CONTRAT DE LICENCE D'UTILISATEUR FINAL SODHA

1. Vue d'ensemble du contrat

1.1 *Introduction*

Ce document est un contrat de licence d'utilisateur final conclu entre les utilisateurs de la plateforme ou de l'infrastructure de l'**Archives de données en sciences sociales et humanités numériques** (ci-après désignés comme « les utilisateurs » ou « l'utilisateur ») et l'Archives de données en sciences sociales et humanités numériques (« SODHA », « l'archive de données ») qui est un service aux Archives de l'État en Belgique agissant au nom et pour le compte du déposant. Ce document stipule le niveau de service attendu tel qu'il doit être presté pour tout utilisateur de la plateforme SODHA.

En tant que service public et gratuit presté au sein d'une institution fédérale, SODHA souscrit aux objectifs de préservation à long terme qui sont la raison d'être de son organisation-mère aussi longtemps que le Service public de programmation Politique scientifique (BELSPO) financera le fonctionnement de l'archive de données.

1.2 *Définitions*

L'**infrastructure SODHA** ou la **plateforme SODHA** s'appuie sur DATAVERSE, un logiciel gratuit et *open source* de gestion de dépôt électronique d'archives et de publication de données de recherche en vue de réutilisations. Ce logiciel a été développé par l'Institut pour la science sociale quantitative (*Institute for Quantitative Social Science*, IQSS) de l'université Harvard. SODHA comprend deux installations Dataverse, une instance de production à l'adresse https://www.sodha.be et une instance de test à l'adresse https://test.sodha.be.

Un **utilisateur** ou **usager** de SODHA est une personne naturelle, inscrite sur la plateforme ou non, via le login institutionnel ou tout autre mode d'inscription, qui se connecte à l'un des sites Web susmentionnés et qui en fait usage ou qui prend contact avec SODHA d'une façon ou d'une autre.

Un **identifiant numérique d'objet** (*digital object identifier*, DOI) est un identifiant unique et permanent fourni par un organisme pour des objets numériques tels que des publications scientifiques au sens courant du terme (articles, livres, chapitres de livre…) ainsi qu'à des jeux de données / datasets.

Un **backup** est une copie d'objets numériques réalisée afin d'en permettre la récupération au cas où les fichiers originaux seraient perdus.

Une **archive de données** est un dépôt d'archives qui fournit des services de gestion électronique des documents (*records management*) et d'archivage à long terme pour des données de recherche. Contrairement à des centres d'archive au sens traditionnel du terme, les archives de données prennent soin d'objets numériques qui ont été produits récemment et que l'on qualifierait de documents « dynamiques » ou « semi-statiques ». Cela signifie que les producteurs de ces documents, ainsi que d'autres utilisateurs, sont susceptibles de réutiliser ces documents dans un avenir proche (par opposition à des documents qualifiés d'« archives définitives ») et qu'il est souvent trop tôt pour documenter sur le plan historique et traiter ces documents (par exemple, en convertissant les fichiers dans des formats ouverts).

2. Contrat de licence

2.1 *Types et niveaux de service*

Le but essentiel de SODHA est de recevoir des collections de données de recherche (des datasets) et de les mettre à disposition en vue de réutilisations. À cette fin, nos principaux services consistent à :

1) Fournir à nos utilisateurs une **interface de versement**

   SODHA s'engage à maintenir une plateforme en ligne sur laquelle des utilisateurs connectés peuvent documenter des datasets (via la rédaction de métadonnées) et téléverser des fichiers de données.

   L'interface de dépôt du logiciel open source sur lequel s'appuie SODHA, Dataverse, a été conçue à partir de la norme internationale de métadonnées pour les sciences sociales, *Data Documentation Initiative* (DDI). En d'autres termes, les champs de métadonnées que les déposants doivent ou peuvent remplir (dépendamment du fait qu'un champ est obligatoire ou non) proviennent directement d'une définition de type de document (*document type definition*, DTD) faisant autorité, et qui a été élaborée par des experts en sciences sociales qui ont déterminé quelles informations devaient être nécessairement enregistrées afin de bien documenter des données de recherche dans leurs disciplines.

2) Fournir à nos utilisateurs un **catalogue de données**

   Le logiciel Dataverse présente d'entrée de jeu, sur sa page d'accueil, une liste des datasets qui ont été publiés sur la plateforme.

   Par défaut, la liste recense les datasets publiés à partir du plus récent. Les utilisateurs peuvent parcourir cette liste, ou bien ils peuvent inverser l'ordre de présentation, ou encore ils peuvent choisir d'afficher les datasets par alphabétique ou antialphabétique (de A à Z ou inversément). Ils peuvent également avoir recours à la fonctionnalité de recherche avancée ainsi qu'aux facettes de recherche (trier les résultats par auteur, par mot-clé, par affiliation d'auteur…).

3) **Préserver** les données de recherche

   SODHA s'engage à fournir à ses utilisateurs une solution de stockage de données sûre et fiable. En tant que de service aux Archives de l'État en Belgique, SODHA bénéfice de l'expertise en préservation numérique de son institution-mère.

   La politique de SODHA en matière de backups consiste à effectuer une copie complète de toutes les données déposées sur la plateforme en début de semaine, puis d'effectuer des backups incrémentaux chaque jour ouvrable qui suit jusqu'à la fin de la semaine. Sous leur forme « active », directement accessible, les données déposées sur la plateforme sont stockées dans un premier local à serveurs tandis que les backups sont, quant à eux, stockées dans un second local. Cependant, les deux locaux en question sont situés sur le même site, sans la distance de 200 kilomètres qui est recommandée pour une procédure de backup de sécurité maximale. Les Archives de l'État travaillent à l'intégration du stockage hors-site à leur infrastructure.

   À l'avenir, certains des datasets publiés sur la plateforme SODHA reviendront de droit aux Archives de l'État, notamment les données provenant d'organisations fédérales, qui sont concernées par la loi sur les archives du 24 juin 1955. À ce moment-là, ces données seront

documentées plus avant et potentiellement intégrées à des collections plus importantes (en termes de volume) par un archiviste qui rédigera un inventaire dédié. Une fois cela fait, et après le transfert des données vers le système de gestion principal des Archives de l'État, ces données ainsi que leurs métadonnées seront copiées sur des bandes LTO et envoyées à la plateforme *Long-Term Preservation* (LTP), une infrastructure financée par BELSPO. La plateforme LTP est une solution d'archivage numérique de pointe qui propose toutes les garanties nécessaires pour l'archivage à long terme.

Cependant, les données qui n'appartiennent pas aux Archives de l'État — typiquement les données déposées par les chercheurs universitaires — ne peuvent être transférées vers la plateforme LTP car, comme dit dans la convention de dépôt SODHA, ces données restent la propriété de leurs ayants droit. La propriété des données peut être transférée aux Archives de l'État via une convention *ad hoc* ou une convention-cadre qui peut être signée entre les Archives de l'État et un centre de recherche au sein d'une université.

Par conséquent, quoique SODHA s'efforce de devenir un dépôt de confiance, nous déconseillons nos utilisateurs de conserver leurs données *uniquement* sur notre plateforme car nos moyens sont encore limités. SODHA adhère aux principes LOCKSS (*Lots Of Copies Keep Stuff Safe*, « Plusieurs copies sont des garanties de sécurité ») et incitent donc les utilisateurs à considérer SODHA comme une option de backup secondaire. Ces contre-indications sont d'autant plus pertinentes dans le cas de données à accès restreint.

4) **Conseiller** les utilisateurs

Dans la mesure de l'expertise au sein de notre archive de données et, plus largement, des Archives de l'État en Belgique, SODHA aura à cœur de répondre aux requêtes de ses utilisateurs concernant l'archivage des données, les pratiques de gestion de données, les formats de fichiers, les plans de gestion de données (*data management plans*, DMP) ou encore la documentation des données, entre autres sujets.

Toutefois, du fait du jeune âge de SODHA, nous ne pouvons pas encore fournir une assistance *ad hoc* pour des problèmes de gestion de données très particuliers, par exemple des problèmes relatifs à la collection de données, aux logiciels de recherche en sciences sociales, aux méthodologies de recherche, aux problèmes d'éthique, à l'anonymisation des données, etc. De façon générale, si nous ne nous sentons pas capables de fournir une réponse à la hauteur des attentes de nos usagers, nous tâcherons de les rediriger vers d'autres services ou organisations plus à même de leur fournir de l'aide.

Notre manuel en ligne, le *SODHA Guide*, contient une rubrique destinée aux Ressources externes, sous la forme d'une liste d'outils, de services et d'organisations susceptibles d'intéresser les chercheurs en sciences sociales et en humanités numériques. Si vous connaissez des outils, services ou organisations qui pourraient être ajoutés à cette liste, n'hésitez pas à nous contacter et nous les y ajouterons volontiers.

2.2 *Accessibilité des services et applications en ligne*

SODHA s'engage à fournir un accès en ligne ininterrompu à ses services et ressources Web dans la mesure des moyens — matériels, logiciels et humains — mis à sa disposition par son organisme de financement.

En cas de panne, nous nous efforcerons, autant que faire se peut, de redonner accès à nos services en ligne le plus vite possible. Cependant, du fait que SODHA fait partie des Archives de l'État, qui

sont une institution fédérale, la possibilité d'intervention technique est limitée aux heures de travail des jours ouvrables du lundi au vendredi.

2.3 *Identifiants d'objets numériques (DOI)*

Les Archives de l'État se sont abonnées à la plateforme DataCite pour l'attribution de DOI aux datasets publiés sur la plateforme SODHA. DataCite est un consortium international sans but lucratif dévoué à la promotion des résultats de recherches scientifiques et qui s'acquitte de cet objectif par la fourniture de DOI, qui ont pour effet d'augmenter considérablement la visibilité en ligne d'objets numériques, notamment aux yeux des robots indexeurs (*crawlers*) des moteurs de recherche tels que Google, Google Scholar, Bing, etc.

Les datasets publiés sur la plateforme SODHA reçoivent un DOI au moment de leur publication. Les fichiers contenus dans les datasets ainsi que les nouvelles versions des datasets ne reçoivent pas de DOI.

Seuls les datasets publiés reçoivent un DOI. Les datasets encore à l'état de « brouillon » (*Draft*) « non publié » (*Unpublished*) n'ont pas encore reçu de DOI, comme on peut le constater lorsque l'on essaie d'accéder à la page Web d'un dataset en cliquant sur son hyperlien DOI avant la publication du dataset.

2.4 *Communication et retours*

SODHA s'engage à réagir aux requêtes de ses utilisateurs en temps et en heure. SODHA s'efforce de fournir à ses utilisateurs un retour constructif, avec des informations pertinentes et/ou une assistance technique pour l'usage des services de notre archive de données de manière à la fois proactive et réactive.

Comme dit précédemment à propos des pannes, il est important de noter que la réactivité de SODHA aux communications de ses usagers est limitée aux heures de travail du lundi au vendredi.

Dans le cas particulier où des utilisateurs demande l'accès à des documents administratifs relatifs au fonctionnement de l'archive de données, SODHA, étant un service compris dans l'institution scientifique fédérale que sont les Archives de l'État, se conformera aux obligations de l'article 32 de la Constitution belge et de la loi fédérale sur la publicité de l'administration du 11 avril 1994, cette dernière imposant une réaction de la part de l'administration aux requêtes présentées par ses utilisateurs endéans les 30 jours après réception de la requête.

SODHA n'effectue aucune forme de discrimination entre ses usagers et traitera chaque demande et chaque utilisateur sur un pied d'égalité.

2.5 *Droits exclusifs*

Comme dit précédemment, le logiciel Dataverse a été développé par l'Institut pour la science sociale quantitative (*Institute for Quantitative Social Science*, IQSS) de l'université Harvard. Ce programme *open source* a été publié en vue de réutilisations avec une licence Apache 2 qui peut être consultée sur GitHub. Voir aussi, pour plus d'informations, [1] et [2].

3. Historique de version de ce document

Date	Numéro de version	Changements effectués
30-07-2021	1.0	Première version