



Yareta

Politique de préservation  
des données de recherche



UNIVERSITÉ  
DE GENÈVE

<b>Introduction</b> .....	<b>3</b>
<b>1 Conformité OAIS</b> .....	<b>4</b>
<b>2 Responsabilités administratives</b> .....	<b>5</b>
2.1 Objectif .....	5
2.2 Mandat .....	5
2.3 Objectifs .....	5
<b>3 Viabilité organisationnelle</b> .....	<b>6</b>
3.1 Champ d'application .....	6
3.2 Principes de fonctionnement .....	7
3.3 Rôles et responsabilités.....	7
3.4 Sélection et acquisition .....	8
3.5 Accès et utilisation .....	11
3.6 Défis et risques .....	12
<b>4 Viabilité financière</b> .....	<b>14</b>
4.1 Engagement institutionnel .....	14
4.2 Coopération et collaboration .....	14
<b>5 Adéquation technologique et procédurale</b> .....	<b>16</b>
<b>6 Sécurité du système</b> .....	<b>18</b>
<b>7 Responsabilité procédurale</b> .....	<b>20</b>
7.1 Audit et transparence .....	20
7.2 Administration du cadre politique .....	20
7.3 Définitions .....	20
7.4 Références.....	23

## INTRODUCTION

La politique de préservation de Yareta (la Politique) sert de cadre politique général pour Yareta, un service d'archivage générique pour la gestion, la préservation et le partage des données de recherche adapté à toute équipe de recherche et aux Hautes Écoles Supérieures (HES) du canton de Genève, en Suisse.

La politique se fonde sur les documents d'orientation suivants dans le domaine de la gestion des données de recherche et de la conservation numérique :

- Le modèle de référence de l'Open Archival Information System (OAIS) (ISO 14721) développé par le Consultative Committee for Space Data Systems (CCSDS) : <https://public.ccsds.org/pubs/650x0m2.pdf>
- Les principes FAIR définis par un consortium de scientifiques et d'organisations : <https://www.go-fair.org/fair-principles>
- Le modèle de cadre politique de préservation numérique préparé par le Consortium interuniversitaire pour la recherche politique et sociale (ICPSR) : <https://www.icpsr.umich.edu/files/ICPSR/curation/preservation/policies/dp-policy-outline.pdf>.
- Le modèle de politique de gestion des données de recherche préparé par le projet de gestion du cycle de vie des données (DLCM), mandaté par swissuniversities : <https://www.dlcm.ch/resources/dlcm-policy>

L'objectif de cette politique est de garantir que les données de recherche de Yareta sont activement gérées pour être reproductibles et réutilisées par d'autres, à l'avenir, conformément aux normes et aux meilleures pratiques en matière de gestion des données et de conservation numérique.

L'organisation de la politique reflète les sept attributs d'un dépôt numérique fiable :

1. Conformité OAIS
2. Responsabilité administrative
3. Viabilité organisationnelle
4. Viabilité financière
5. Adéquation technologique et procédurale
6. Sécurité des systèmes
7. Responsabilité procédurale

Version 1.2

Préparé par : Lydie Echernier, conseillère en recherche et gestionnaire de services pour Yareta, équipe e-Research, UNIGE  
David Gold, développeur et gestionnaire de services pour Yareta, équipe e-Research, UNIGE  
Mathieu Vonlanthen, ingénieur de développement pour Yareta, équipe e-Research, UNIGE

Validé par : Pierre-Yves Burgi, Chef du Service de la Recherche et de l'Information Scientifique (RISE), Division IT, UNIGE, et représentant de Yareta  
Hugues Cazeaux, Chef de l'équipe e-Research, UNIGE, et architecte de solution pour Yareta

Date de création : 24 octobre 2022

Date de la dernière révision : 31 octobre 2022

Date de publication : 31 octobre 2022

## 1 CONFORMITÉ OAIS

Yareta et la technologie sous-jacente de gestion du cycle de vie des données (DLCM) ont été conçus en conformité avec le modèle de référence OAIS (ISO 14721) et diverses normes ouvertes et internationales afin de faciliter leur intégration avec d'autres systèmes d'information et d'éviter le verrouillage technologique. Son architecture prend en charge l'acquisition, la validation, l'ingestion, le stockage, la gestion, la préservation et la diffusion de tous les types de données de recherche selon les meilleures pratiques actuelles de préservation.

Outre le modèle de référence OAIS utilisé pour assurer la conservation à long terme des données, Yareta a été conçu dans le respect des [principes FAIR](#), un ensemble de principes directeurs qui régissent les meilleures pratiques de gestion des données de recherche afin de rendre les données trouvables, accessibles, interopérables et réutilisables. Il met également en œuvre des technologies et des normes ouvertes telles que :

- Le format d'emballage de fichier hiérarchique [BagIt](#) de la Library of Congress, le Metadata Encoding & Transmission Standard ([METS](#)) et le Preservation Metadata Maintenance Activity ([PREMIS](#)) pour produire des paquets d'informations archivistiques de haute qualité et conformes aux normes.
- Le protocole [OAI-PMH](#) (Open Archive Initiative Protocol for Metadata Harvesting) pour faciliter l'échange de métadonnées.
- Le [schéma de métadonnées de DataCite](#) et l'identificateur d'objet numérique ([DOI](#)) pour assurer l'identification correcte et unique des ressources.
- Le File Information Tool Set ([FITS](#)) de Harvard, le [schéma PUID](#) des Archives nationales du Royaume-Uni et le Recommended Formats [Statement](#) de la Bibliothèque du Congrès, afin d'encourager l'utilisation de normes ouvertes tout en maintenant la solution ouverte à la plus large gamme de formats de fichiers utilisés dans les différentes disciplines scientifiques.
- Le programme [LOCKSS](#) (Lots of Copies Keep Stuff Safe) de l'université de Stanford pour assurer collectivement la préservation, la vérification de l'intégrité et la réparation des données sur un réseau distribué et décentralisé.
- L'Open Researcher and Contributor ID ([ORCID](#)), le Research Organization Registry ID ([ROR ID](#)) et le Software Package Data Exchange ([SPDX-ID](#)) pour améliorer la visibilité et l'attribution des données.

### 2.1 OBJECTIF

Yareta a pour objectif de rendre FAIR toutes les données de recherche dans le canton de Genève en utilisant et en combinant des technologies ouvertes, des normes et des principes conformes aux politiques européennes et aux développements internationaux, afin de créer des services de données durables, transparents et vérifiables qui facilitent la gestion, la préservation et le partage des données de recherche selon le modèle de référence OAIS (ISO 14721).

### 2.2 MANDAT

Yareta est un dépôt de données de recherche générique proposé à toutes les équipes de recherche et aux Hautes Écoles Supérieures (HES) situés dans le canton de Genève, dans le cadre de la mission de l'Université de Genève (UNIGE) de rendre les données de recherche aussi largement disponibles que possible, en particulier lorsque des fonds publics ont été reçus pour la réalisation du projet de recherche (§14 de la [Politique institutionnelle de l'UNIGE sur la gestion des données de recherche](#)).

Yareta est également le dépôt de données institutionnel de l'UNIGE. A ce titre, ce document est accompagné de :

- [Les directives de l'UNIGE sur l'intégrité dans la recherche scientifique](#)  
visent à garantir l'intégrité dans la recherche scientifique, à promouvoir une recherche de haute qualité, à assurer un cadre cohérent pour l'intégrité, à sensibiliser les équipes de recherche aux risques découlant des conflits d'intérêts et à définir une procédure pour les violations présumées de l'intégrité.
- [Politique institutionnelle de l'UNIGE sur la gestion des données de recherche](#)  
Affirme l'importance d'une bonne gestion des données de la recherche et rappelle aux parties prenantes leurs devoirs respectifs.
- [Charte de l'UNIGE pour la science ouverte](#)  
Marque l'engagement de l'institution et de la communauté universitaire en faveur du partage des connaissances scientifiques, et affirme l'adhésion de l'Université aux principes de la science ouverte, qui visent à garantir un accès libre et facile aux publications scientifiques, aux données de recherche et aux méthodologies utilisées pour générer ces données.
- [Directives sur l'application à l'Université de la loi sur l'information du public, l'accès aux documents et la protection des données personnelles \(LIPAD\)](#)  
Vise à informer les membres de la communauté universitaire de la loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD), et à promouvoir son application cohérente.

### 2.3 OBJECTIFS

Yareta soutient la gestion, la préservation et le partage de données de recherche de haute qualité dans tous les établissements d'enseignement supérieur de Genève et leur fournit les bases de la gestion des données - collecte, documentation, partage et publication - en :

- Développant et maintenant une infrastructure de haute qualité pour faciliter une **gestion, une préservation et une publication responsables des données de recherche**, conformément aux dispositions légales, aux exigences des éditeurs et des bailleurs de fonds et aux principes FAIR (Findable, Accessible, Interoperable and Reusable).
- Garantissant un accès sécurisé aux données tout en **promouvant les bonnes pratiques et les normes internationales**, pour une approche inclusive de la gestion, de la préservation et du partage des données.

- Collaborant avec les équipes de recherche pour élaborer des solutions pratiques pour la préservation et la publication de données volumineuses ou sensibles afin de **garantir la reproductibilité des arguments et des résultats scientifiques** à long terme, en respectant les différentes pratiques et processus de travail au sein des HES, des facultés, des départements et des groupes de recherche.
- **Mettant en valeur les données de recherche en tant que résultats de recherche indépendants** en rendant visibles les contributions aux projets de recherche des équipes de recherche de l'UNIGE et du canton de Genève.

## 3 VIABILITÉ ORGANISATIONNELLE

### 3.1 CHAMP D'APPLICATION

Yareta est un dépôt de données générique adapté à toutes les équipes de recherche et institutions du canton de Genève.

Cette politique s'applique à :

- Toutes les données de recherche sont collectées et conservées dans le but d'être conservées et partagées à long terme par Yareta.
- Métadonnées de préservation et identifiants persistants (DOI) associés aux archives et collections de données.
- Réseau de stockage distribué, répliqué sur deux nœuds institutionnels de l'UNIGE.

Ce que fait Yareta (selon le modèle de référence OAIS) :

- Portail : Fournit une interface utilisateur pour transférer les données et les métadonnées.
- Module de pré-ingestion : Effectuer l'identification et l'évaluation des fichiers, y compris l'analyse de virus, l'identification du format des fichiers, l'évaluation du niveau de conformité des données et le filtrage des fichiers.
- Module d'ingestion : Préparer les données pour l'archivage en utilisant les normes et les technologies des meilleures pratiques.
- Module de stockage d'archives : Gère le traitement et le transfert des archives pour leur conservation à long terme et assure le stockage sécurisé des données.
- Module de gestion des données : Rend les archives trouvables et enregistre les DOI dans les registres de données.
- Module d'accès : Rendre les archives accessibles à des fins de réutilisation, conformément aux droits d'accès définis pour les données par les équipes de recherche.
- Module de planification de la préservation : Gérer les périodes et les politiques de conservation des archives et des collections.
- Module d'administration : Surveiller l'environnement technique pour planifier et mettre en œuvre toute mesure nécessaire pour atténuer les risques.

Ce que Yareta ne fait pas :

- Fournir des espaces de travail collaboratif pour la gestion des données actives (stockage à court terme).
- Accepter les soumissions qui consistent uniquement en des données secondaires (publication, document de données, documentation).
- Effectuer la migration ou la normalisation du format des fichiers, la dérivation et la transformation des données.
- Assurer la conformité des données avec les questions d'éthique, de droit, de copyright et de droits de propriété intellectuelle.

Il est souligné que les chercheurs et les chercheuses restent responsables du contenu des archives et des collections qu'ils créent sur Yareta. Il peut être demandé aux chercheurs et aux chercheuses d'apporter les modifications et/ou les ajouts nécessaires à leurs données ou métadonnées pour se conformer à la présente politique. En cas de non-conformité,

l'UNIGE pourra prendre certaines mesures et en tirer des conséquences à la discrétion de l'UNIGE. Les Directives de l'UNIGE sur l'intégrité dans la recherche scientifique et les Directives sur l'application à l'Université de la Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD) indiquent les mesures restrictives (sanctions) en vigueur.

### 3.2 PRINCIPES DE FONCTIONNEMENT

Yareta s'engage à développer et à mettre en œuvre des procédures qui sont appropriées, proportionnées, fondées sur des preuves, pratiques, rentables et durables, et dans le meilleur intérêt du renforcement de sa mission. Pour ce faire, elle opère conformément aux principes suivants :

- Conformité aux normes et pratiques reconnues au niveau international, telles que les principes OAIS et FAIR.
- Alignement et intégration des technologies pour une meilleure conformité aux exigences des organismes statutaires, universitaires, de financement et d'édition.
- Développement et maintenance d'une architecture et de services fiables pour garantir l'accès aux archives et aux collections.
- Adoption de normes techniques et de métadonnées pour l'interopérabilité qui permettent l'échange de données et de métadonnées.
- Promotion de normes de qualité pour les données et les métadonnées afin d'accroître la valeur des données et de faciliter leur découverte, leur accès et leur réutilisation.
- Observation de la communauté de recherche, afin de suivre l'évolution des exigences en matière de fonctionnalités répondant à leurs besoins en matière de gestion, de préservation et de partage des données.

### 3.3 RÔLES ET RESPONSABILITÉS

Yareta est exploité au sein de l'UNIGE par le Service de l'information scientifique et de la recherche (RISe) de la Division informatique, dirigé par Pierre-Yves Burgi - agissant comme représentant de Yareta - et développé et maintenu par l'équipe e-Recherche, dirigée par Hugues Cazeaux - agissant comme architecte de solution pour Yareta. L'équipe e-Recherche comprend actuellement 6 développeurs et 2 responsables de service, et travaille en étroite collaboration avec la Bibliothèque de l'UNIGE, le Comité des données de recherche ouvertes de l'Université - qui comprend des représentants de différents services de l'Université énumérés ci-dessous - et les équipes de recherche des Facultés de l'UNIGE et des autres HES du canton de Genève.

Les parties suivantes jouent un rôle essentiel dans l'accomplissement de la mission de Yareta visant à protéger le patrimoine de données de la recherche dans le canton de Genève.

**La division informatique de l'UNIGE** est censée :

- Coordonner la création et le développement de services de haute qualité sur la préservation et le partage des données.
- Fournir une infrastructure de qualité pour faciliter la bonne gestion, la préservation et la publication des données, notamment :
  1. Services standard pour aider à la gestion des données importantes et sensibles.
  2. Infrastructure et services personnalisés si nécessaire.
- Assurer une gestion sécurisée de l'accès aux données conformément aux normes internationales et aux directives en matière de sécurité informatique.

La **bibliothèque de l'UNIGE** est censée :

- Coordonner un réseau de spécialistes des données intégrés dans les facultés.

- Travailler avec les facultés pour fournir un soutien et une formation en matière de gestion, de préservation et de publication des données.

**D'autres services de l'université**, notamment le :

Rectorat,  
 Service de soutien à la recherche,  
 Département des affaires juridiques,  
 Commission Universitaire pour une Recherche Ethique à l'Université de Genève (CUREG),  
 Centre de Compétences en Sciences des Données (CCSD),  
 Bureau de Transfert de Technologie (UNITEC),

On attend d'eux qu'ils :

- Fournir une contribution experte sur les questions politiques et pratiques liées à la protection et à la propriété des données.
- Veiller à ce que la compréhension des bonnes pratiques en matière de données de recherche soit reconnue comme faisant partie de la culture universitaire.
- Concevoir des stratégies pour traiter les aspects économiques de l'archivage à long terme des données.

On attend des **chercheurs principaux et des chercheuses principales** ou des personnes qui occupent des fonctions équivalentes dans les facultés de l'UNIGE et dans les autres Hautes Écoles du canton de Genève :

- Demander la création d'une "Unité organisationnelle", représentant leur équipe de recherche, leur projet, leur département, leur laboratoire, etc., qui correspond au contexte spécifique de leur environnement de recherche.
- Être le principal contact pour toutes les questions financières/facturation et budgétiser les coûts de la préservation des données de recherche au stade de la demande de l'unité organisationnelle.
- S'assurer que les exigences en matière de gestion des données de recherche imposées par les organismes statutaires, universitaires, de financement et de publication sont respectées, y compris le traitement approprié des données personnelles et confidentielles, comme l'exige la loi sur l'information du public, l'accès aux documents et la protection des données personnelles ([LIPAD](#)), le règlement général sur la protection des données ([RGPD](#)) et la loi fédérale sur la recherche impliquant des êtres humains ([LRH](#)), selon le cas - consulter la Commission Cantonale d'Éthique et de Recherche ([CCER](#)) ou la Commission Universitaire pour une Recherche Ethique à l'Université de Genève ([CUREG 2.0](#)) pour plus d'informations.
- Comprendre qui est propriétaire des données résultant de leurs recherches et les implications pour la gestion, la préservation et la publication des données, et veiller à ce que les accords contractuels avec des tiers sur la confidentialité des données soient toujours respectés.
- Suivre les demandes d'accès et les notifications, et assurer une gestion et un contrôle adéquats de leurs archives, y compris la qualité et la cohérence des données de recherche et des métadonnées associées.

**Les chercheurs et les chercheuses des Facultés de l'UNIGE et des autres HES du canton de Genève** sont censées :

- Suivre la stratégie et les pratiques de gestion des données de recherche de leur institution, de leurs facultés, de leurs départements et de leurs groupes de recherche, et suivre une formation à la gestion des données si nécessaire.
- Veiller à ce que les données de recherche soient documentées de manière appropriée, conformément aux principes FAIR (Findable, Accessible, Interoperable and Reusable).
- Veillez à ce que les métadonnées liées aux données soient pertinentes et utiles même lorsque les données ne sont pas publiques, avec une déclaration expliquant pourquoi l'accès est restreint ou fermé, qui peut utiliser les données et dans quelles circonstances.

### 3.4 SÉLECTION ET ACQUISITION



Yareta est un repository générique, conçu pour la préservation de tout type de format de fichier et ouvert aux équipes de recherche de toutes les hautes écoles du canton de Genève. Il accepte les données originales générées et enregistrées par les équipes de recherche ainsi que les données préexistantes collectées auprès d'autres sources. Les équipes de recherche doivent évaluer et sélectionner les données à conserver à long terme en fonction des politiques ou des exigences légales auxquelles elles sont soumises, et choisir parmi un ensemble de politiques de soumission, de préservation et de diffusion en conséquence. Yareta s'assure que toutes les données sont accompagnées de métadonnées administratives, descriptives, structurelles et techniques suffisantes en effectuant des évaluations et des contrôles automatisés des fichiers pour garantir la cohérence avec les politiques. Toutes les données qui échouent à l'évaluation et/ou aux contrôles de qualité automatisés sont renvoyées au chercheur pour un examen plus approfondi, aboutissant soit à une amélioration (par exemple, des métadonnées améliorées) et à une nouvelle soumission, soit à une élimination. Bien que tous les formats de fichiers soient acceptés, Yareta comprend un système d'étoiles qui informe les chercheurs et les chercheuses du niveau de conformité de chaque fichier de données avec les formats de fichiers recommandés pour la conservation à long terme, les sensibilisant ainsi aux risques liés à l'obsolescence. À la fin de la période de conservation des données, un processus d'élimination aide les aides à déterminer le sort de leur datasets, ce qui aboutit soit à une prolongation de la période de conservation, soit à l'élimination des données, tout en préservant leur identifiant (DOI) et les métadonnées associées comme une "pierre tombale".

Les listes suivantes détaillent l'ensemble minimal d'exigences, les fonctions personnalisables, les fonctions de conservation et l'assistance dont disposent les équipes de recherche au cours du processus de sélection et d'acquisition dans Yareta.

## EXIGENCES MINIMALES

Yareta s'efforce de faciliter le travail des chercheurs et des chercheuses avec des exigences minimales :

- Les chercheurs et les chercheuses demandent la création d'une unité organisationnelle pour eux-mêmes ou leur équipe, ou demandent à être affectés à une ou plusieurs unités organisationnelles existantes pour gérer les dépôts, les archives, les métadonnées, les contributeurs et contributrices, les politiques, la validation et l'accès.
- Les chercheurs et les chercheuses fournissent un ensemble minimal de métadonnées requises par le schéma ouvert de métadonnées descriptives DataCite pour attribuer un identifiant permanent aux archives.
- Un identifiant d'objet numérique (DOI) est automatiquement attribué à une archive de données et peut être réservé par les équipes de recherche avant le traitement et le stockage des archives pour une conservation à long terme.
- Les chercheurs et les chercheuses définissent un niveau d'accès à leurs archives, soit public, restreint ou fermé, avec une période d'embargo optionnelle.
- Pour les archives en accès public, les chercheurs et les chercheuses choisissent une licence, avec un choix par défaut de licences Creative Commons (par exemple CC0 ou CC-BY, selon la configuration).

## FONCTIONS PERSONNALISABLES

Les fonctions personnalisables permettent aux équipes de recherche de poursuivre leur travail comme d'habitude avec le flux de travail qui leur convient :

- Lors de la création de leur unité organisationnelle, les chercheurs et les chercheuses définissent les politiques qui les régissent :
  - La soumission des données, avec un flux de validation facultatif.
  - La conservation des archives, avec un délai de conservation garanti allant de 5 ans à l'infini, ou personnalisable en fonction des besoins des chercheurs et des chercheuses.
  - La licence par défaut, y compris la suite Creative Commons de CC0, CC-BY, CC-BY-SA, CC-BY-NA, et les licences individuelles personnalisables.
  - La diffusion d'archives personnalisées à la demande (selon le protocole IIIF, une norme de diffusion et de partage des objets numériques).
- Des métadonnées supplémentaires riches et facultatives permettant une meilleure facilité de recherche et de réutilisation sont disponibles, notamment :

- Métadonnées descriptives supplémentaires avec des champs complémentaires provenant du pool de métadonnées DataCite.
- La possibilité d'associer un schéma de métadonnées personnalisé en XML et JSON, avec une validation automatique du type de fichier (format) et du contenu (structure) par rapport au schéma.
- Tags de domaine de recherche basés sur des registres suisses et internationaux : domaines et disciplines de recherche du Fonds National Suisse de la recherche scientifique (FNS), sujets de Re3data, ou tout autre registre sur demande.
- Des mots-clés pour catégoriser et filtrer les données.
- Types d'archives (basés sur l'attribut DataCite resourceType) permettant la préservation de différents types d'objets numériques.
- Des éléments graphiques personnalisés, tels que des logos, des avatars et des vignettes, pour améliorer la convivialité du portail et la lisibilité du contenu.
- Des API RESTful bien documentées permettent d'intégrer Yareta à d'autres applications afin qu'il s'intègre directement dans les flux de travail de la recherche - voici quelques exemples d'intégration :
  - Instruments de laboratoire (systèmes d'information de laboratoire et cahiers de laboratoire électroniques).
  - Outils de visualisation, d'analyse ou toute autre application permettant d'extraire et de synchroniser automatiquement les données (par exemple, Jupyter Notebooks).
  - Les centres de données externes doivent conserver une copie des données pour se conformer à une politique locale.

## FONCTIONS DE CURATION

Les fonctions de curation aident les équipes de recherche à assurer et à contrôler la qualité.

Fonctions de curation **automatisée** :

- Un module de pré-ingestion capable d'extraire, de répertorier et de tester les données compressées d'un fichier ZIP, tout en conservant la hiérarchie des dossiers et la structure des données.
- Une évaluation automatisée du dossier par le système, y compris :
  - Identification du format du fichier de données (FITS).
  - Examen des formats et des noms de fichiers pour trier automatiquement les fichiers qui doivent être exclus d'une archive (par exemple, les fichiers système) ou qui nécessitent une validation manuelle (par exemple, les noms de fichiers contenant "old").
  - Évaluation du niveau de conformité des données à l'aide du système PUID des Archives nationales du Royaume-Uni.
  - Recommandation de la Library of Congress Format Statement, pour encourager l'utilisation de normes ouvertes tout en gardant la solution ouverte à tous les types de fichiers.
  - Contrôle impliquant une série de fonctionnalités différentes (par exemple, cohérence des niveaux d'accès et des niveaux de sensibilité).
- Filtrage fin pour permettre aux chercheurs et chercheuses de valider, d'ignorer ou d'exclure des fichiers en fonction d'un statut spécifique.

Curation **humaine** :

- Les chercheurs et chercheuses peuvent choisir d'inclure un flux de validation facultatif avec un contrôle renforcé effectué par un responsable des données (expert du domaine, archiviste, bibliothécaire, etc.) permettant un examen manuel avec la possibilité d'approuver, de rejeter ou de renvoyer les ensembles de données à l'édition.
- Les chercheurs et chercheuses bénéficient d'un modèle DataTag intégré au système pour le partage et l'utilisation de données sensibles d'une manière normalisée et responsable.
- Les chercheurs et chercheuses peuvent modifier les métadonnées d'une archive et établir des liens avec les résultats de recherche connexes :
  - Indiquer qu'une archive est référencée par une autre archive ou par une publication (DOI référencé par)
  - Indiquer qu'une archive est identique à une autre (DOI identique à)

- Ajout du DOI d'une autre archive qui remplace l'archive actuelle (DOI rendu obsolète par)
- Les chercheurs et chercheuses peuvent créer des collections reliant des archives avec un référencement qualifié dans les métadonnées DataCite.
- Les chercheurs et chercheuses bénéficient d'un soutien spécifique pour les aider à planifier la préservation de leurs données, en tenant compte du coût de l'archivage à long terme des données et des exigences des programmes d'encouragement de la recherche suisses (FNS) et européens (Horizon).

## SOUTIEN

Une assistance est disponible pour aider les chercheurs et les chercheuses à se familiariser avec le portail Yareta et ses nombreuses fonctionnalités.

- Une documentation étendue avec un guide de "démarrage", un "guide de l'utilisateur" complet, des FAQ, une série de clips vidéo, des info-bulles et des visites guidées - proposées sous forme de tutoriels - est disponible en libre-service.
- Des ressources institutionnelles sont allouées au soutien à la gestion des données dans chaque établissement d'enseignement supérieur du canton de Genève, avec des possibilités de formation sur place et en ligne.
- Une assistance personnalisée est disponible pour la soumission et la gestion de datasets volumineux et/ou sensibles.

Yareta offre également des fonctionnalités supplémentaires, comme le support multilingue et la visualisation de documents, d'images, d'audio, de vidéo, etc., y compris de molécules chimiques en 3D.

## 3.5 ACCÈS ET UTILISATION

Yareta s'efforce de faire en sorte que les archives de données restent accessibles aux équipes de recherche. Il s'agit notamment de présenter des métadonnées descriptives par le biais d'une interface de recherche personnalisée et d'API publiques, et de suivre l'évolution des technologies, des normes et des exigences de la communauté de recherche pour continuer à répondre à ses besoins. Les utilisateurs et les utilisatrices de Yareta comprennent, sans s'y limiter, des équipes de recherche de tous les établissements d'enseignement supérieur du canton de Genève et de leurs partenaires en Suisse et à l'étranger, des facultés et des étudiants de l'UNIGE, des décideurs politiques, des journalistes et toute personne souhaitant accéder à des archives de données de recherche. Les identifiants ORCID, ROR ID et SPDX-ID permettent de désambigüiser les noms des chercheurs et des chercheuses, leur institution et la licence. Ceci donne plus de visibilité à leurs données en connectant leurs publications à travers différentes plateformes.

Yareta maintient un accès public, permanent et fiable aux métadonnées descriptives de l'archive, indépendamment du contenu de l'archive, accompagné de mesures permettant de suivre l'utilisation et l'impact des données, ainsi qu'une fonction de citation des données pour garantir une attribution correcte, améliorer la reproductibilité, la découvrabilité des données et contribuer à leur reconnaissance en tant que résultat de la recherche. Yareta consacre également des ressources importantes au développement et à la mise en œuvre de fonctions garantissant la confidentialité et la protection des données personnelles ou sensibles. L'accès au contenu des archives est régi par un contrôle d'accès basé sur les rôles et le niveau d'accès aux archives, les archives publiques pouvant être téléchargées sans mur de connexion, tandis que les archives restreintes ou fermées restent disponibles sur demande après connexion, selon le cas.

Yareta utilise les technologies et les normes actuelles pour fournir une gamme de services d'accès :

- Yareta offre une facilité de découverte grâce à l'indexation et à la récolte des métadonnées descriptives DataCite par le protocole OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting).
- Des API RESTful ouvertes et bien documentées permettent l'intégration aux infrastructures informatiques existantes.
- Les techniques de recherche traditionnelles sont complétées par un système de navigation à facettes personnalisable.

- Un affichage métrique complète chaque archive avec son nombre de fichiers, son nombre de vues, son nombre de téléchargements et la possibilité de noter la qualité et l'utilité de chaque archive.
- L'authentification et le contrôle d'accès basé sur les rôles permettent aux chercheurs et aux chercheuses d'accorder les autorisations appropriées aux membres de leur équipe en fonction de leurs flux de travail, et au-delà de leur propre équipe pour faciliter le partage et la collaboration avec leurs partenaires de la communauté internationale.
- Des listes de contrôle d'accès (ACL) attribuent des autorisations à des personnes individuelles ou à des groupes, et un système de notification permet aux chercheurs de traiter les demandes d'accès au cas par cas (demander, accorder ou refuser l'accès).
- Les citations de données peuvent être générées automatiquement en suivant les formats standards courants (APA 6e édition, MLA 7e édition et Chicago 16e édition).

Yareta permet aux utilisateurs de télécharger les données associées à une archive par différentes méthodes :

- Téléchargement direct de toutes les données associées à une archive publique : les données sont disponibles immédiatement pour le téléchargement et les fichiers sont accessibles à partir d'un lien affiché dans le navigateur.
- Commande groupée : les données de plusieurs archives sont ajoutées à un "panier" et sont disponibles pour le téléchargement sous forme de commande après que Yareta a fini de traiter la demande.
- Demande d'accès aux données d'une archive sous un niveau d'accès restreint ou fermé : les données sont disponibles par téléchargement direct ou par commande groupée si et quand l'accès est accordé par le propriétaire de l'archive.

### 3.6 DÉFIS ET RISQUES

Yareta doit relever les défis suivants en matière de préservation numérique :

#### ÉVOLUTION TECHNOLOGIQUE

La multiplication des formats de fichiers et le choix institutionnel de ne pas effectuer de migration ou de normalisation des formats de fichiers implique que les fichiers de données existants conservés dans Yareta sont confrontés à un risque d'obsolescence. Yareta doit donc s'adapter à une technologie en constante évolution pour garantir la continuité de l'accès à ces fichiers. Yareta doit également maintenir des systèmes et des flux de travail évolutifs (c'est-à-dire le personnel, la technologie, l'infrastructure, les pratiques) pour faire face à la croissance rapide et évolutive du volume de données.

#### SENSIBILISATION ET FORMATION

La majorité du personnel de l'UNIGE et des HES à Genève n'a pas la préservation numérique et la gestion des données comme partie explicite de leur mandat. Yareta s'engage à sensibiliser le personnel de l'UNIGE aux questions de préservation numérique et de gestion des données et à fournir une formation appropriée au personnel de l'UNIGE et à la communauté plus large des chercheurs et des chercheuses, des gestionnaires de l'information, des archivistes et des utilisateurs et utilisatrices de données. Yareta s'efforce également d'accroître la collaboration entre les divisions de l'UNIGE et les départements informatiques afin d'utiliser plus efficacement l'expertise et les ressources disponibles du personnel et de promouvoir une administration cohérente du système.

#### GESTION ET RESSOURCES

L'UNIGE élargit les types de contenus numériques collectés pour inclure des collections et des biens provenant de galeries, bibliothèques, archives et musées, ce qui signifie que les technologies, politiques, procédures et pratiques existantes devront être révisées. Yareta nécessite également un soutien financier durable et continu, notamment l'allocation de ressources technologiques, administratives et en personnel pour répondre à tous les besoins de développement des utilisateurs actuels et futurs.



### 4.1 ENGAGEMENT INSTITUTIONNEL

Yareta est accueilli à l'Université de Genève (UNIGE), fondée en 1559. L'UNIGE est un établissement public doté d'un statut moral, placé sous la tutelle du Conseil d'Etat genevois, qui exerce cette tutelle par l'intermédiaire du Département de l'instruction publique. L'UNIGE s'organise, fixe ses priorités et ses méthodes d'action et est responsable de sa gestion dans le cadre de la [Loi sur l'université du 13 juin 2008](#) (entrée en vigueur le 17 mars 2009), qui règle également son financement.

La préservation des données dans Yareta est une responsabilité conjointe entre Yareta et les équipes de recherche et les institutions déposantes. Si l'UNIGE met fin à Yareta pour quelque raison que ce soit, l'UNIGE restituera les archives et les métadonnées de préservation aux personnes et aux institutions déposantes selon des conditions mutuellement convenues. Au moment du transfert, l'UNIGE s'assurera que le support de transfert et le format de diffusion seront pertinents et compatibles avec les meilleures pratiques et normes actuelles.

### 4.2 COOPÉRATION ET COLLABORATION

La solution a été mise au point dans le cadre de deux projets plus vastes visant à développer divers services de gestion des données pour les équipes de recherche et les HES en Suisse :

#### PL 12146

Le projet de loi cantonale genevoise "Infrastructure numérique pour la recherche" ([PL 12146](#)) 2018-2024 :

L'État de Genève, l'UNIGE et la Haute École Spécialisée de Suisse occidentale (HES-SO) à Genève se sont associés pour créer une convention-cadre sur le numérique. Cette convention vise à faciliter et encourager la collaboration dans les domaines de l'enseignement, de la formation continue, de la recherche appliquée et des services liés à la transformation numérique. Cette collaboration permet la mise en place d'activités communes telles que l'offre de stages, l'organisation d'événements numériques, des collaborations avec les étudiants sur ces sujets, ainsi que la collaboration à des projets de recherche cantonaux, nationaux et internationaux. L'État de Genève a accordé à l'UNIGE une subvention de plus de 15 millions de francs, allant de 2018 à 2022 et prolongée jusqu'en 2024, pour financer le développement d'infrastructures de recherche et la mise en place de services numériques associés, notamment dans les domaines du calcul haute performance et de la gestion des données de recherche.

Ce projet s'adresse à tous les établissements d'enseignement supérieur de Genève et vise à :

1. Renforcer et développer l'infrastructure numérique existante.
2. Développer de nouvelles solutions et de nouveaux services numériques pour la recherche.
3. Utiliser les technologies de pointe conformément aux politiques de recherche nationales et internationales.
4. Assurer la qualité de l'information, de la documentation et du support associés aux services.
5. Écouter, soutenir et conseiller les chercheurs et les chercheuses sur les questions numériques.

Pour renforcer l'équipe, plusieurs personnes ont été engagées pour la durée du projet. Elles sont au nombre de 6 dans l'équipe e-Research chargée du développement de Yareta et de la gestion du service à l'UNIGE. Les 3 autres membres de l'équipe sont sous contrat permanent afin d'assurer la continuité des services.

#### DLCM

Le projet national de gestion du cycle de vie des données ([dlcm.ch](#)) 2015-2021 :

La technologie DLCM qu'utilise Yareta est le résultat d'un effort coordonné de l'UNIGE avec neuf partenaires - huit HES suisses, à savoir l'[EPFL](#), l'[ETHZ](#), la [HEG](#), la [HES-SO](#), l'[UNIBAS](#), l'[UNIL](#) et l'[UNIZH](#) depuis 2015, et la [ZHAW](#) depuis 2018, et un prestataire de services, [SWITCH](#) - visant à proposer des solutions au niveau national couvrant l'ensemble du DLCM (Data Life-Cycle Management) adaptées aux équipes de recherche et aux institutions, afin de les aider à gérer, publier et préserver leurs données en toute sécurité sur le très long terme (20 ans et plus) tout en traitant la plupart des outils et des formats dans toutes les disciplines scientifiques.

En 2014, comme fondation du projet DLCM, 49 entretiens ont été menés dans 30 disciplines différentes de 6 HES pour identifier les besoins primaires des équipes de recherche et les solutions existantes en place. Au cours de la période DLCM 2015-2018, le projet a identifié les composants de base qui ont contribué au développement d'un référentiel de données générique adapté à tous les chercheurs et institutions à travers la Suisse. Avec un partenaire supplémentaire impliqué, le projet a ensuite mené un certain nombre de projets pilotes centrés sur l'utilisateur pendant la phase d'application en 2019 et 2020 pour affiner l'ensemble des services de base avec des fonctionnalités supplémentaires. En 2021, le projet a abouti à la mise à disposition de dépôts de données conformes aux normes FAIR et OAIS intégrés aux solutions des partenaires, ainsi qu'à des services de formation et de conseil fournis par un réseau national d'experts.

Yareta, l'instance genevoise de la solution DLCM, a été lancée le 26 juin 2019. Son nom provient d'une plante de la famille des Apiacées originaire d'Amérique du Sud, connue pour être l'un des plus anciens organismes vivants sur terre - plus de 3'000 ans selon certaines estimations. Yareta est enregistré auprès de [Re3data](#), [BASE](#) (Bielefeld Academic Search Engine), [OpenAIRE](#) et [OpenDOAR](#) (Directory of Open Access Repositories), et peut bénéficier d'un financement pour le téléchargement de données aux chercheurs dans le cadre d'une [subvention du FNS](#).

[OLOS.swiss](#), l'instance nationale de la solution DLCM, offre des services similaires à toutes les équipes de recherche suisses et aux établissements d'enseignement supérieur en dehors du canton de Genève.

## RÉSEAU DE SOUTIEN RDM

Le réseau national de soutien à la gestion des données de recherche ([RDM support](#)) :

Les membres de l'équipe Yareta participent activement à un réseau de soutien national né du projet DLCM et d'autres projets locaux ou nationaux financés par des agences nationales de financement de la recherche et le Secrétariat d'État à l'éducation, à la recherche et à l'innovation (SERI). Il rassemble des intervenants des équipes de soutien à la gestion des données de recherche au sein des universités, des universités de sciences appliquées, des universités de formation des enseignants et des instituts de recherche. Le réseau se réunit deux fois par an sur des questions liées à :

1. Soutien à la planification de la gestion des données (DMP).
2. Formation à la gestion des données de recherche.
3. Consultations sur les questions juridiques et éthiques liées à la gestion des données de recherche.
4. Infrastructures, services et outils de gestion des données de recherche.

## 5 ADÉQUATION TECHNOLOGIQUE ET PROCÉDURALE

En examinant les actions de préservation technologiques et procédurales appropriées, Yareta a l'intention de conserver les données dans leur format de fichier original au sein des archives, et de permettre leur diffusion.

Les actions de préservation sont guidées par les principes suivants :

- Maintenir les technologies existantes et évaluer les technologies en évolution pour les développements futurs.
- Suivre et s'adapter aux nouvelles technologies en fonction de l'évolution des meilleures pratiques en matière de préservation numérique.
- Normaliser la technologie et les flux de travail entre les applications ayant des exigences de préservation similaires.
- Suivre les meilleures pratiques actuelles en matière de gestion des données et conseiller les équipes de recherche et les HES en conséquence.

Yareta s'engage à préserver les archives publiées conformément à la présente politique, aux procédures décrites sur le [site web de l'UNIGE e-Research](#) et aux [spécifications Yareta](#) qui énumèrent les normes, protocoles, registres, technologies et applications open-source utilisés par les fonctionnalités Yareta. Yareta permet d'accéder aux données de recherche grâce à la maintenance des archives, des métadonnées descriptives et des identifiants associés, et fournit des citations et des mesures d'accès, qui comprennent les pages vues, les téléchargements et les évaluations. Yareta s'engage à assurer ce niveau de conservation, d'accès et de rapport pendant une période d'au moins 5 ans pour toutes les archives, et d'au moins 20 ans pour les archives destinées à être conservées "pour toujours", à compter de leur date de publication sur Yareta. A la fin de la période de conservation, Yareta informe les chercheurs pour leur donner le choix entre :

- Extension de la période de préservation : Yareta conservera les données associées à ses archives publiées au-delà de l'engagement initial pendant une période d'au moins 5 ans.
- Destruction : Yareta détruira en toute sécurité les données associées à ses archives publiées, tout en conservant les métadonnées descriptives et l'identifiant correspondants.

En l'absence de réponse, Yareta s'efforcera de conserver les archives publiées au-delà de l'engagement initial, mais se réserve le droit d'évaluer la viabilité à long terme de l'archive pour déterminer s'il convient de conserver l'archive au-delà de l'engagement initial. Yareta peut également retirer des données des fichiers d'archives en utilisant le processus d'élimination pour des raisons de maintenance de routine (par exemple, si l'on découvre qu'une archive est un doublon ou qu'une version haute résolution remplace un fichier de qualité inférieure) ou à la suite d'une réévaluation de la valeur d'une archive à la lumière des critères de sélection définis par la recherche, tout en conservant l'enregistrement de métadonnées descriptives et l'identifiant correspondants et en maintenant une référence à la nouvelle archive qui remplace l'ancienne (DOI rendu obsolète par).

Les actions de préservation, d'accès et de communication de Yareta sont soutenues par des fonctions de sécurité dédiées, qui permettent également la visibilité, la réutilisation et la reproductibilité des résultats de la recherche :

- Des normes ouvertes de préservation des données pour un stockage et un transfert fiable (PREMIS, METS et BagIt) afin d'éviter le verrouillage et de soutenir les stratégies de sortie.
- Mesures de sécurité telles que les sommes de contrôle, la redondance et la séparation géographique des copies (par défaut, les données restent à Genève, mais les nœuds d'archivage peuvent être configurés chez tout autre fournisseur de stockage sur demande).
- Les pistes d'audit capturent et enregistrent toutes les activités des utilisateurs et du système concernant les fichiers et les transactions. Les activités enregistrées comprennent les actions sur les unités organisationnelles (journal des modifications avec noms et horodatages), les archives (création, contrôles automatisés, validation, approbation, soumission, validation technique, édition de métadonnées, achèvement, nettoyage) et les fichiers individuels (création, traitement, recherche de virus, identification du format de fichier).
- Connecteurs à plusieurs systèmes de stockage d'archives tels que le File System, le protocole S3 et les bandes.
- Conteneurisation (Docker, Kubernetes) pour améliorer la sécurité, réduire les coûts et assurer la portabilité des nuages.



- Mesures techniques soutenant une " Privacy by Default and by Design " conforme aux législations suisse et européenne, telles que la loi sur l'information du public, l'accès aux documents et la protection des données personnelles ([LIPAD](#)) et le règlement général sur la protection des données ([RGPD](#)).

## 6 SÉCURITÉ DU SYSTÈME

L'intégrité des données et la continuité de l'accès aux données archivées à Yareta sont garanties par les éléments suivants :

1. Infrastructure technique
2. Gestion des données
3. Politiques de sauvegarde
4. Installations de stockage et sécurité
5. Prévention des catastrophes et reprise après sinistre

### INFRASTRUCTURE TECHNIQUE

Yareta est alimenté par la technologie DLCM, qui est un [logiciel libre](#) et fonctionne sur des technologies libres. La technologie DLCM est publiée sous une licence [GNU General Public License v2.0 ou ultérieure](#). Les actions de développement et de maintenance réalisées par l'équipe e-Research de l'UNIGE utilisent la méthodologie agile Scrum.

La technologie DLCM a été développée avec le framework Solidify, développé par l'UNIGE, étendu à partir de [Spring Boot](#).

- Système d'archivage : Solution DLCM (Java, Spring Boot, MariaDB, Elasticsearch, ActiveMQ, ClamAV, FITS)
- Frontend : Portail Yareta (Angular)
- Système d'exploitation : Debian Linux

Yareta gère trois instances indépendantes :

- Le système de production.
- Un système d'assurance qualité du développement, garantissant que les changements peuvent être testés et validés avant d'être appliqués au système d'assurance qualité de la formation.
- Un système d'assurance qualité de la formation, permettant aux équipes de recherche et aux membres du personnel responsables de la formation de tester et de valider les nouveaux développements avant leur déploiement dans le système de production.

### GESTION DES DONNÉES

La gestion des données dans Yareta repose sur la réplication asynchrone, qui consiste à partager les données entre deux ressources redondantes afin de garantir la cohérence, d'améliorer la fiabilité et la tolérance aux pannes. Chaque fichier de données est associé à :

- Un contrôle de fixité documenté.
- Un contrôle de virus au moment de l'ingestion initiale dans Yareta.
- Une procédure documentée de vérification des objets.

Pour chaque fichier de données, nous stockons des sommes de contrôle MD5, SHA1 et SHA256 indépendantes.

- La réplication asynchrone est effectuée quotidiennement.
- Le contrôle de fixité est effectué chaque semaine.
- Un contrôle de réplication est effectué tous les mois.
- La vérification des objets est effectuée chaque année.

Les métadonnées sont indexées dans un cluster Elasticsearch pour une recherche rapide et puissante. Toutes les modifications apportées aux métadonnées d'archive sur Yareta sont classées par version.

### POLITIQUES DE SAUVEGARDE

En outre, chaque copie de fichier possède deux répliques de sauvegarde situées sur des serveurs de disques différents. Les sauvegardes sont créées et maintenues localement et hors site selon les directives suivantes :

1. Les données et métadonnées nouvelles et modifiées dans Yareta sont immédiatement enregistrées de manière redondante sur le disque.
2. Des sauvegardes quotidiennes sont conservées pendant deux mois et stockées localement.
3. Des sauvegardes hebdomadaires sont conservées pendant deux ans et stockées localement et hors site.
4. Les sauvegardes annuelles sur des supports autres que des disques (bandes) sont conservées pendant 10 ans et stockées hors site.

## INSTALLATIONS DE STOCKAGE ET SÉCURITÉ

Physiquement, l'infrastructure technique de Yareta est située dans les locaux de l'UNIGE :

- Localement à Uni Dufour, Rue Général-Dufour 24, 1204 Genève, Suisse.
- Hors site à Campus Biotech, Chemin des Mines 9, 1202 Genève, Suisse.

Les supports de stockage comprennent les disques durs, NVME, bandes, gérés par Hitachi HCP et IBM Spectrum Protect.

Yareta suit les politiques établies par la Division informatique de l'UNIGE pour maintenir la sécurité de l'environnement technique en conformité avec les lois et règlements applicables en matière de sécurité des données. Ces politiques sont définies dans la Politique de sécurité du système d'information ([PSSI](#)) de l'UNIGE. L'UNIGE emploie des collaborateurs exclusivement dédiés à la sécurité des réseaux et des systèmes. Seuls les collaborateurs autorisés et impliqués dans la maintenance informatique ont accès au système d'archivage. Les datacenters sont climatisés et protégés par un système anti-incendie. Les privilèges d'accès aux datacenters sont limités par un système d'accès par badge du personnel.

Yareta utilise également les procédures de sécurité suivantes :

- Authentification : [SWITCHaai](#) et Shibboleth
- Autorisation : OAuth2 avec Spring Authorization Server
- Antivirus : ClamAV
- VPN : FortiGate
- Scanner de sécurité : Nessus
- Analyse du code : SonarQube

## PRÉVENTION DES CATASTROPHES ET REPRISE APRÈS SINISTRE

Tous les systèmes et services fournis par l'UNIGE, y compris Yareta, font l'objet d'une analyse des risques et des vulnérabilités sur une base annuelle. Des documents de politique dédiés traitent des aspects spécifiques de la protection à long terme de chacun des actifs numériques de l'UNIGE.

Yareta, en tant qu'infrastructure de conservation numérique de l'UNIGE, est considéré comme « mission-critical » et dispose de niveaux appropriés de surveillance, de réponse et de récupération. Un plan de reprise après sinistre dédié à Yareta couvre les procédures techniques et administratives visant à garantir une sécurité des données cohérente et systématique, y compris les exigences de sécurité du système, les exigences opérationnelles et les audits et examens réguliers.

### 7.1 AUDIT ET TRANSPARENCE

Les actions de préservation sont revues périodiquement, par cycles de 12 mois. La durée du cycle de révision dépend de l'évolution de l'environnement technologique, de l'évaluation des risques, des ressources financières et humaines disponibles et des besoins des équipes de recherche et des HES.

Yareta utilise deux modèles de maturité pour la préservation numérique :

Les niveaux de préservation numérique ([LoP](#)) de la National Digital Stewardship Alliance (NDSA), qui se concentrent sur 5 aspects de la préservation :

1. Stockage et localisation géographique
2. Fixité des fichiers et intégrité des données
3. Sécurité de l'information
4. Métadonnées
5. Formats de fichiers

Le modèle d'évaluation rapide ([RAM](#)) de la Digital Preservation Coalition (DPC), un outil conçu pour permettre une évaluation comparative rapide des capacités de préservation numérique, qui se concentre sur 6 capacités organisationnelles et 5 capacités de service :

1. Viabilité organisationnelle
2. Politique et stratégie
3. Base juridique
4. Capacité informatique
5. Amélioration continue
6. Communauté
7. Acquisition, transfert et ingestion
8. Préservation du flux binaire
9. Préservation du contenu
10. Gestion des métadonnées
11. Découverte et accès

### 7.2 ADMINISTRATION DU CADRE POLITIQUE

La politique de préservation des données de recherche de Yareta a été achevée en octobre 2022, approuvée par le chef du service de la Recherche et de l'Information Scientifique (RISe) de la Division informatique de l'UNIGE et représentant pour Yareta, et par le chef de l'équipe e-Research et architecte de solutions pour Yareta, le 31 octobre 2022. Les modifications préconisées seront portées à la connaissance des collaborateurs et des collaboratrices responsables de Yareta au sein de la Division informatique de l'UNIGE pour validation et au Rectorat de l'UNIGE pour information. Chaque nouvelle version de la politique portera un numéro de version et un cachet de date.

### 7.3 DÉFINITIONS

#### UNITÉ ORGANISATIONNELLE

Une unité organisationnelle est une entité logique qui peut représenter une unité structurelle, comme votre laboratoire, votre institut ou votre département, ou encore votre groupe de recherche ou votre projet. Elle permet de gérer les membres de l'équipe et leurs autorisations de dépôt et d'archivage des données de recherche.

## DÉPÔT

Un dépôt est l'endroit où les utilisateurs et les utilisatrices téléversent leurs ensembles de données et y associent des métadonnées. Il est entièrement modifiable jusqu'à ce qu'il soit soumis

## ARCHIVES

Une fois soumis, un dépôt devient une archive. L'archive est publiée sur Yareta et est liée par un DOI. Seules ses métadonnées peuvent être modifiées. Ses fichiers de données sont immuables.

## COLLECTION

Yareta permet aux utilisateurs et aux utilisatrices de combiner plusieurs archives dans une collection qui a ses propres DOI.

## ARCHIVE PIERRE TOMBALE

La pierre tombale est ce qui reste d'une archive après qu'elle ait été éliminée. Les fichiers de données associés aux archives publiées sont détruits de manière sécurisée. Les métadonnées descriptives et le DOI de l'archive sont conservés.

## RÔLES DES UNITÉS ORGANISATIONNELLES

Les utilisateurs peuvent avoir différents rôles au sein d'une organisation jusqu'à ce que. Ces rôles sont les suivants :

- Manager
- Steward
- Approver
- Creator
- Visitor

Vous trouverez des informations sur ces rôles dans le [Guide d'intégration Yareta](#).

## DEMANDE DE TÉLÉCHARGEMENT

Yareta permet aux utilisateurs et aux utilisatrices d'ajouter plusieurs archives à une "demande de téléchargement" qui téléchargera toutes les archives sélectionnées en une seule fois lorsqu'elle sera soumise.

## DONNÉES PRIMAIRES

Données collectées directement, par opposition aux données collectées à partir de recherches effectuées précédemment. Les données primaires sur Yareta sont classées comme suit :

- Observation
- Expérimental
- Référence
- Simulation
- Dérivé
- Numérisé

## DONNÉES SECONDAIRES

Données collectées à partir de recherches effectuées précédemment. Les données secondaires sur Yareta sont classées comme suit :

- Publication
- Papier de données
- Documentation

## 7.4 RÉFÉRENCES

Les liens suivants sont affichés par ordre d'apparition.

- Modèle de référence OAIS (ISO 14721) : <https://public.ccsds.org/pubs/650x0m2.pdf>
- Principes du FAIR : <https://www.go-fair.org/fair-principles>
- Cadre de la politique de conservation numérique du Consortium interuniversitaire pour la recherche politique et sociale (ICPSR) : <https://www.icpsr.umich.edu/files/ICPSR/curation/preservation/policies/dp-policy-outline.pdf>
- Modèle de politique de gestion des données de recherche préparé par le projet DLCM : <https://www.dlcm.ch/resources/dlcm-policy>
- Le format d'emballage des fichiers hiérarchiques BagIt de la Bibliothèque du Congrès : <https://www.rfc-editor.org/rfc/rfc8493>
- Norme de transmission et d'encodage des métadonnées (METS) : <http://www.loc.gov/standards/mets/>
- Protocole de l'Open Archive Initiative pour la collecte de métadonnées (OAI-PMH) : <https://www.openarchives.org/pmh/>
- Schéma de métadonnées DataCite : <https://schema.datacite.org/>
- Identificateur d'objet numérique (DOI) : <https://www.iso.org/standard/43506.html>
- File Information Tool Set (FITS) : <https://projects.iq.harvard.edu/fits/home>
- Schéma PUID des archives nationales britanniques : <https://www.nationalarchives.gov.uk/aboutapps/pronom/puid.htm>
- Formats recommandés Déclaration de la Bibliothèque du Congrès : <https://www.loc.gov/preservation/resources/rfs/>
- LOCKSS (Lots of Copies Keep Stuff Safe) : <https://www.lockss.org/>
- Open Research and Contributor ID (ORCID) : <https://orcid.org/>
- Registre de l'organisme de recherche (ID ROR ID) : <https://ror.org/>
- Software Package Data Exchange (SPDX-ID) : <https://spdx.dev/>
- Politique institutionnelle de l'UNIGE sur la gestion des données de recherche : <https://memento.unige.ch/doc/0320>
- Directives de l'UNIGE sur l'intégrité dans la recherche scientifique : <https://memento.unige.ch/doc/0003>
- Charte de l'UNIGE pour la science ouverte : <https://www.unige.ch/openscience/en/open-science-2>
- Directives sur l'application à l'Université de la loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD) : <https://memento.unige.ch/doc/0160>
- Loi sur l'information du public, l'accès aux documents et la protection des données personnelles (LIPAD) : <https://silgeneve.ch/legis/index.aspx>
- Règlement général sur la protection des données (RGPD) : <https://gdpr-info.eu/>
- Loi fédérale sur la recherche impliquant des êtres humains (LRH) : <https://www.fedlex.admin.ch/eli/cc/2013/617/en>
- Commission cantonale d'éthique et de recherche (CCER) : <https://www.ge.ch/organisation/commission-cantonale-ethique-recherche>
- Comité de l'UNIGE pour la recherche éthique (CUREG 2.0) : <https://cureg.unige.ch/>
- Yareta dans Re3data : <https://www.re3data.org/repository/r3d100013423>
- Yareta dans BASE (Bielefeld Academic Search Engine) : <https://www.base-search.net/Search/Results?q=dccoll:ftugeneveyareta>
- Yareta dans OpenAIRE : [https://explore.openaire.eu/search/dataprovider?datasourceId=re3data\\_::263180f806632244a29e88306d7eed5f](https://explore.openaire.eu/search/dataprovider?datasourceId=re3data_::263180f806632244a29e88306d7eed5f)
- Yareta dans OpenDOAR : <https://v2.sherpa.ac.uk/id/repository/10042>
- Fonds national suisse de la recherche scientifique (FNS) : <https://www.snf.ch/en>
- OLOS. Suisse : <https://access.olos.swiss/portal/>
- Réseau national suisse de soutien à la gestion des données de recherche (RDM support) : <https://forschungsdaten.info/fdm-im-deutschsprachigen-raum/schweiz/rdm-support-stakeholders/>
- Modèle DataTag : <http://datatags.org/datatags-compliant>

- Politique de sécurité des systèmes d'information de l'UNIGE : <https://memento.unige.ch/doc/0174>
- Site web de l'UNIGE e-Research : <https://www.unige.ch/eresearch/en/services/yareta/>
- Page gitlab du DLCM : <https://gitlab.unige.ch/dlcm/community/dlcm-backend>
- UNIGE Loi sur l'université du 13 juin 2008 : <https://silgeneve.ch/legis/index.aspx>
- Projet de loi cantonale genevoise "Infrastructure numérique pour la recherche" (PL 12146) 2018-2024 : <https://ge.ch/grandconseil/memorial/seances/010409/48/3/>
- Spécification de Yareta : <https://www.unige.ch/eresearch/en/services/yareta/specifications/>
- Guide d'utilisation de Yareta : <https://yareta.unige.ch/doc/Yareta-QuickStartGuide.html>
- Guide d'intégration de Yareta : <https://admin.yareta.unige.ch/administration/docs/DLCM-IntegrationGuide.html>
- Documentation sur les API RESTful de Yareta : <https://admin.yareta.unige.ch/administration/docs/DLCM-APIs.html>
- Guide des outils de Yareta : <https://admin.yareta.unige.ch/administration/docs/DLCM-ToolsGuide.html>
- SWITCHaai : <https://www.switch.ch/en/aai/>



Ce texte est mis à disposition sous la licence CC-BY-SA.

Crédit photographique : UNIGE, Jacques Erard, 2018 <https://phototheque.unige.ch/unige:121641>

Contact :

Recherche en ligne

Université de Genève

Division du système et des technologies de l'information et de la communication / IT Division

24 rue du Général-Dufour

1211 Genève 4

[eresearch@unige.ch](mailto:eresearch@unige.ch)

Octobre 2022