

BESOCIAL

Towards a sustainable social media archiving strategy for Belgium

Contract - B2/191/P2/BESOCIAL

Résumé

Contexte

L'archivage des réseaux sociaux (SMA) n'est ni nouveau ni inconnu. Au cours de la dernière décennie, les réseaux sociaux, qui sont des *big data*, sont devenus l'une des principales sources de recherche quantitative (et qualitative). Il a été crucial de révéler des informations fascinantes sur une variété de questions relatives au comportement humain. En plus de collecter les données des réseaux sociaux uniquement à des fins de recherche, les institutions culturelles telles que les bibliothèques nationales s'efforcent également d'archiver et de préserver ce type de données qui évolue rapidement, en créant un cadre pour l'archivage du web et/ou des réseaux sociaux.

De juillet 2020 à septembre 2022, la Bibliothèque royale de Belgique (KBR) s'est engagée avec l'UGent, l'UNamur et l'UCLouvain dans un partenariat visant à développer une stratégie durable pour l'archivage et la préservation des réseaux sociaux en Belgique. Ce projet de recherche BESOCIAL a été financé par le Politique Scientifique Fédérale Belge (BELSPO) dans le cadre de son programme BRAIN.be. La Bibliothèque royale de Belgique (KBR) était le coordinateur de ce projet qui a été géré en étroite collaboration avec le CRIDS (Centre de recherche en information, droit et société) de l'Université de Namur, CENTAL (Centre de traitement automatique du langage), de l'UCLouvain, IDLab (Internet Technology & Data Science Lab), GhentCDH (Ghent Centre for Digital Humanities), et MICT (Research Group for Media, Innovation and Communication Technologies) de l'Université de Gand. L'interdisciplinarité du réseau de recherche a permis de couvrir les aspects techniques, juridiques et opérationnels ainsi que les aspects liés aux besoins des utilisateurs et a favorisé l'enrichissement mutuel au sein du projet.

Objectifs

Le projet de recherche BESOCIAL a été divisé en un certain nombre de lots de travaux (WP) et de tâches (T) décrivant une approche étape par étape pour le développement d'une stratégie d'archivage durable des réseaux sociaux pour la Belgique.

Le premier objectif du projet, "passer en revue les projets d'archivage des réseaux sociaux existants en Belgique et à l'étranger", était lié au WP1, où quatre tâches dédiées visaient à fournir un état de l'art (inter)national concis de l'archivage des réseaux sociaux (SMA). Deux objectifs supplémentaires du projet BESOCIAL étaient de mettre en place des pilotes pour l'archivage des réseaux sociaux et de fournir un

accès à l'archive des réseaux sociaux. Le WP2 a servi de phase de préparation à la mise en place d'un pilote pour l'archivage des réseaux sociaux dans le WP3, et d'un pilote pour fournir l'accès aux archives des réseaux sociaux dans le WP4.

Les résultats finaux de ces objectifs ont été consignés dans une section de recommandations en 6 parties (WP5) avec le cadre légal, les exigences techniques et fonctionnelles, un modèle d'entreprise ou *business model*, une définition de SMA en Belgique, l'intégration institutionnelle de la SMA à KBR et la définition des procédures.

Tout au long du projet, les résultats (préliminaires) et les processus ont été partagés au niveau national et international sous forme de participation à des conférences ou de leur organisation, de présentations et de publication d'articles (WP6). Une autre tâche cruciale en cours a été la mise en place d'un service d'assistance juridique où le partenaire juridique de BESOCIAL a offert une consultation continue concernant les questions principales et pertinentes sur la vie privée et le droit des TIC rencontrées au cours du projet. Cela a conduit à un document interne de FAQ résumant ces questions et réponses concernant le SMA.

Méthodologie

Objectif 1 : Analyse des projets d'archivage des réseaux sociaux existants en Belgique et à l'étranger : pour étudier la représentation du paysage belge et international du SMA, nous avons adopté une approche de recherche descriptive. Cela a impliqué un certain nombre d'étapes : trouver des initiatives qui collectent des données de réseaux sociaux via une recherche secondaire, collecter des données de ces initiatives via une recherche documentaire, des enquêtes et des entretiens semi-structurés, et analyser et interpréter les données, en utilisant une approche d'analyse thématique qualitative.

Objectifs 2 et 3 : Mettre en place des pilotes pour l'archivage des réseaux sociaux, et fournir un accès aux archives des réseaux sociaux : pour la phase de préparation (WP2) et la phase d'exécution (WP3 et WP4), nous avons utilisé une approche de méthode mixte. Un cadre juridique a été créé à l'aide d'une recherche documentaire. Des entretiens semi-structurés ont été utilisés pour mieux comprendre les besoins des utilisateurs. Nous avons également expérimenté avec certains outils au cours des études de faisabilité afin de commencer la collecte des données. Une méthode expérimentale a également été mise en œuvre pour la création d'une plate-forme d'accès pilote.

Conclusions

Objectif 1 : Analyse des projets d'archivage des réseaux sociaux existants en Belgique et à l'étranger : La cartographie du paysage belge et international du SMA est une première étape vers la compréhension et l'identification des caractéristiques communes des différentes approches d'archivage des réseaux sociaux appliquées. Nos résultats ont montré que de nombreuses institutions sont engagées dans le SMA, mais que le stade d'avancement et les efforts varient à la fois en taille et en portée.

Globalement, nous pouvons conclure que les choix et les différentes étapes du processus d'archivage des réseaux sociaux dépendent des ressources. Il y a trois pierres d'achoppement : i) le temps nécessaire pour

explorer le SMA à côté des tâches en cours au sein des institutions culturelles, ii) les connaissances (techniques) en interne, iii) le manque de ressources humaines et financières.

En outre, le cadre juridique joue également un rôle important dans les processus du SMA. En ce qui concerne les considérations juridiques qui constituent des obstacles à l'évolution de la société numérique, on peut dire que cette évolution s'est produite à un rythme si rapide qu'il a été difficile pour le droit de suivre les derniers développements.

Au niveau international, les tendances communes suivantes ont été identifiées : i) Twitter est la plateforme de réseaux sociaux la plus souvent archivée ; ii) les collections de comptes et/ou de hashtags se concentrent sur des personnes, des organisations et des événements importants ; iii) la priorité est donnée aux collectes sélectives ; iv) les outils open source sont le plus souvent utilisés ; v) le format de données WARC est le plus souvent utilisé, et enfin, vi) un manque évident de compréhension commune des concepts de préservation (par exemple, "formats de préservation" ou "normes de préservation") a été observé.

Au niveau belge, nous avons conclu que Facebook est le plus souvent archivé, suivi par Twitter et Instagram. Cette préférence est motivée par l'aspect thématique ; les organisations constatent que les comptes qu'elles veulent capturer sont plus susceptibles d'être trouvés sur Facebook. Nous avons également notifié que le nombre de comptes et/ou de hashtags est bien équilibré entre le nombre de plateformes de réseaux sociaux. Dans le cas où l'on souhaite archiver un grand nombre de comptes et de hashtags, les organisations optent pour moins de plateformes. Nos résultats ont montré que la plupart des collectes sont effectuées au moyen d'outils open source, l'accent étant mis sur la capture de l'aspect visuel et de la convivialité d'un compte.

Au fil du temps, le concept de réseaux sociaux et la façon dont la société utilise ces plateformes vont sans doute considérablement évoluer. Il reste à voir si les institutions seront en mesure de surveiller et de suivre ces changements. Sur la base de ces conclusions, les praticiens devraient envisager de mener des recherches de suivi en mettant à jour la vue d'ensemble belge et internationale afin de mieux comprendre les implications de certaines décisions du SMA pour l'avenir.

Objectif 2 : Mettre en place des pilotes pour l'archivage des réseaux sociaux : En menant plusieurs études de faisabilité dans le WP2, nous avons constaté que le projet BESOCIAL devrait se concentrer sur le texte, principalement pour éviter les implications juridiques et techniques. Twitter et Instagram ont été mis en avant comme les plateformes à collecter avec l'aide des outils Social Feed Manager et Instaloader. La sélection du corpus a été déterminée par une politique double (T2.1) : i) l'équipe de BESOCIAL a créé plusieurs listes de sélection (*seed lists*), et ii) l'avis du public a également été pris en compte. Pour ce dernier, une campagne de crowdsourcing a été mise en place, qui a donné lieu à de nombreuses suggestions ainsi qu'à une publicité pour BESOCIAL, l'archivage des réseaux sociaux en Belgique et KBR en général. Le thème des collections portait sur le patrimoine culturel en Belgique.

Dans la phase de collecte (WP3), des directives ont été créées pour surmonter la technicité des outils utilisés (T3.1). Une meilleure compréhension des besoins et des exigences (T2.2) d'un large éventail de parties prenantes lors de l'utilisation d'une archive de réseaux sociaux a également été obtenue. Les

conclusions suivantes ont été tirées : i) l'existence des archives web (de réseaux sociaux) est mal connue, ii) il est nécessaire d'inclure le domaine universitaire dans les décisions et les politiques de sélection, iii) nous devons nous mettre d'accord sur la manière dont le contenu archivé doit être consultable, et iv) nous devons ouvrir les références à des méthodologies spécifiques ou à des logiciels ou outils spécifiques. Ces observations ont été prises en compte lors du contrôle de la qualité du contenu des données récoltées (T3.2). Notre analyse montre que i) une expérience préalable des fichiers .csv ou .json, ou, plus généralement, la connaissance des données est essentielle et vitale pour la gestion, l'accès et l'analyse critique des données et du processus de collecte des données. ii) davantage d'informations contextuelles devraient être fournies à l'avance sur l'objet de l'ensemble de données afin d'identifier la connaissance spécifique du domaine nécessaire, et iii) que le critère "licence", le critère "conditions d'utilisation" et le critère "prototypes et documentation" peuvent être encore améliorés.

Toutes ces informations ont été prises en compte lors de la création d'un plan de préservation et d'un flux de travail dans un scénario idéal à KBR (T3.3). Cela inclut des recommandations pour KBR, telles que la fourniture d'une meilleure documentation afin d'archiver et de préserver notre corpus de manière durable à long terme.

Objectif 3 : Pilote pour fournir l'accès à l'archive des médias sociaux : Dans le WP4, une analyse a été menée sur la base des besoins des utilisateurs (T4.3) lors de recherches sur une interface de médias sociaux (basée sur l'interface existante de CENTAL). Nous avons identifié que le contenu archivé devait pouvoir être interrogé, et que l'idée d'une interface de recherche classique ne suffirait pas pour la recherche sur les médias sociaux. Les critères suivants doivent être pris en compte :

- Orientation (par exemple, un nouvel utilisateur qui visite le site web et l'interface utilisateur sans aucune connaissance préalable a besoin d'informations contextuelles sur ce qu'il peut faire ici, sur les données disponibles via l'interface utilisateur, etc.)
- Vérification (par exemple, l'utilisateur doit être aidé à créer sa requête de recherche).
- Construire (par exemple, disposer de visualisations sous forme de diagrammes à barres et de diagrammes circulaires pour saisir rapidement les données de l'utilisateur).

Une maquette de plateforme d'accès a été créée (T4.2) par le partenaire technique sur la base de ces recommandations, des besoins de BESOCIAL (par ex. NLP) et des recommandations légales (T4.1).

Dans l'ensemble, nous pouvons conclure que les résultats du projet BESOCIAL constituent une première étape importante vers la mise en œuvre d'une stratégie d'archivage des médias sociaux à long terme pour la Belgique.

Mots-clés

Humanités numériques, Archivage de réseaux sociaux, Collections nativement numériques, Institutions patrimoniales, Préservation numérique.