

Covid-19 Museum

Compte rendu rédigé par ANDSI & Pierre Delort

En bref...

Yves ROZENHOLC présentera le projet Covid-19 Museum, véritable outil électronique dédié au suivi et à la compréhension des impacts de la pandémie sur la société. Après avoir exposé les objectifs du projet, il détaillera notamment son fonctionnement ainsi que son architecture.

L'Association Nationale des Directeurs des Systèmes d'Information organise des débats et en diffuse des comptes-rendus, les idées restant de la seule responsabilité de leurs auteurs. Elle peut également diffuser les commentaires que suscitent ces documents.

Yves ROZENHOLC explique avoir imaginé le concept de Covid-19 Museum le 20 mars 2020, soit quatre jours après le début du confinement. Ce projet a été mis en ligne le 2 avril 2020 et est disponible sur l'adresse <https://covid-19.museum>.

Introduction

La crise sanitaire du Covid-19 correspond à la première pandémie de l'ère numérique. Pour la première fois, il est possible de suivre, comprendre, comparer au sein de chaque pays l'évolution d'une maladie émergente, les réactions sociétales qu'elle entraîne qu'elles soient individuelles, politiques, médicales, et ce, à travers le matériel numérique produit. Cette pandémie a conduit à l'émergence d'un corpus électronique sans précédent, notamment à travers une créativité quotidienne ou une réflexion revêtant la forme de recherches, de textes ou d'interviews.

Les objectifs du projet

Le projet propose d'unifier les collectes des traces numériques ou numérisables au sein d'une base de données sans déplacer ou copier ces traces et leurs métadonnées puis d'en offrir un accès facilité et interconnecté. La construction de ce musée électronique sera favorisée par le développement de la volonté participative et l'aide du soutien public.

Le musée sera accessible par tous et pour tous à travers un site web pour dépôt et consultation, avec un accès à tout ou partie de la base de données possible sur requête pour recherche. L'idée est de couvrir les aspects culturels, éducatifs et de recherche en lien avec la période à travers la construction d'un outil didactique, analytique et mémoriel permettant de suivre et comprendre l'impact du Covid-19 sur la société.

Un tel outil permettra de réfléchir aux modèles sociétaux actuels et d'en imaginer de nouveaux. En effet, cette pandémie a induit des changements comportementaux. Ce musée pourra aider à la construction de modèles sociétaux mieux préparés, plus résilients.

Le fonctionnement : principes généraux

Ce musée repose sur une brique informatique intitulée [CaMPA], pour *Collaborative archive for Memory, Pedagogy and Analysis*. Une [CaMPA] correspond à un logiciel en *open source* déployable sur toute ressource d'archivage ou de calcul virtuel.

Cette brique élémentaire du Covid-19 Museum contient une ressource d'archivage et/ou de calcul ainsi qu'une interface d'accès et de requête. Elle donne accès aux diverses briques interconnectées. Son interface reprend les éléments d'identité du musée et précise par qui cette [CaMPA] est opérée. Elle donne accès aux traces digitales dédiées à la pandémie, protégées par l'acteur qui l'a installée sur sa collecte. La [CaMPA] est développée en logiciel libre et pourra être modifiée et installée par les acteurs internationaux comme l'UNESCO, l'Europe, etc. ; dans chaque pays (villes, universités, musée, etc) mais aussi chez les particuliers qui souhaiteraient à titre individuel participer comme pour les projets *seti@home* et *folding@home*.

Schématiquement, le Covid-19 Museum unifie des personnes qui collectent, des personnes qui mettent à disposition des ressources de calcul et des personnes qui effectuent de la recherche.

En synthèse, le Covid-19 Museum est un outil d'analyse au service des sociétés humaines respectueux des acteurs qui le constitue. Il participe notamment de l'unification de façon ouverte des collectes de traces numériques liées à la pandémie, de l'opportunité d'analyser de façon reproductible ces traces ou de la protection de la connaissance qui en sera extraite. En revanche, il n'organise pas de collecte et ne sera pas ainsi juge et partie.

Exemple d'application

A titre d'illustration, il est possible de considérer les dessins réalisés par des enfants pendant la période de confinement et de les rapporter au concept de [MPA] ou *Memory, Pedagogy, Analytic*. Le fait de les déposer dans le Covid-19 Museum aboutit aux éléments suivants :

- [M] : les dessins forment un corpus mémoriel ;
- [P] : la symbolique des dessins exprime *a posteriori* le ressenti des enfants pendant le confinement ;
- [A] : les trajectoires d'évolution des dessins deviennent source de recherche (IA, psychologie, etc.)

La constitution d'un tel musée permet de favoriser le questionnement sociétal et de capitaliser sur les expériences qui ont fonctionné.

Principes et détails

Le projet veut respecter les principes de la science ouverte et favoriser la recherche reproductible dans le contexte des sciences humaines et sociales.

Les codes sources permettant de générer la base de données et le site web sont développés sous une licence *open source* qui autorise le partage (copier, distribuer et communiquer le matériel par tous moyens et sous tous formats) et l'adaptation (remixer, transformer et créer à partir du matériel pour toute utilisation, y compris commerciale). Le code est déposé et développé à partir d'un serveur de type GitHub. Les groupes de programmeurs suivent les directives de la direction de l'équipe technique du musée et participent aux choix techniques au travers de forums dédiés à chaque tâche. Les objets y sont déposés sous licence de type *Creative Commons*.

De surcroît, l'accès privilégié sur requête suivra les principes de la recherche reproductible. Une trace exacte de l'accès sera archivée au sein du musée et permettra la reconstitution de l'accès *a posteriori* sans avoir à dupliquer les données concernées et la vérification par un relecteur de la conformité des résultats d'une prépublication déposée sur le serveur du musée.

Le musée contient un serveur de *preprints* en libre accès pour consultation. Les produits réalisés après requête en accès au musée doivent y être déposés. Assez tôt, les mathématiciens ont imaginé la construction de serveurs de *preprints* sur lesquels les chercheurs pouvaient déposer une version avant publication, même si ces travaux restaient pendant plusieurs années en *reviewing*.

Lors d'une requête en accès privilégié, le requérant s'engage à déposer sur le serveur de *preprint* du musée une copie du produit de sa recherche réalisée à partir des données auxquelles il a accès. Cette trace ainsi que les prépublications réalisées à partir de l'accès privilégié deviennent partie intégrante des données protégées par le musée. S'agissant du RGPD, le Covid-19 Museum ne sera pas hébergeur de données, mais offre l'opportunité de la fouille. Il suit une RGPD simplifiée.

Il convient ensuite de mentionner la nécessité d'une *blockchain* pour garantir :

- la traçabilité des dépôts et de leur origine ;
- la qualité de ceux-ci ;
- le lien entre l'objet et ses métadonnées ;
- l'usage fait des objets ;
- la protection de l'accès en requête, du code utilisé et de leur produit.

Par ailleurs, la muséographie et le système de mots-clés de chaque acteur sont protégés par le Covid-19 Museum. L'accès privilégié sur requête permet l'usage des ressources des sciences de données. Les projets *open source* d'apprentissage statistique ou d'IA y sont disponibles. Le Covid-19 Museum permet de construire le lien entre les empreintes numériques d'un objet et l'objet lui-même. C'est là une occasion d'offrir au grand public et aux chercheurs non spécialistes la capacité de mieux comprendre les enjeux de l'Intelligence Artificielle.

Architecture réseau

Techniquement, l'interconnexion par la [CaMPA] va suivre les liens d'Internet. Elle permettra une extraction de connaissances à impact écologique réduit. La copie des traces numériques est limitée. Les calculs sur les traces numériques sont réalisés au plus près de chaque trace. Seuls les résultats sont agrégés.

Une association a été constituée autour de ce projet. Parmi les co-fondateurs, il convient de mentionner Elisabeth Belmas, Professeur émérite d'histoire moderne ; Christophe Cérin, Professeur d'informatique ou Xavier Coumoul, Professeur de biologie, Juliette Vion-Dury, Professeur de littérature générale et comparée ainsi que Gérard Ben Arous, Professeur de mathématiques de New York University.

Les acteurs qui se sont intéressés à ce projet peuvent être des institutions académiques à l'instar d'universités, mais également des entreprises (Orange, Dassault Systèmes, etc.). A noter également l'intérêt du Musée de l'Éducation Nationale ou des Archives régionales comme celle des Vosges.

Yves Rozenholc enfin explique qu'il cherche de l'aide auprès de notre communauté pour développer ce projet.

Débat

Int : Concernant l'IA citée, envisagez-vous, pour l'entraînement des algorithmes, un cadre de Split-Learning, de Federated Learning..., étant donné que les Data demeurent chez les producteurs ?

YR : En effet. Ce point est fondamental. Vous ne pouvez pas demander à l'INA ou à la BnF de communiquer une copie des données.

Int : Disposez-vous de visibilité concernant le planning du projet ? Plus particulièrement, à quel moment comptez-vous avoir un premier résultat visible pour les particuliers ?

YR : Au moment où j'ai imaginé le projet, j'envisageais une durée de deux ans. Il serait positif qu'un tel projet fonctionne au bout de dix ans. Il serait également positif de bénéficier d'une bêta version d'ici cinq à huit ans.

Par ailleurs, nous conduisons actuellement une négociation avec l'ANRS MIE à propos d'un projet intitulé « Covid-19 soignants » et dans lequel la brique [CaMPA] serait intégrée. De surcroît, le *board* de Time Machine Europe souhaiterait que le Covid-19 Museum devienne un point d'entrée de Time Machine Europe.

Int : D'autres incitatives Nord-Américaines ont-elles été prises ?

YR : J'ai constaté une initiative nord-américaine réalisée par des *assistants professors*, mais celle-ci veut seulement documenter des liens web vers les collectes. On peut douter que ces collègues aient les moyens de réaliser leur projet.

Une vidéo de présentation de Time Machine est projetée.

Il existe déjà des dizaines de milliers de collections en France et des centaines de milliers dans le monde. L'idée avec Covid-19 Museum est de refuser de choisir. Il est indispensable que les historiens de demain disposent du plus grand nombre de données possibles afin d'effectuer leur travail. Une telle boîte à outils offre de nombreuses applications.

Int : Cette absence de choix suppose que vous acceptiez tous les volontaires.

YR : Le risque est qu'un groupe de personnes soit en mesure de diffuser une information fausse dont le poids est suffisamment important pour faire basculer la vérité. Mathématiquement, il est possible de contrer ce problème jusqu'à 50 %.

Int : Je constate que de nombreux acteurs de la data ou du numérique s'intéressent à ce projet, mais aucun acteur médical. Je suppose que vous devez être en rapport des laboratoires ou d'autres organismes médicaux.

YR : Pendant un an, je n'ai pas contacté l'INSERM et l'APHP car les acteurs médicaux étaient saturés. A ce jour, le premier acteur qui pourrait apporter de l'argent pour construire un démonstrateur de la [CaMPA] correspond à l'ANRS MIE.

Présentation des orateurs

Yves Rozenholc est professeur en science des données à la faculté de Pharmacie de l'Université de Paris. Il dirige l'UR 7537 BioSTM « Biostatistique, Traitement et Modélisation des données biologiques » qui réunit des enseignants-chercheurs tous tournés vers les interactions pluridisciplinaires. Son parcours va des mathématiques appliquées aux applications des mathématiques pour les sciences du vivant.