

Claire CARLIN
University of Victoria
ccarlin@uvic.ca

Ça dépend : Le défi de la conservation des projets numériques

En 2016, le Conseil de recherches en sciences humaines du Canada a subventionné *Endings : Concluding, Archiving, and Preserving Digital Projects for Long-Term Usability*. Notre équipe (trois chercheuses, trois programmeurs et trois bibliothécaires, tous de l'Université de Victoria en Colombie Britannique) finit de construire en ce moment des « boîtes à outils » axées sur les pratiques exemplaires en conservation numérique. La conscience croissante de la fragilité des sites et autres projets numériques souligne l'urgence de cette démarche.

Nos études de cas ont confirmé une évidence: les questions relatives à la pérennité numérique sont d'ordre technique d'une part et d'ordre humain de l'autre. Du côté technique, nos programmeurs ont déjà présenté leurs stratégies de base dans des communications et des articles¹ : il s'agit fondamentalement de créer une version statique de chaque page d'un projet, réécrivant les applications particulières en utilisant seulement des technologies dont l'avenir est plus ou moins assuré (HTML5, CSS et Javascript).

Le défi le plus complexe a pourtant à voir avec la nature humaine. Pour avoir une idée plus précise des obstacles à la conservation numérique, nous avons effectué un sondage à choix-multiple en 2017-2018. Sur les 139 réponses, nous avons sélectionné 25 collègues à consulter par Skype. Une dizaine de thèmes ressortent de ces discussions, dont les questions relatives à l'hébergement des projets à long terme. Les enjeux relevés sont emblématiques de toute une gamme de surprises, la plupart du temps mauvaises : attentes peu réalistes, inconscience des dépendances, incompréhension de l'ampleur du défi et ainsi de suite. Les pièges rencontrés par les collègues illustrent la difficulté de voir venir les embûches même si on est très expérimenté en humanités numériques.

Voici quelques exemples phares extraits de nos conversations avec les collègues interrogés :

La précarité de l'hébergement \propto la précarité de l'emploi

Ce sont souvent les collègues détenant un doctorat relativement récent et qui n'ont pas de poste permanent qui innovent en humanités numériques – et s'ils obtiennent un poste, les projets initiés lors de leurs études ne sont pas toujours accueilli par l'université qui les embauche ; rares sont les collègues qui peuvent faire du soutien d'un projet numérique le *sine qua non* de l'acceptation d'un poste. Plutôt que d'abandonner leurs projets, ces chercheurs ont souvent choisi de confier leur travail à un serveur non-institutionnel, augmentant le risque de sa disparition. **Maintenant Zenodo ou GitHub représentent des solutions plus sûres, mais** la précarité ne disparaît pas pour autant.

¹ Voir <https://projectendings.github.io/papers/> et <https://projectendings.github.io/articles/>.

Il faut des fonds

L'absence d'un financement pour l'hébergement à long terme représente un défi auquel presque tous les chercheurs ont tôt ou tard affaire. Faute de fonds institutionnels, plusieurs collègues subventionnent l'hébergement de leur travail de leur propre poche. Le maintien général et la mise à jour du code rajoutent des dépenses supplémentaires. Une entreprise de conseil technologique a proposé un prix de 6 000\$/an à une collègue qui cherchait un soutien en dehors du contexte universitaire – deux fois rien dans le monde des affaires mais la fortune pour les enseignants-chercheurs.

À qui la propriété intellectuelle ?

Un autre scénario courant parmi les collègues consultés est la perte de l'accès à leurs propres données lors d'un changement de poste. Les collègues qui n'examinent pas soigneusement les contrats d'hébergement peuvent faire face à une mauvaise surprise quand ils changent d'emploi. Qui est le propriétaire d'un site ou d'un projet créé avec les ressources d'une institution quelconque ? Pour éviter ce genre de dilemme, il incombe aux chercheurs de garder une copie de leurs données, conservés dans un endroit sûr sous leur propre contrôle. Une autre solution est de partager régulièrement les données parmi les membres de l'équipe dans la mesure du possible, par exemple en se servant d'un système de contrôle de versions. L'équipe « Endings » était plutôt étonnée d'apprendre combien de collègues ne possédaient pas de copie de leur propre travail.

Les remous de la politique institutionnel

Même si on arrive à établir l'hébergement d'un projet « en permanence », un changement de politique est plus que probable sur la longue durée car un projet peu entretenu crée une vulnérabilité dans le système où il réside. Un projet actif à l'heure actuelle peut être transféré plus tard dans un coin invisible d'un dépôt (« dark storage ») sans que les auteurs en soient avisés, surtout si le contenu ne correspond plus aux priorités institutionnelles. Les morales de l'histoire : l'archivage dans plus d'un endroit, et (encore une fois) un contrat écrit plutôt qu'un accord verbal pour l'hébergement à long terme (une période à définir avec clarté). Pour citer un des collègues consultés, « Il ne m'est jamais venu à l'esprit que l'université qui hébergeait le projet dès le début le laisserait tomber ; j'ai pris pour acquis le soutien à long terme ».

Une infrastructure peu sûre

Ce ne sont évidemment pas toutes les institutions académiques qui ont la capacité financière et technique d'héberger des projets d'ampleur. La mésestimation de la quantité et du genre de ressources nécessaires à l'entretien à long terme d'un projet numérique est assez typique. Un scénario courant dans nos entretiens était la dépendance sur un technicien individuel : si cette personne quitte son poste, celle qui la remplace sera-t-elle capable de reprendre le projet en main ? Sans une infrastructure solide dans toutes ses branches, les projets numériques sont vulnérables.

L'équipe évoluera

Les démissions, les mutations et autres promotions sont quasiment inévitables dans le monde universitaire actuel.

Plusieurs projets dépendaient d'assistants de recherche lors de leur construction sans avoir bien documenté les décisions prises par ces étudiants, qui sont maintenant introuvables.

Le travail virtuel permet en principe à une équipe de continuer à travailler à distance, mais la réalité des obligations professionnelles des uns et des autres mène souvent à l'abandon d'un projet, surtout si les responsabilités ne sont pas déterminées avant que l'équipe ne se disperse. Dans le contexte de l'hébergement par l'institution où le projet est né, il y avait cinq cas parmi les 25 où il ne restait personne de l'équipe d'origine, d'où la mise en question de la présence du projet sur les serveurs universitaires.

Le projet éternel

Plusieurs d'entre les collègues consultés lors de notre enquête ne voyaient pas venir la fin d'un projet déjà en cours depuis plusieurs années. Dans ce cas-là la planification d'une solution à long terme pour son hébergement était loin d'être leur priorité. Nous avons appelé notre projet « Endings » précisément parce que finir est un des plus grands défis du travail numérique. Trouver une – ou idéalement plusieurs – solutions pour l'hébergement à long terme devrait devenir la première question à régler plutôt que la dernière.

Recherché, mort ou vivant

Un projet fonctionnant à partir d'applications « boutiques », d'outils spécialement créés ou d'une plateforme particulier risque de connaître de plus en plus de pannes au cours du temps s'il n'est pas entretenu. Dans les cas où un projet dépend en partie d'outils ou de logiciels créés par les grandes sociétés ou autres entreprises (e.g. Google Earth), il faut savoir que le projet dépend entièrement de leur bonne volonté pour rester opérationnel.

Un site internet en libre accès est normalement la forme de diffusion préférée des créateurs d'un projet numérique, mais certains d'entre collègues interrogés ont décidé que la meilleure stratégie à long terme est de conserver seulement leurs données de base, sans trop se soucier de leur rendu. Cependant, la perte de lisibilité est parfois suivie de l'impossibilité totale de récupérer les données. Utiliser des dépôts comme GitHub ou Zenodo représente une solution pour la conservation basique, mais ils ont des limites importantes par rapport à la quantité (50GB maximum pour Zenodo) et à la fonctionnalité (n'étant pas des serveurs Web). Les collègues européens ont l'énorme avantage de pouvoir faire héberger et archiver leurs données dans le « très grande infrastructure de recherche » Huma-Num (<https://www.huma-num.fr/>) qui offre des services numériques pérennes.

Pour sa part, « Project Endings » vous propose de garder le beurre et l'argent du beurre grâce aux stratégies décrites à la page « Principes » de notre site : <https://projectendings.github.io/principles/>. Le code est également consultable à <https://github.com/projectEndings/>.

Conclusions

Même si une bibliothèque universitaire ou autre institution accepte l'hébergement « permanent » d'un projet, il n'y a pas de garantis qu'il soit accessible et utilisable sur la longue durée. Notre stratégie fournit à peu de frais une chance pour une cinquantaine d'années à venir, dans la quasi-certitude que HTML5, CSS et Javascript continueront à être lisibles.

Malgré la création d'infrastructures exemplaires dans certains pays et dans certaines institutions, avant que les sciences humaines entrent davantage dans l'ère numérique les chercheurs universitaires doivent absolument acquérir le savoir nécessaire à la conservation de leurs données à long terme. « Project Endings » a pour but de participer pleinement à ce mouvement, et en tant que programme canadien, de diffuser les résultats de nos recherches en français aussi bien qu'en anglais. Veuillez noter que nous dirigeons un numéro spécial de *Digital Humanities Quarterly*, à paraître en 2022.