



HAL
open science

Plan de gestion du projet "Dictionnaire Électronique des Synonymes" du laboratoire CRISCO de l'Université de Caen

Laurette Chardon

► To cite this version:

Laurette Chardon. Plan de gestion du projet "Dictionnaire Électronique des Synonymes" du laboratoire CRISCO de l'Université de Caen. Université de Caen Normandie. 2023. hal-04317573

HAL Id: hal-04317573

<https://hal.science/hal-04317573>

Submitted on 1 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

DMP du projet "DES - Dictionnaire Électronique des Synonymes "

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré" fourni par Science Europe.

Renseignements sur le plan

Titre du plan	DMP du projet "DES - Dictionnaire Électronique des Synonymes "
Version	Version finale
Domaines de recherche (selon classification de l'OCDE)	
Langue	fra
Date de création	2022-05-31
Date de dernière modification	2023-12-01
Licence	Creative Commons Attribution 4.0 International
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	<ul style="list-style-type: none">• site web : https://crisco.unicaen.fr/des

Renseignements sur le projet

Titre du projet	DES - Dictionnaire Électronique des Synonymes
Acronyme	DES

Résumé

L'objectif du projet au départ est de développer un Dictionnaire Électronique des Synonymes sur la base de la symétrie et de relations partielles : si le mot A est synonyme du mot B dans un contexte donné, alors le mot B est synonyme du mot A.

Le Dictionnaire Électronique des Synonymes du CRISCO (DES) contient à ce jour plus de 50 000 entrées et 209 000 relations synonymiques réciproques.

La base de départ dans les années 1990 a été constituée à partir de sept dictionnaires classiques. Un premier travail, réalisé par l'INALF (Institut National de la Langue Française), a permis d'en extraire les relations synonymiques.

Le laboratoire ELSAP, qui est devenu par la suite le CRISCO, a ensuite concaténé, homogénéisé et symétrisé les données pour en créer une base (Cf DES-Bases de données). Depuis 1994, un important travail de correction se perpétue par l'ajout ou la suppression de liens synonymiques et antonymiques (Cf DES-maintenance).

Outre un accès en ligne public depuis le début des années 2000 (Cf DES-frontend et DES-Backend) et d'une version autonome accessible depuis 2015 (Cf DES-autonome), le projet s'est étoffé de nombreux outils de visualisations en 2D et 3D (espace sémantique, graphe d'adjacence, regroupement automatique de sens, ... : Cf DES-Visualisations), de communication (vidéo de présentation, devinettes : cf DES-communication) et d'extraction et de traitement des mots les plus recherchés (Cf DES-DataMotsRech. et DES-Trait.MotsRech.).

Le DES comptabilise de 150.000 à 200.000 requêtes par jour.

Le projet a débuté dans les années 90 sous la direction de deux chercheurs : Sabine Ploux et Bernard Victorri.

Il a ensuite été maintenu et amélioré par plusieurs ingénieurs : Jean-Luc Manguin, Michel Morel et actuellement Laurette Chardon.

Plusieurs stagiaires et prestataires de service sont également intervenus. Voir la page sur [l'historique du projet](#) sur le site web du CRISCO.

Sources de financement

- Université de Caen :

Date de début

1996-01-01

Produits de recherche :

1. Programmes pour l'accès en ligne du DES (frontend) (Logiciel)
2. Bases des graphies et des variantes, des liaisons synonymiques et antonymiques (Jeu de données)
3. Procédures de maintenance mensuelle (Workflow)
4. Programmes en backend pour l'accès en ligne (Logiciel)
5. Programme d'utilisation du DES sans accès internet (autonome)
6. Programmes de visualisation des données du DES (Logiciel)
7. Éléments de publications et de communication

Contributeurs

Nom	Affiliation	Rôles
Chardon Laurette - https://orcid.org/0000-0003-2091-884X	UNICAEN	<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact pour les données (Maintenance, AccèsBackend, Version Autonome, Données, Communication, DESVisu, AccèsFrontend) • Responsable de la documentation des données (Données) • Responsable du dépôt et de la diffusion des données (Données) • Responsable du plan de gestion de données • Responsable du traitement et de l'analyse des données (Données, Maintenance)

DMP du projet "DES - Dictionnaire Électronique des Synonymes "

Programmes pour l'accès en ligne du DES (frontend)

Nom	Programmes pour l'accès en ligne du DES (frontend)
Description	<p>Les données sont accessibles par une interface web : crisco.unicaen.fr/des Le mot recherché est saisi avec l'aide d'une complétion automatique. La requête retournée donne : la liste des synonymes, la liste des antonymes, le classement des premiers synonymes et la liste des cliques. DES-frontend contient les programmes en PHP exécutés à partir du lien d'accès direct (crisco.unicaen.fr/des). Cela concerne:</p> <ul style="list-style-type: none">• l'interface principale de requête (https://crisco.unicaen.fr/des)• l'interface des propositions (https://crisco4.unicaen.fr/des/proposition.php)
Type	Logiciel
Mots clés (texte libre)	frontend; php; des; dictionnaire ; synonymes
Langue	fra
Date de publication	2000-01-01
Identifiant pérenne	Pas d'identifiant pérenne
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui
Justification	Il n'est pas prévu de réutiliser les programmes. Le code est ancien et il n'y a aucun intérêt à le réutiliser.
Nom de la méthode	Aucune méthode
Description	Une documentation technique interne existe pour les programmes "AccèsFrontend" présente sur le git privé de l'Université : https://git.unicaen.fr/crisco-des/crisco-des-maintenance
Code langue des métadonnées	fra
Description	Les programmes ont été développés en interne au laboratoire. La personne en charge du DES assure la qualité des éventuelles modifications par ses tests.
Description	Aucune données à caractère personnel
Description	Aucune contrainte juridique
Description	Aucun aspect éthique

Description	La personne en charge du DES met en place ses propres moyens pour gérer les programmes.
Besoins de stockage	Les programmes sont régulièrement sauvegardés sur le git de l'Université.
Modalités de partage	Les programmes ne sont pas partagés
Potentiel de réutilisation	Aucun potentiel de réutilisation
Justification	Pour l'instant seul le git de l'Université de Caen conserve les programmes sur https://git.unicaen.fr/crisco-des/crisco-des-maintenance

Bases des graphies et des variantes, des liaisons synonymiques et antonymiques

Nom	Bases des graphies et des variantes, des liaisons synonymiques et antonymiques
Description	<p>Les données du DES :</p> <ul style="list-style-type: none"> - sont issues initialement de 7 dictionnaires. Ce processus a été initié par Bernard Victorri et Sabine Ploux en partenariat avec l'ATILF dans les années 90, - sont mises à jour mensuellement à partir des propositions des utilisateurs et à partir de calculs, - sont stockées dans une base de données, - sont extraites de la base de données dans un premier format texte, sont traitées par programme pour obtenir un second format texte utilisé dans la version en ligne (Cf le paragraphe Données FAIR plus bas). Un autre traitement créé d'autres formats (texte ou csv) pour la version en ligne (en cas de panne du moteur - Cf DES-backend) et pour la version autonome (sans accès internet), - sont utilisées pour créer les fichiers de l'espace sémantique, - sont utilisées pour créer d'autres formats de données notamment des graphiques. <p>Les données du DES sont accessibles sur la plate-forme ORTOLANG moyennant authentification pour la communauté scientifique.</p>
Type	Jeu de données
Mots clés (texte libre)	synonymes;base de données
Langue	fra
Date de publication	2022-11-07
Identifiant pérenne	https://www.ortolang.fr/market/lexicons/des
Type d'identifiant	URL
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui
Justification	Cette base de données des synonymes, variants et antonymes du DES peut être réutilisée pour tout travaux de recherche gratuitement et par contrat avec des entreprises privées.
Description	L'organisation, la structuration et la documentation des données est sur la plate-forme ORTOLANG .
Code langue des métadonnées	fra

Description	Pas d'aspect juridique particulier à prendre en compte.
Description	Pas d'aspect éthique particulier à prendre en compte.
Description	Les données sont traitées par des programmes (python et php principalement) détaillé dans le produit de recherche "Maintenance".
Besoins de stockage	Les données sont stockées sur le portable du développeur lors des traitements et transférés mensuellement sur le serveur git privé de l'Université de Caen (https://git.unicaen.fr/crisco-des/crisco-des-maintenance)
Volume estimé des données	1
Unité	Go
Modalités de partage	Les données sont accessibles auprès de la communauté scientifique sur la plate-forme ORTOLANG.
Potentiel de réutilisation	Tout traitement autour de la synonymie et polysémie.
Justification	Il n'y a pas de système d'archivage, les données étant mensuellement mises à jour. Le stockage sur le git privé de l'Université de Caen et le dépôt sur la plate-forme ORTOLANG assure actuellement une certaine pérennité.

Procédures de maintenance mensuelle

Nom	Procédures de maintenance mensuelle
Description	<p>Le DES-maintenance concerne toutes les opérations nécessaires à la mise à jour mensuelle du DES. Il contient :</p> <ul style="list-style-type: none"> - d'un fichier tableur des liens proposés.xls de type libreoffice calc contenant les propositions mensuelles des utilisateurs avec des commentaires. Un indicateur de traitement est ajouté : 0= regardé, 1=valide, 2= retenue ou déduite - d'un fichier tableur liens probables.xls de type libreoffice calc contenant les propositions probables calculées (Voir le calcul des liens manquants probables sur le site web du laboratoire), - de plusieurs programmes python regroupés dans un dépôt git privé de l'Université de Caen - d'une documentation technique générée automatiquement par l'outil sphinx dans le dépôt git qui décrit toutes les étapes à suivre : <ul style="list-style-type: none"> • la récupération des propositions des utilisateurs du mois précédent dans un nouvel onglet et leur traitement • l'exécution du programme sur les liens probables et leur traitement. • l'exécution des scripts et programmes python de mise à jour automatique de la BD • la mise à jour des statistiques à partir de l'outil Matomo • la mise à jour de l'espace sémantique en ligne • la mise à jour de la version du DES autonome • la récupération à partir des logs des mots les plus recherchés par jour et leur traitement
Type	Workflow
Mots clés (texte libre)	
Date de publication	2000-01-01
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui
Nom de la méthode	maintenance
Description	Les programmes de maintenance évoluent peu : quelques modifications mineures en fonctions de l'évolution des systèmes d'exploitation (Linux Ubuntu) et des langages de programmation (python, php et java).
Description	Les programmes sont documentés dans le code lui-même et une documentation pdf/html est générée avec l'outil sphinx installé dans le dépôt git.
Code langue des métadonnées	eng
Description	Pas de données à caractère personnel dans les programmes
Description	Pas d'aspect éthique
Description	Les programmes sont utilisés et mis à jour par le développeur qui utilise son propre environnement de développements adaptés

Besoins de stockage	Les programmes sont stockés sur le serveur git de l'université : https://git.unicaen.fr/crisco-des/crisco-des-maintenance
Volume estimé des données	30
Unité	Go

Programmes en backend pour l'accès en ligne

Nom	Programmes en backend pour l'accès en ligne
Description	DES-backend est le programme en C exécuté en tant que processus en arrière plan. Il lit les fichiers texte créés lors de la maintenance et les stocke en mémoire. Il répond aux requêtes du programme principal frontend en PHP. Le programme est dans un dépôt git privé de l'université de caen. Ce programme est optionnel : le DES peut fonctionner sans qu'il soit lancé. Il apporte simplement une fonctionnalité supplémentaire à savoir connaître les synonymes communs de deux mots.
Type	Logiciel
Mots clés (texte libre)	
Langue	fra
Date de publication	2008-01-01
Identifiant pérenne	Pas d'identifiant pérenne
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

Description Il n'y a pas de modifications majeures sur ce produit.

Description Pas de métadonnées. Documentation interne sur le git de l'Université.

Description Seul la développeuse en charge du produit met en place une méthode pour maintenir le programme.

Description Pas de contrainte juridique

Description Pas d'aspects éthiques à prendre en compte.

Description Le programme est géré par les moyens propres à la personne en charge du produit.

Besoins de stockage	Le programme est sauvegardé dans une instance privée sur le serveur git de l'Université.
Volume estimé des données	10
Unité	Mo
Mesures prises pour la sécurité des données	La sécurité est assurée par le service informatique de la DSI
Modalités de partage	le programme n'est pas partagé.
Potentiel de réutilisation	Pas de potentiel de réutilisation
Justification	La conservation à long terme est gérée par le service informatique de l'Université.

Programme d'utilisation du DES sans accès internet (autonome)

Nom	Programme d'utilisation du DES sans accès internet (autonome)
Description	Le DES autonome est une archive java (.jar) contenant les programmes en java et les données nécessaires pour une utilisation du DES sans connexion internet. L'ensemble du code est dans un dépôt git privé de l'université.
Mots clés (texte libre)	
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

Programmes de visualisation des données du DES

Nom	Programmes de visualisation des données du DES
Description	DES-visualisations contient l'ensemble des programmes javascript, programmes python autonomes et plug-ins python associés au logiciel Blender nécessaires pour les visualisations interactives en 2D et 3D. Ces visualisations concernent l' espace sémantique réalisées en javascript et un ensemble de visualisations accessibles sur le site web créées à partir du logiciel Blender .
Type	Logiciel
Mots clés (texte libre)	
Langue	fra
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

Éléments de publications et de communication

Nom	Éléments de publications et de communication
Description	DES-communication contient les vidéos, flyer de présentation, un jeu de devinettes (au format papier et avec une version interactive). Voir les actualités du DES sur le site web. Il contient également les publications réalisées sur le DES .
Mots clés (texte libre)	
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui