs, donc que la dématérialisation ne résout pas tout), en violet les data centers, sachant que les fournisseurs ont travaillé et travaillent encore beaucoup dessus. Ensuite vient le poste de travail, extrêmement polluant – sa fabrication notamment.

Elle témoigne alors du management pour faire passer ces messages en interne. « Lorsqu'il s'agit de supprimer les vieux mails, c'est assez sexy et les gens vous suivent, mais dès que l'on parle du non renouvellement d'outils numériques, ou de supprimer un écran, qui vont à l'encontre d'un certain confort, c'est beaucoup plus complexe », témoigne-t-elle.

Elle revient alors sur la politique mise en place en termes de gestion du matériel, portée par ses convictions et les relais qu'elle a au sein de l'entreprise : une gestion des équipements orientée autour des 4 R : Réduire, Réutiliser, Réparer, Recycler. Cela inclut : Fournitures de smartphones reconditionnés, Arrêt de la politique d'attribution systématique de seconds écrans, Extinction automatique des postes de travail le soir après 22h, Pilotage optimisé des écrans des salles de réunion, Extinction automatique des écrans de communication après 19h, réflexion autour de l'extinction des serveurs hors prod le soir et le week-end, gestion de la climatisation des locaux techniques, éco-conception des sites internet... Ainsi qu'une meilleure communication et sensibilisation autour de ce sujet.

14/ Cloud et sustainability: The path to 24/7 carbon-free energy

Anne-Elisabeth Caillot, Infrastructure Modernization and Data Management - Practice Lead, Google Cloud

[Vidéo ici :](#)

Anne-Elisabeth Caillot rappelle que Google a comme ambition de devenir net zéro carbone 24 heures sur 24 et sept jours sur sept, en 2030. Elle va donc présenter la démarche et les bonnes pratiques de l'entreprise pour y parvenir. Elle présente ensuite la réalité industrielle de Google :

35 Cloud Régionaux, dont six en projets, pour un total de 18 TWh de consommation électrique (2GW de consommation instantanée). Mais face à ces chiffres, cela permet de jouer avec l'effet d'échelle. Chaque avancée et chaque pourcentage de gagné représente in fine quelque chose de très important à l'échelle. Elle revient ensuite sur la démarche engagée par Google depuis 2008 et la première construction de son propre Data Center, pour mieux optimiser sa consommation selon ses besoins, et de les développer pour les seules fonctions de l'entreprise, mais aussi établir une meilleure gestion des eaux selon leur disposition géographique, ou encore revoir l'infrastructure logiciel pour mutualiser au maximum les travaux sur les machines. Elle détaille également la politique de recyclabilité des machines et des data centers.

Côté énergie, l'entreprise utilise à 100% des énergies renouvelables depuis 2017, sur la consommation annuelle. Elle présente alors la politique d'achat et de production d'énergie de Google sur la planète.

Elle précise alors, par l'exemple du Chili, le mode de calcul du « 100% énergie renouvelable » revendiqué par Google, en détaillant la différence entre le calcul horaire, et le calcul de consommation de masse sur l'année, ainsi qu'avec la compensation faite par les investissements additionnels de Google dans les énergies renouvelables. Avec l'utilisation de batteries, qui stockent les surplus, l'entreprise finit par arriver à 100% pour le data center du Chili.

D'autres pistes sont en cours de travaux, comme l'utilisation de la géothermie, ou l'utilisation de l'IA pour optimiser les services et faire tourner certains calculs pendant les périodes de moindre consommation énergétique. Deepmind travaille également sur les prévisions du vent pour mieux anticiper les périodes de fabrication d'énergie éolienne. Tout cela pour devenir Carbon net zero à horizon 2030.

A force de travailler ce sujet, de réaliser des partenariats, Google a développé un savoir-faire que l'entreprise met à disposition de ses clients, comme un outil de mesure de l'impact environnementale par utilisation de Google Cloud Platform, et permettre de piloter sa propre utilisation pour minimiser son impact. « Il faut préciser ici que l'on parle de mesures réelles, et non d'estimations », ajoute-t-elle.

15/ Jean Zay : exemple d'un supercalculateur particulièrement efficient et écoresponsable

Rafael Medeiros, responsable adjoint production, CNRS/IDRIS

[Vidéo ici :](#)

Rafael Medeiros va présenter la politique de numérique responsable du supercalculateur Jean Zay, situé à Orsay, créé en 1993 sur le campus de Paris Saclay.

C'est un centre de calcul majeur pour le CNRS, pour le calcul de haute performance (HPC) mais aussi de plus en plus à l'IA (environ 60% vs 40%). En termes de chiffres, le supercalculateur représente 36,85 Pflap/s, 89976 cœurs CPU, 2712 GPU (NVIDIA V100), 440 GPU (NVIDIA A100) et 468 To mémoire.

Après avoir présenté quelques caractéristiques techniques du supercalculateur Jean Zay, Rafael Medeiros présente les évolutions récentes de l'infrastructure qui est passé de Babel en 2008, à Turing en 2012, pour devenir Jean Zay en 2019. En 12 ans, la performance a été multipliée par 33, la puissance de crête par 265 et la consommation par seulement 8.

Sur un diagramme de présentation de la consommation électrique depuis 2014, nous voyons bien la différence lors de l'arrivée de Jean Zay, au premier trimestre 2019, et ses extensions en 2020.

Les équipes ont travaillé avec Ecoinfo pour exécuter le bilan carbone de l'infrastructure. Rafael Medeiros relève que les fabricants ne donnent que très peu d'information sur la fabrication des calculateurs (GPU) ce qui ne permet que de faire des estimations sur le bilan global du supercalculateur, car la fabrication représente la majeure partie du bilan carbone, face à l'usage. Selon la structure du calculateur, il reste possible d'estimer le bilan carbone de chaque calculateur. Les équipes sont arrivées ainsi à 1,41 gramme de CO2 par heure d'utilisation, pour les CPU, 41,79 grammes de CO2 par heure pour les GPU (x4) et 35,54 pour les GPU (x8).

A l'heure actuelle, sa forte dotation en GPU (plus efficace) ainsi que son système de refroidissement par eau tiède (30 degrés entrée, 36 degrés en sortie), ainsi que son système de récupération de chaleur pour le bâtiment, et prochainement pour le campus Paris-Saclay, en font un des supercalculateurs les plus efficaces en production. Il aurait un PUE théorique de 0,65. (1,21 sans prise en compte de la récupération de chaleur.)

Rafael Medeiros présente alors le mécanisme de récupération de chaleur pour Paris Saclay en détail. Au total, le système permet de récupérer 4000 méga watt heure par an, soit l'équivalent des besoins en chaleur de 1000 logements neufs.

Il revient alors sur le système énergétique global du campus.

16/ ECLAIRION : Calculer plus vite pour un monde meilleur

Charles Huot, président de Cap Digital et chargé du projet ECLAIRION

[Vidéo ici :](#)

Charles Huot présente le projet Eclairion, qui est un projet d'hébergement de supercalculateur en containers, près de Bruyères Le Châtel, et qui s'établira sur quatre hectares, dont trois opérationnelles. Le but est de créer un centre sans bâtiment, toutes les infrastructures sont disposées dans des containers, sur une plateforme. « Les technologies de data centers en containers sont déjà bien éprouvées », indique Charles Huot. Le premier intérêt c'est la modularité, et la rapidité dans le déploiement, et une efficacité énergétique importante.

Le président de Cap Digital revient alors sur les capacités techniques du projet : Hébergement haute densité de 30KW à plus de 200KW, pour une puissance totale de 60 MW. Le site permet d'héberger 160 containers pour un total de 10 000 mètres carrés, un refroidissement à air ou par immersion (il n'y a pas d'imposition pour les clients, le projet reste souple selon les besoins), et une bonne connectivité par fibre optique.

Il présente alors l'impact environnemental du site : tout béton a été supprimé. La seule composition résidera sur des poutrelles en acier sur lesquelles sont disposées les containers. La quantité de terre déplacée a été minimisée et sera réutilisée, l'éclairage pour les animaux nocturnes, et les caméras infrarouge pour ne pas gêner la vision des reptiles, ont été adaptées. Tous les matériaux utilisés seront entièrement recyclables. En termes d'énergies, on sera sur du 100% décarboné, 50% renouvelable. Des réflexions sont en cours pour récupérer la chaleur pour un projet hôtelier dans un ancien château réaménagé.

L'ouverture est prévue fin 2023.