



DGP

/ PROJET GINCO V3 : GRAPHE-CULTURE

/ ATELIERS DE RESTITUTION - 22/11/2018

/ CONVENTION CADRE - MINISTÈRE DE LA
CULTURE & INRIA



Équipe Projet :

DGP

INRIA

JÉRÔME DAVID – ENSEIGNANT-CHERCHEUR COORDINATEUR DU PROJET - ÉQUIPE-PROJET MOEX

MINISTÈRE DE LA CULTURE :

DÉPARTEMENT DES SYSTÈMES D'INFORMATION PATRIMONIAUX

KATELL BRIATTE – COORDINATRICE MOA ET EXPERT GINCO

JIHEN LANDOLSI - CHEFFE DE PROJET MOA

SOUS-DIRECTION DES SYSTÈMES D'INFORMATION

FRÉDÉRIQUE ROLLAND – CHEF DE PROJET MOE

DÉPARTEMENT DE L'INNOVATION NUMÉRIQUE

MARIE-VÉRONIQUE LEROI - CHARGÉE DE MISSION - SPONSOR DU PROJET

- / Présentation du projet Ginco V3 Graphe-Culture : Objet, Contexte, Objectifs
- / GINCO
- / La Fonction d'alignement dans GINCO V2
- / Équipe MOEX-INRIA
- / Fonctionnement des Outils INRIA sur les vocabulaires du ministère
- / Unités de réalisation
- / Macro-Planning
- / Recommandations

« Gincó V3 : des outils d'aide à l'alignement pour enrichir le Graphe-Culture »

- / **BÉNÉFICIAIRES** : Producteurs et Utilisateurs d'informations culturelles
- / **OBJET** : Mettre en relation des vocabulaires culturels
- / **COMMENT** : Créer des correspondances entre des données culturelles de vocabulaires scientifiques différents
- / **POURQUOI** : Afin de créer de la connaissances culturelles

/ Lignée du programme HADOC :

Harmonisation des structures des données

Harmonisation des contenus par le partage de référentiels

/ En phase avec les orientations stratégiques du MC :

Feuille de Route Stratégique Métadonnées et Web 3.0 de 2014

Convention Cadre INRIA / MC de 2016

CONTEXTE

/ Contribuer à la création d'un Graphe-culture :

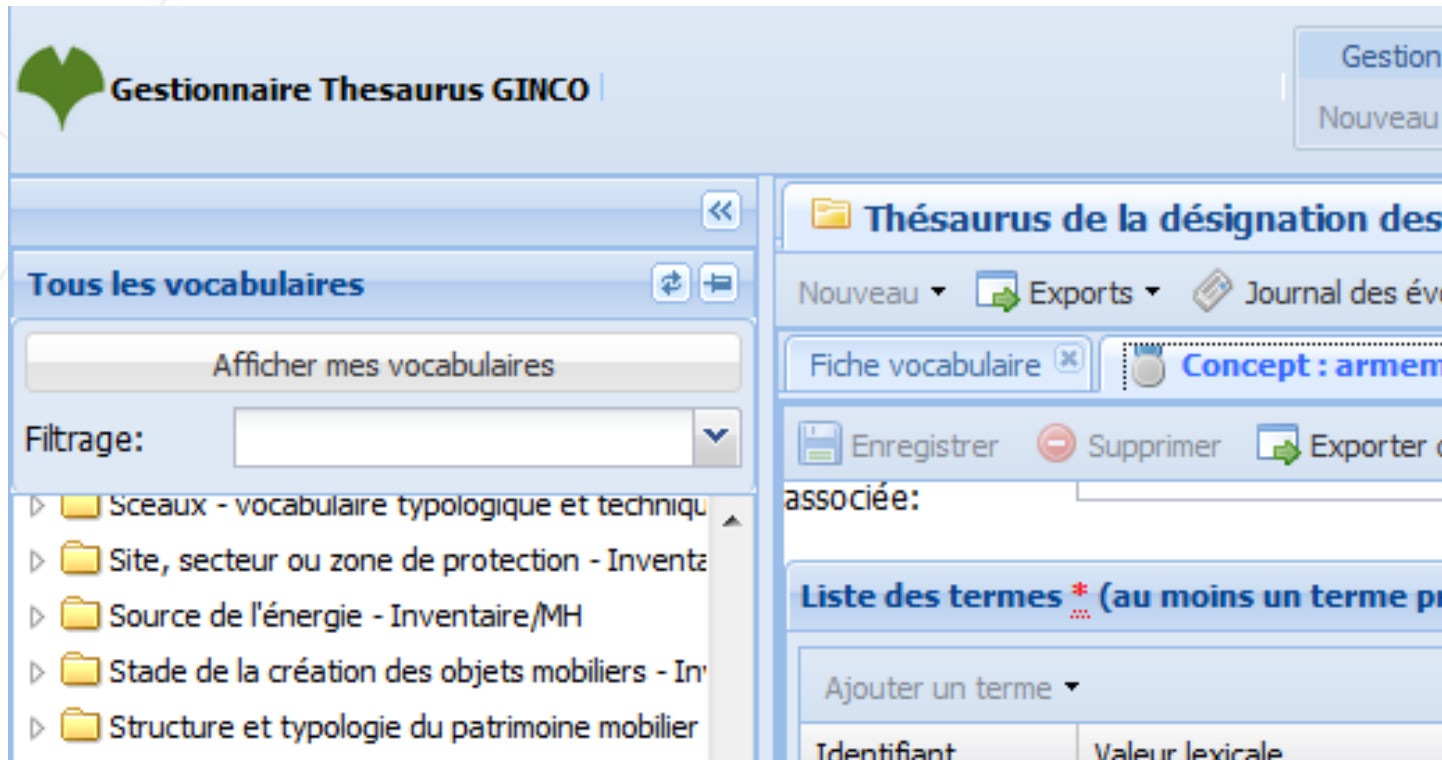
Interconnexion des grands référentiels culturels initiée par la stratégie Web 3.0 du ministère de la culture

/ Industrialiser la production d'alignements dans GINCO :

Amélioration de la productivité et Émergence de Vocabulaire « Pivots »

OBJECTIFS

- / Première brique applicative du programme HADOC
- / Outil de création et de gestion de Vocabulaires Scientifiques développé par le ministère de la culture et disponible sur la forge github.
- / GINCO V2 : fonction d'alignements réalisés «à la main»



The screenshot shows the 'Gestionnaire Thesaurus GINCO' interface. On the left, there is a sidebar titled 'Tous les vocabulaires' with a search bar 'Afficher mes vocabulaires' and a 'Filtrage:' dropdown. Below this is a list of folders: 'Sceaux - vocabulaire typologique et technique', 'Site, secteur ou zone de protection - Inventaire', 'Source de l'énergie - Inventaire/MH', 'Stade de la création des objets mobiliers - Inventaire', and 'Structure et typologie du patrimoine mobilier'. The main area on the right shows a selected folder 'Thésaurus de la désignation des' with a 'Nouveau' dropdown, 'Exports' button, and 'Journal des évènements' button. Below this is a 'Fiche vocabulaire' tab for 'Concept : armement'. It includes buttons for 'Enregistrer', 'Supprimer', and 'Exporter ce concept'. A section labeled 'associée:' is visible, followed by a 'Liste des termes * (au moins un terme par concept)'. At the bottom of this list, there is an 'Ajouter un terme' dropdown and a table with columns 'Identifiant' and 'Valeur lexicale'.

FONCTION D'ALIGNEMENT GINCO V2

The screenshot displays the GINCO V2 interface with three overlapping windows:

- GINCO Admin:** Shows the 'Concept: action sociale' page with sections for 'Concepts Parents (TG)', 'Concepts Enfants (TS)', 'Concepts de tête', 'Concepts Associés (TA)', and 'Alignements'. The 'Alignements' table lists external URIs for alignment:

Identifiant	Type	Relation	Concepts	Actions
http://data.cult...	Equivalence exa...	OU	Allez à http://data.culture.fr/thesaurus/resources/	✖
http://data.cult...	Equivalence exa...	OU	Allez à http://data.bnf.fr/ark:/12148/eb13318308/	✖
http://data.cult...	Equivalence exa...	OU	Allez à http://fr.dbpedia.org/resource/Action_sociale	✖

- data.bnf.fr:** Shows the 'Action sociale' entry with metadata:
 - Thème: Action sociale
 - Origine: RAMEAU - Notice originale
 - Note: Voir aussi la subdivision Participation des citoyens [+ subd. géogr communs]
 - Domaines: Problèmes et services sociaux. Criminologie
 - Autres formes du thème: Intervention (action sociale), Intervention sociale, Interventions sociales, Voir plus
- fr.dbpedia.org:** Shows the 'About: Action sociale' page with the definition: 'L'action sociale désigne l'ensemble des moyens par lesquels une société agit sur elle-même pour préserver sa cohésion, notamment par des dispositifs législatifs ou réglementaires et par des actions visant à aider les personnes ou les groupes les plus fragiles à mieux vivre, à acquérir ou à préserver leur autonomie et à s'adapter au milieu social environnant.'

Red arrows indicate the flow of information from the GINCO alignment table to the external data sources.

/ MOEX : équipe d'experts au sein d'INRIA qui a développé une suite d'outils afin de réconcilier les données d'ontologies complexes et de maintenir ces alignements.



Inria 50 ANS inventeurs du monde numérique

Scientifiques | Entrepreneurs | Candidats | Journalistes | Nous contacter

Inria Recherche Innovation En région Europe & International Agenda

EQUIPES DE RECHERCHE

Accueil > Recherches > Structures de recherche > Perception, Cognition, Interaction > Représentation et traitement des données et des connaissances > MOEX

Présentation Publications HAL

Equipe de recherche MOEX

Evolution de la connaissance

Responsable : Jerome Euzenat
Type : équipe
Centre(s) de recherche : Grenoble
Domaine : Perception, Cognition, Interaction
Thème : Représentation et traitement des données et des connaissances

Les "équipes" Inria sont en général des groupes de chercheurs engagés dans la définition d'un projet et d'objets partagés, qui donneront lieu à la création d'une équipe-projet. Les équipes-projets ainsi créées peuvent être commune avec des partenaires (universités ou organismes de recherche)

Partager

f Partager t Tweeter
in Partager e Envoyer

Contact

Responsable de l'équipe
Jerome Euzenat (voir toutes les équipes)
Tél: +33 4 76 61 53 66

Secrétariat
Tél: +33 4 76 61 54 59

API d'Alignement : Interface pour exprimer, manipuler, stocker et partager des alignements

AROMA : Algorithme automatique d'alignement

EDOAL : langage qui étend le format d'alignement pour des alignements complexes

/ COMMENT FONCTIONNENT LES ALGORITHMES INRIA ?

- Téléchargement de deux vocabulaires complets (dump)
- - Lancement d'un algorithme qui compare les deux ontologies selon une stratégie spécifique décidée en amont :

• **méthodes simples : *eqalign, subalign, relalign***

• **méthodes complexes = association de plusieurs méthodes et/ou « alignement d'alignement »**

- - Obtention des résultats pour chaque alignement : liste de correspondances trouvées selon les critères de la méthode choisie.
- L'api d'alignement est extensible : possibilité d'ajouter des méthodes d'alignements.

/ EXPÉRIMENTATION DES MÉTHODES

- Tests probants sur les vocabulaires du MC et des grands référentiels culturels – Attention toutefois les algorithmes n'ont pas été conçus pour traiter ce type de vocabulaire et un effort d'adaptation sera impérativement fait.

/ **Unité de Réalisation 1 :**

- Réalisation de l'outil d'administration « GINCO ALIGN »

/ **Unité de Réalisation 2 et 3 :**

- Réalisation d'une interface d'exploitation des correspondances
- Evolution version GINCO PROD

/ **Unité de Réalisation 4 :**

- Evolution de GINCO DIFF

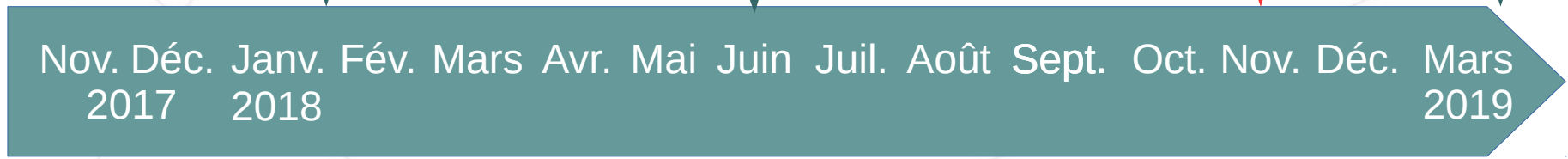
- Passage en COMOP
- Validation note d'opportunité

Lancement du projet

- Passage en COMOP
- Validation Note Cadrage

Réalisation

Contractualisation



Mise en place convention

Expérimentation des algorithmes & Expertises

Analyse du marché & Exp. Besoins

CDC & CCTP

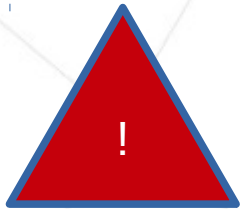
Opération 1

Opération 2

/ POINT DE VIGILANCE : Transfert de Technologie

Nous recommandons de consacrer du temps à l'étape de cadrage pour ce type de projet qui peut s'avérer être plus coûteux que prévu initialement.

Transformation d'une technologie Existante



≠

Élaboration d'un outil sur mesure

Exemples: algorithmes initialement non prévus pour comparer vocabulaires comme les nôtres – il faudra donc adapter l'API d'alignement (GINCO ALIGN) pour qu'il puisse s'intégrer dans nos environnements techniques et fonctionnels et puis ensuite adapter les suites GINCO PROD et GINCO DIFF pour exploiter les possibilités de cet outil.

- / Projet cœur de cible : Conforter les orientations Stratégique du MC
- / Valeur/Bénéficiaire d'un accompagnement et d'une expertise technique sur un domaine de niche (ex : Diagnostic éclair plus riche) qu'on espère maintenir lors de la phase de réalisation avec le prestataire.
- / Préserver la souplesse que laisse le dispositif en matière d'organisation (exemple : autonomie du planning projet etc.)

Merci de votre écoute
/ DISCUSSION

ANNEXES

Opération 1

ADAPTATION DES OUTILS MOEX AU CHAMP CULTUREL

Cette opération correspond à un transfert de technologies vers le Ministère de la Culture.

Mise en œuvre d'une suite d'outils permettant de mettre en relation des vocabulaires scientifiques dans le domaine culturel.

Opération 2

ÉVOLUTION DE L'OUTIL GINCO

Exploiter ces technologies pour faciliter l'interconnexion des grands référentiels culturels

Permettre à l'application GINCO d'exploiter les jeux de correspondances proposés par ces outils afin d'enrichir le graphe culture.

- / RUN – Utilisateurs : Valider/rejeter des propositions de correspondances candidates au sein de chaque fiche concept
- / BUILD - Resp. Vocabulaire : Valider / rejeter des propositions de correspondances candidates par vocabulaire
- / ADMINISTRATION : Disposer d'une interface de gestion back office : administration des campagnes, statistiques, etc.
- / **EXIGENCE** : l'outil final aura la souplesse de permettre l'ajout de nouveaux algorithmes d'alignements.

- / Gain de temps / coûts RH pour les producteurs de données
- / Gain en Qualité des données
- / Impulser GINCO au rang d'outil de référence dans les domaines de la gestion des terminologies scientifiques
- / Mise en valeur des institutions culturelles françaises
- / Démocratiser l'accès aux technologies du web sémantique
- / Permettre à l'ensemble des producteurs de données culturelles de bénéficier des apports du programme HADOC à moindre coût.
- / Offrir aux usagers une base de connaissance ouverte et réutilisable

- / Dépendance forte sur la qualité et la stabilité de la donnée : Engagement des producteurs de vocabulaires à maintenir leurs référentiels.
- / Renforcement de la fonction d'administration fonctionnelle et technique : augmentation de la charge de travail, animation d'une communauté d'utilisateurs
- / Disponibilité des acteurs pour contribuer à la construction du graphe culture

Refalign : est l'alignement de référence extrait à partir des exactMatch.

Eqalign : Deux concepts sont alignés si ils possèdent au moins un label en commun.

Subalign : Deux concepts sont alignés si ils partagent des fils (directs) qui sont alignés entre eux. Un alignement de départ doit être fourni. Dans cette expérience l'alignement de départ est eqalign.

Relalign : Deux concepts sont alignés si ils partagent des skos :related qui sont alignés entre eux. Un alignement de départ doit être fourni. Dans cette expérience l'alignement de départ est eqalign.

Tests des Algorithmes INRIA pour vérifier leur pertinences sur nos vocabulaires

Voc 1 : Objets mobiliers

Voc 2: Architecture + AAT + DBPedia + DataBNF

Méthodes : Refalign, Ealign, subalign, relalign

Objet : Comparer nos alignements actuels à ceux des méthodes INRIA → Déterminer les taux de Précision/rappel (*ie % des correspondances faites à la mains, retrouvées par la méthode utilisée et le % de celles trouvées en plus*)

La méthode simple qui consiste à établir une correspondance lorsque deux concepts partagent au moins label en commun permet de **retrouver la majorité des correspondances manuelles existante.**

Les résultats les plus faibles ont été obtenus sur Rameau. Cela a pu être partiellement résolu en supprimant les « s » en fin de mots.

Les méthodes subalign et relalign permettent de trouver quelques correspondances ignorées par eqalign mais semble générer pas mal de bruit. *Cela reste à vérifier par des experts des jeux de données.* Le filtrage entre bonnes correspondances et bruit semble être facilement possible *en seuillant sur les valeurs de confiances* retournées par les méthodes.