

Le projet Paleosaurus et les données ouvertes

Valérie Daux, Ludovic Hamiaux, Henri Bretel

Paleosaurus: un thesaurus de paléoclimatologie

Le thesaurus est...

un répertoire structuré de termes pour l'analyse de contenu et le classement de documents.

qui hiérarchise un vocabulaire...

- contrôlé (concepts définis, uniques et univoques)
- spécifique (définitions valides dans le contexte du thesaurus et non en dehors)

et qui évolue.

Le thesaurus à pour ambition...

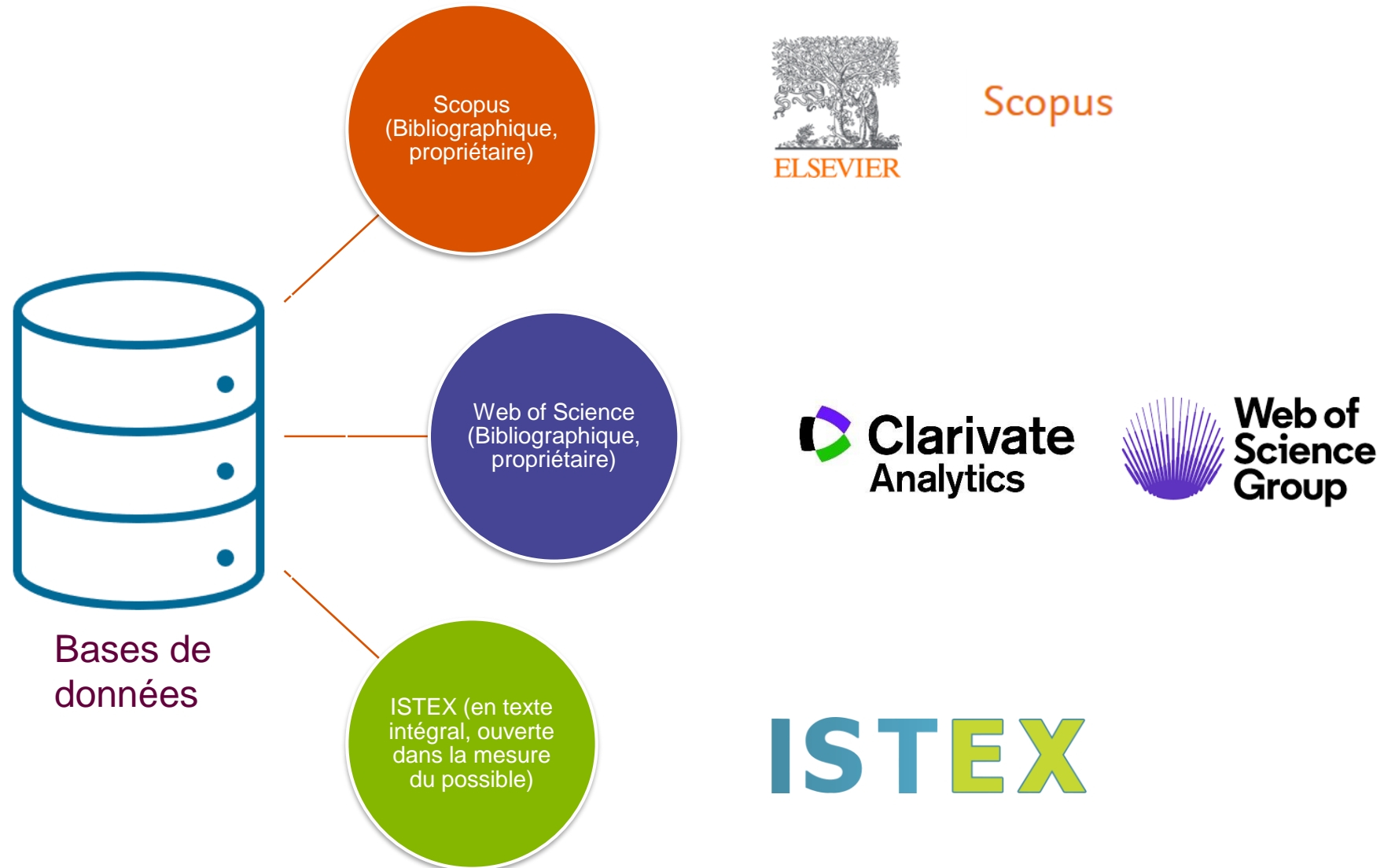


Améliorer le partage des données et des productions scientifiques en paléoclimatologie sur le web

La collecte de données pour Paléosaurus

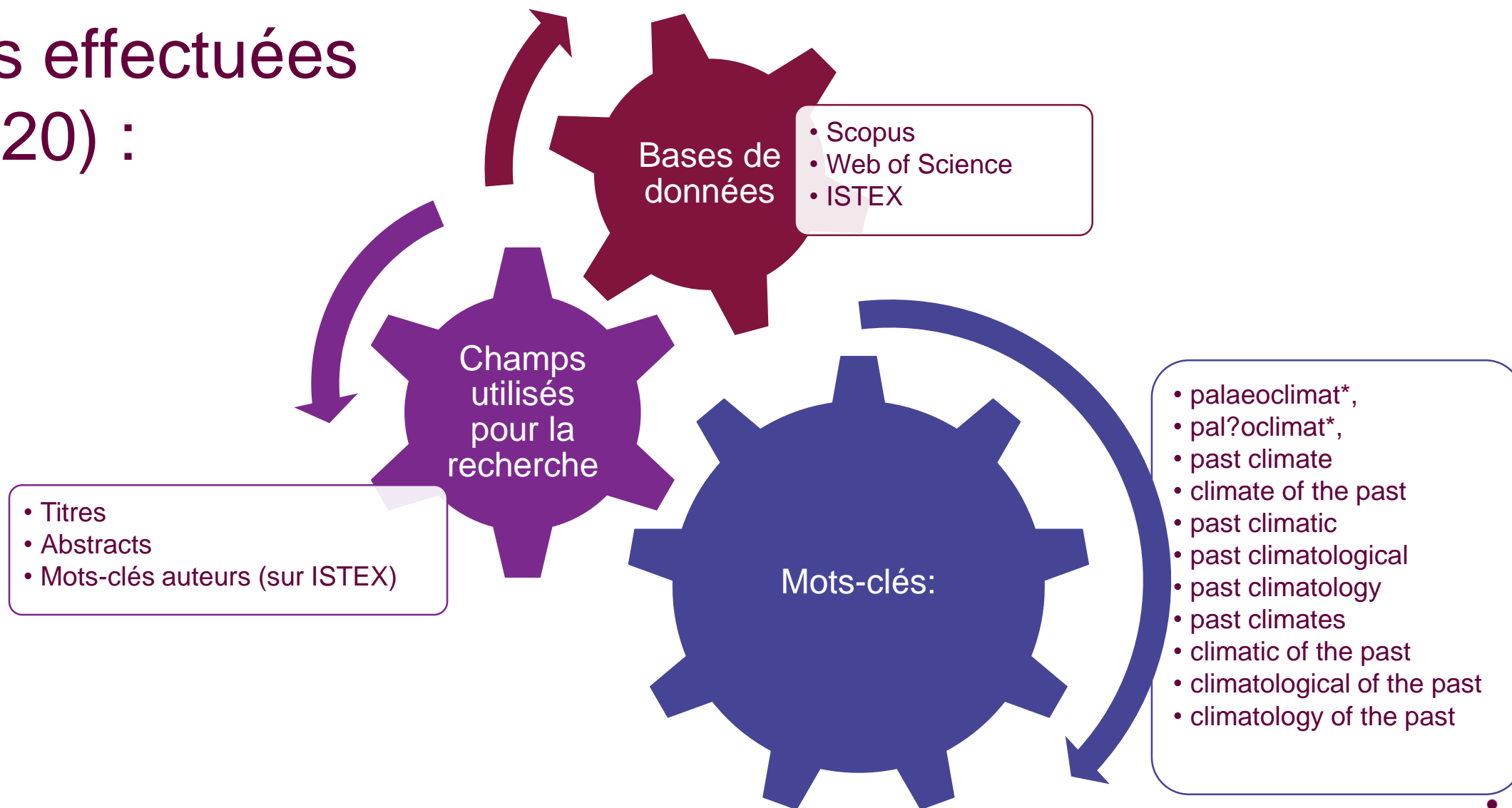
Henri Bretel

Première étape : définition d'un corpus

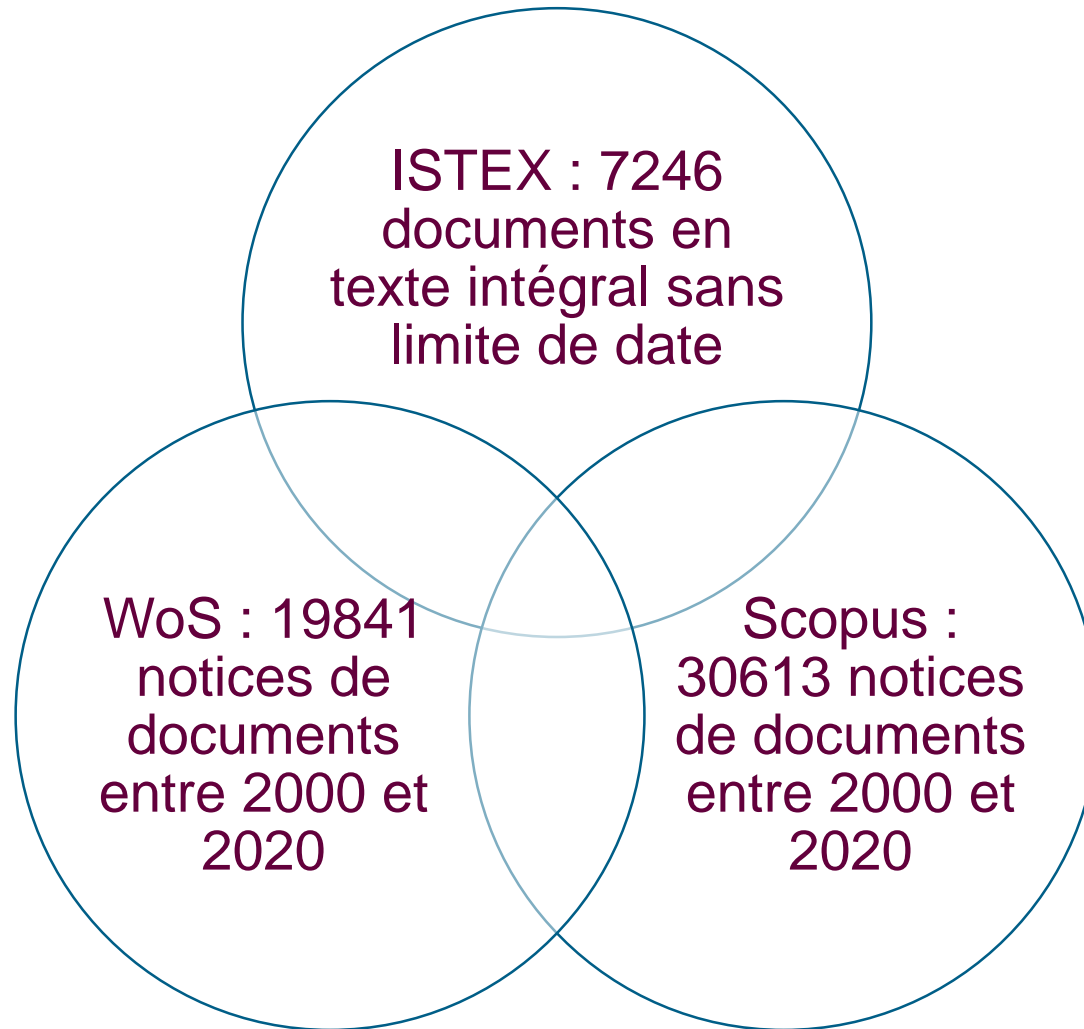


Première étape : définition d'un corpus

Requêtes effectuées
(mars 2020) :



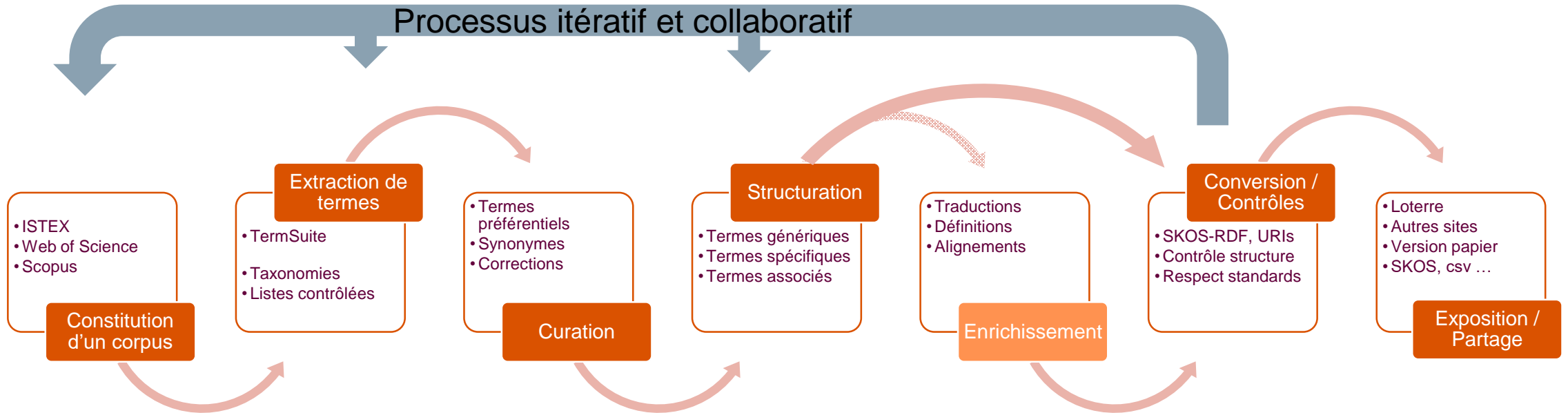
Première étape : définition d'un corpus



Extraction terminologique, affinage et structuration des données pour Paleosaurus

Ludovic Hamiaux

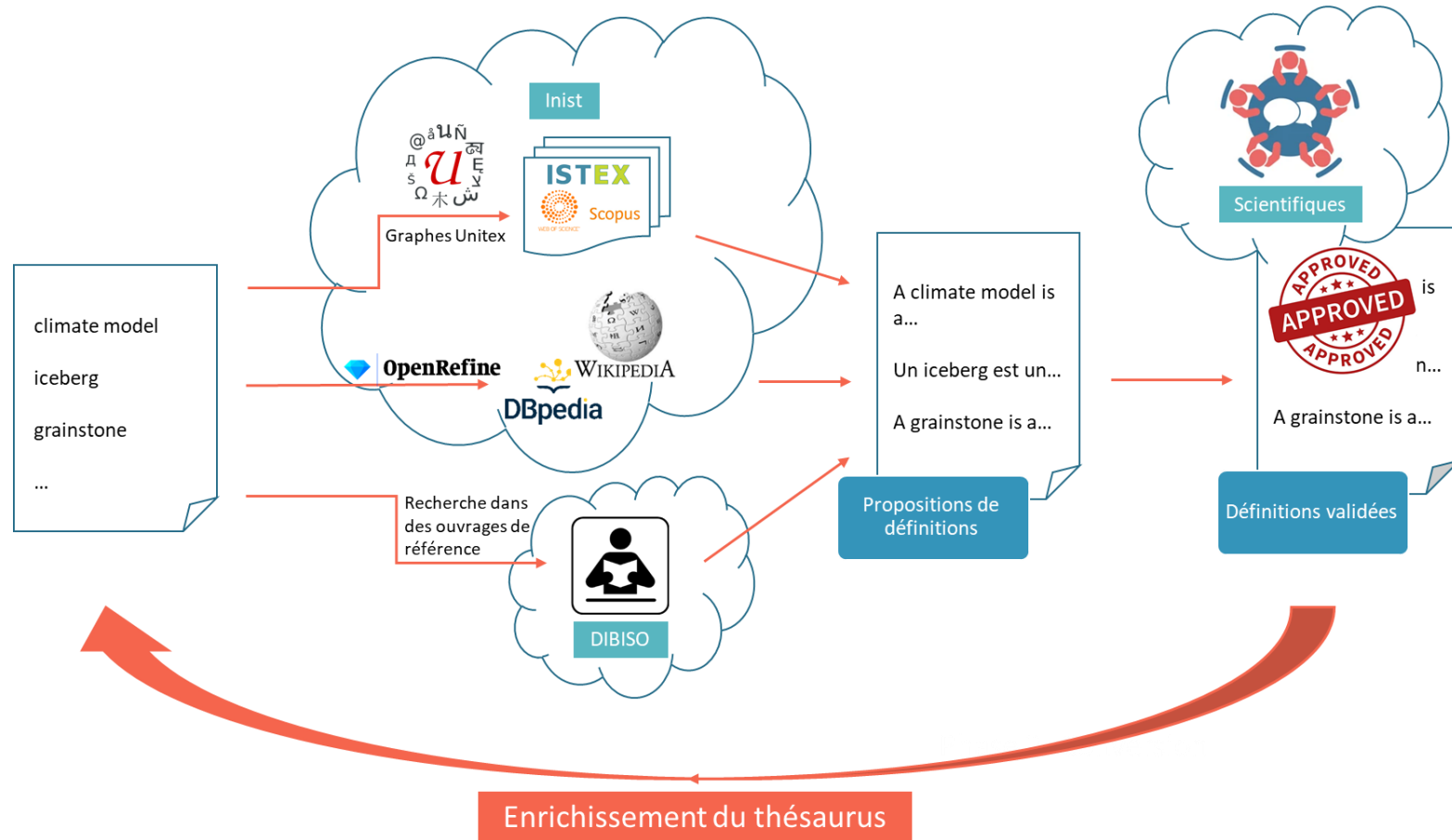
Etapes de création du thésaurus



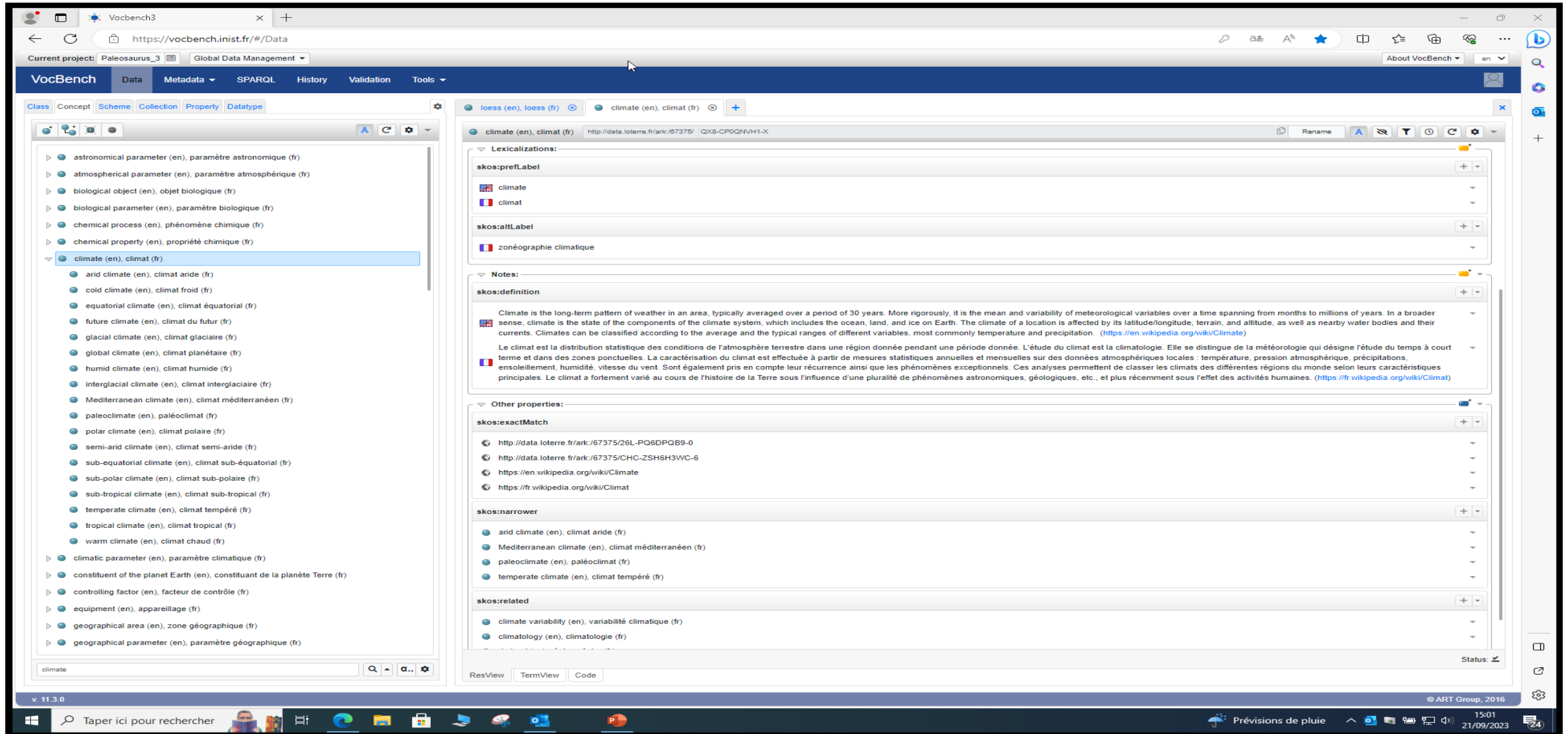
Constitution du thésaurus selon un processus itératif et collaboratif

- **Extraction** des termes à l'aide du logiciel TermSuite : obtention d'environ 120 000 termes et 24 000 variantes dont sont ressorties 46 000 formes après un premier tri, parmi lesquelles seulement 7 000 (d'une fréquence supérieure ou égale à 100) ont en définitive été retenues ;
- **Curation** (par les chercheurs) ayant abouti à environ 2 000 concepts constituant les entrées du futur thésaurus ;
- **Structuration** proprement dite à l'aide du logiciel VocBench : choix d'une trentaine de Top concepts, termes génériques, termes spécifiques, termes associés... ;
- **Enrichissements** : traduction des termes en français, ajout de synonymes et de définitions, à la fois en anglais et en français (par le biais des logiciels OpenRefine et Unitex), et alignements en direction de Wikipédia, du vocabulaire Inist des Sciences de la Terre et du thésaurus Inist sur le changement climatique ;
- **Conversion** (SKOS-RDF), **contrôles** et **exposition** sur la plateforme LOTERRE (<https://skosmos.loterre.fr/QX8/fr/>)

Enrichissement des concepts sur l'exemple des définitions



Exemple du concept « Climat » dans VocBench

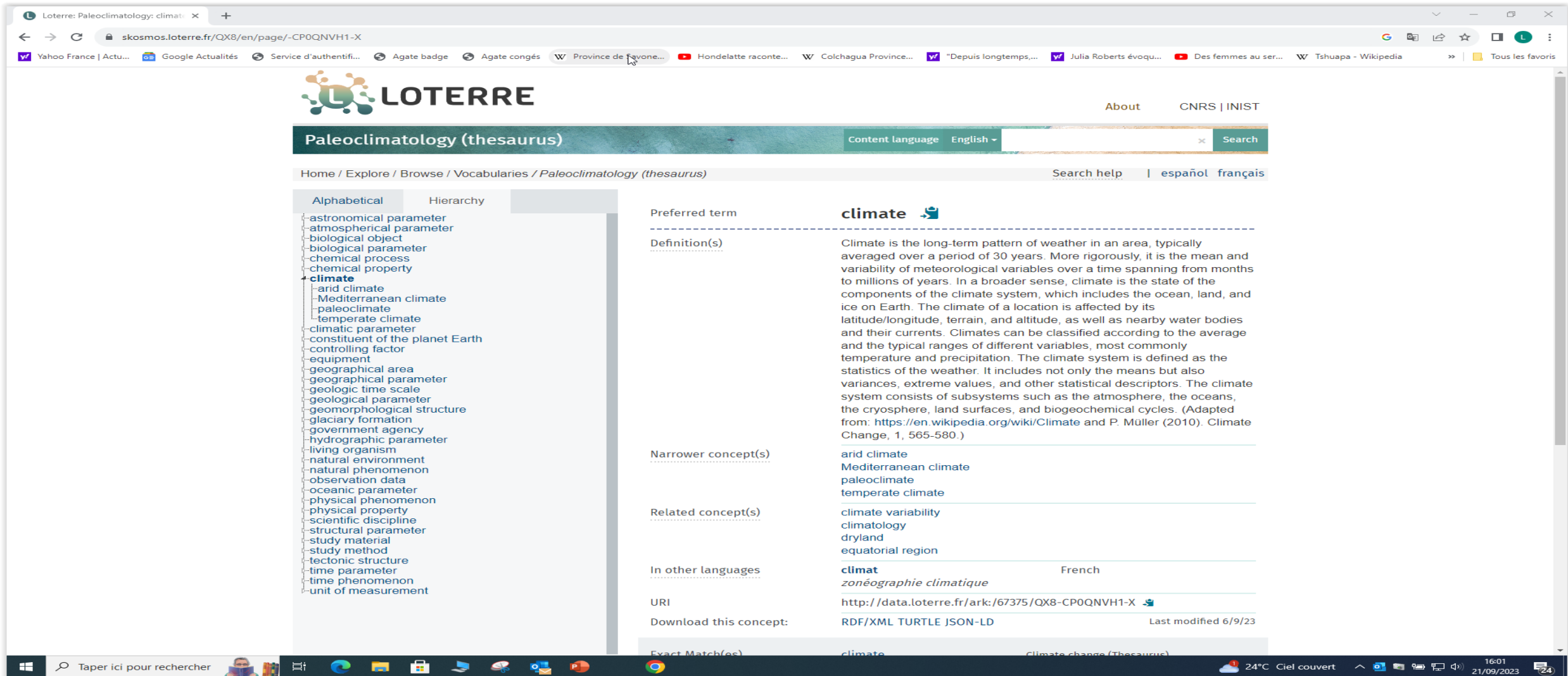


The screenshot displays the VocBench interface for the concept 'climate'. The left sidebar shows a hierarchical tree of concepts, with 'climate (en), climat (fr)' selected. The main content area is divided into several sections:

- Lexicalizations:**
 - skos:prefLabel:** Lists 'climate' (English) and 'climat' (French).
 - skos:altLabel:** Lists 'zonéographie climatique' (French).
- Notes:**
 - skos:definition:** Contains a detailed English definition: "Climate is the long-term pattern of weather in an area, typically averaged over a period of 30 years. More rigorously, it is the mean and variability of meteorological variables over a time spanning from months to millions of years. In a broader sense, climate is the state of the components of the climate system, which includes the ocean, land, and ice on Earth. The climate of a location is affected by its latitude/longitude, terrain, and altitude, as well as nearby water bodies and their currents. Climates can be classified according to the average and the typical ranges of different variables, most commonly temperature and precipitation." followed by a French translation: "Le climat est la distribution statistique des conditions de l'atmosphère terrestre dans une région donnée pendant une période donnée. L'étude du climat est la climatologie. Elle se distingue de la météorologie qui désigne l'étude du temps à court terme et dans des zones ponctuelles. La caractérisation du climat est effectuée à partir de mesures statistiques annuelles et mensuelles sur des données atmosphériques locales : température, pression atmosphérique, précipitations, ensoleillement, humidité, vitesse du vent. Sont également pris en compte leur récurrence ainsi que les phénomènes exceptionnels. Ces analyses permettent de classer les climats des différentes régions du monde selon leurs caractéristiques principales. Le climat a fortement varié au cours de l'histoire de la Terre sous l'influence d'une pluralité de phénomènes astronomiques, géologiques, etc., et plus récemment sous l'effet des activités humaines." with a link to the French Wikipedia page.
- Other properties:**
 - skos:exactMatch:** Lists URIs for 'http://data.loterre.fr/ark:/67375/26L-PQ6DPQB9-0', 'http://data.loterre.fr/ark:/67375/CHC-ZSH6H3WC-6', 'https://en.wikipedia.org/wiki/Climate', and 'https://fr.wikipedia.org/wiki/Climat'.
 - skos:narrower:** Lists 'arid climate (en), climat aride (fr)', 'Mediterranean climate (en), climat méditerranéen (fr)', 'paleoclimate (en), paléoclimat (fr)', and 'temperate climate (en), climat tempéré (fr)'.
 - skos:related:** Lists 'climate variability (en), variabilité climatique (fr)' and 'climatology (en), climatologie (fr)'.

At the bottom of the interface, there are tabs for 'ResView', 'TermView', and 'Code', and a status bar indicating 'v. 11.3.0' and '© ART Group, 2016'.

Exemple du concept « Climat » dans LOTERRE



The screenshot shows a web browser displaying the LOTERRE website. The page title is "Paleoclimatology (thesaurus)". The main content area is titled "climate" and includes a definition, narrower concepts, related concepts, and other languages. The left sidebar shows a list of terms under "Alphabetical" and "Hierarchy" views.


LOTTERRE About CNRS | INIST

Paleoclimatology (thesaurus) Content language English Search

Home / Explore / Browse / Vocabularies / Paleoclimatology (thesaurus) Search help | español français

Alphabetical Hierarchy

- astronomical parameter
- atmospherical parameter
- biological object
- biological parameter
- chemical process
- chemical property
- climate**
- arid climate
- Mediterranean climate
- paleoclimate
- temperate climate
- climatic parameter
- constituent of the planet Earth
- controlling factor
- equipment
- geographical area
- geographical parameter
- geologic time scale
- geological parameter
- geomorphological structure
- glaciary formation
- government agency
- hydrographic parameter
- living organism
- natural environment
- natural phenomenon
- observation data
- oceanic parameter
- physical phenomenon
- physical property
- scientific discipline
- structural parameter
- study material
- study method
- tectonic structure
- time parameter
- time phenomenon
- unit of measurement

climate 

Preferred term

Definition(s)

Climate is the long-term pattern of weather in an area, typically averaged over a period of 30 years. More rigorously, it is the mean and variability of meteorological variables over a time spanning from months to millions of years. In a broader sense, climate is the state of the components of the climate system, which includes the ocean, land, and ice on Earth. The climate of a location is affected by its latitude/longitude, terrain, and altitude, as well as nearby water bodies and their currents. Climates can be classified according to the average and the typical ranges of different variables, most commonly temperature and precipitation. The climate system is defined as the statistics of the weather. It includes not only the means but also variances, extreme values, and other statistical descriptors. The climate system consists of subsystems such as the atmosphere, the oceans, the cryosphere, land surfaces, and biogeochemical cycles. (Adapted from: <https://en.wikipedia.org/wiki/Climate> and P. Müller (2010). Climate Change, 1, 565-580.)

Narrower concept(s)

- arid climate
- Mediterranean climate
- paleoclimate
- temperate climate

Related concept(s)

- climate variability
- climatology
- dryland
- equatorial region

In other languages

Language	Term
French	climat
	zonéographie climatique

URI

<http://data.loterre.fr/ark:/67375/QX8-CP0QNVH1-X>

Download this concept:

RDF/XML TURTLE JSON-LD Last modified 6/9/23

Exact Match(es)

- climate
- Climate change (Thesaurus)

En résumé (au sujet de la structuration) :

Import des concepts dans VocBench (outil collaboratif d'édition et de gestion de terminologies) après conversion en SKOS, puis enrichissements et hiérarchisation à l'aide de cet outil.

Travaux en cours :

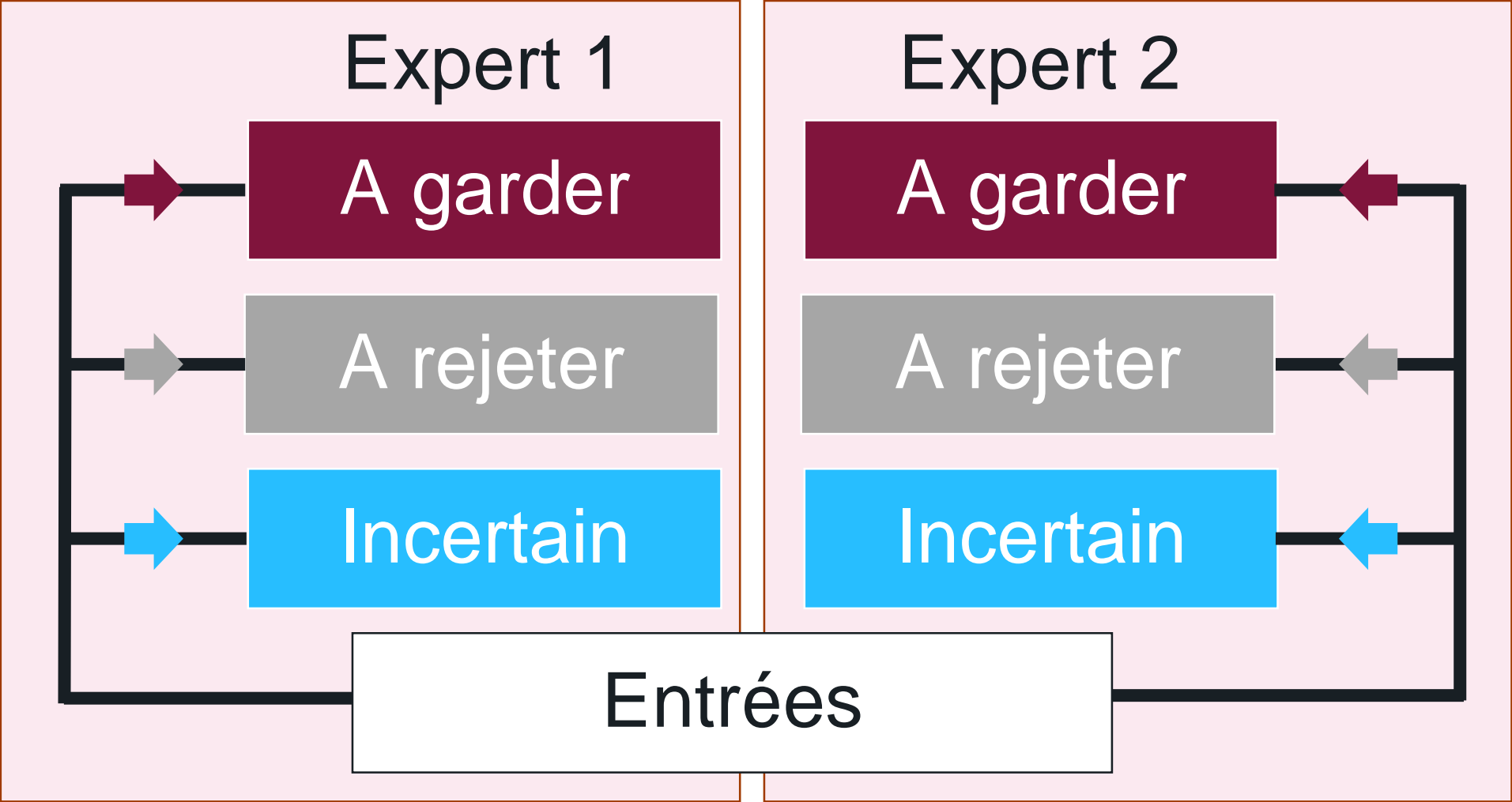
Ajouts d'entrées supplémentaires (génériques et spécifiques), de synonymes, de définitions, de relations de type termes associés... Complétion des listes de systématiques (roches, sols, minéraux, organismes végétaux et animaux, instruments et méthodes d'analyse, composés chimiques...), et mises à jours régulières de LOTERRE.



La sélection du vocabulaire scientifique pour Paleosaurus

Valérie Daux

- **7000 termes** à inspecter, classer, associer
- une dizaine de **spécialistes des paléoclimats** des Universités Paris-Saclay (LSCE et GEOPS) et de Montpellier



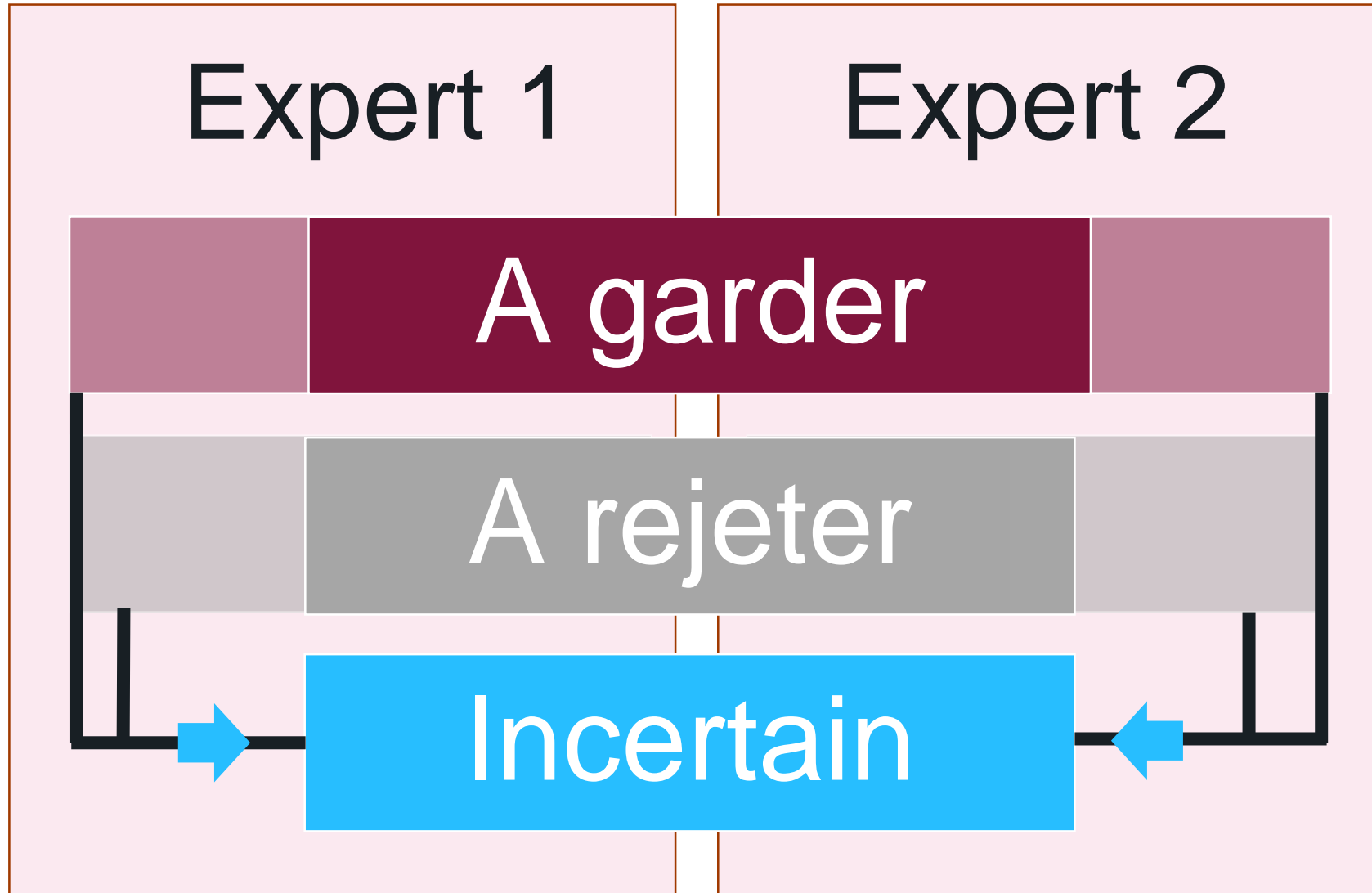
Expert 1

Expert 2

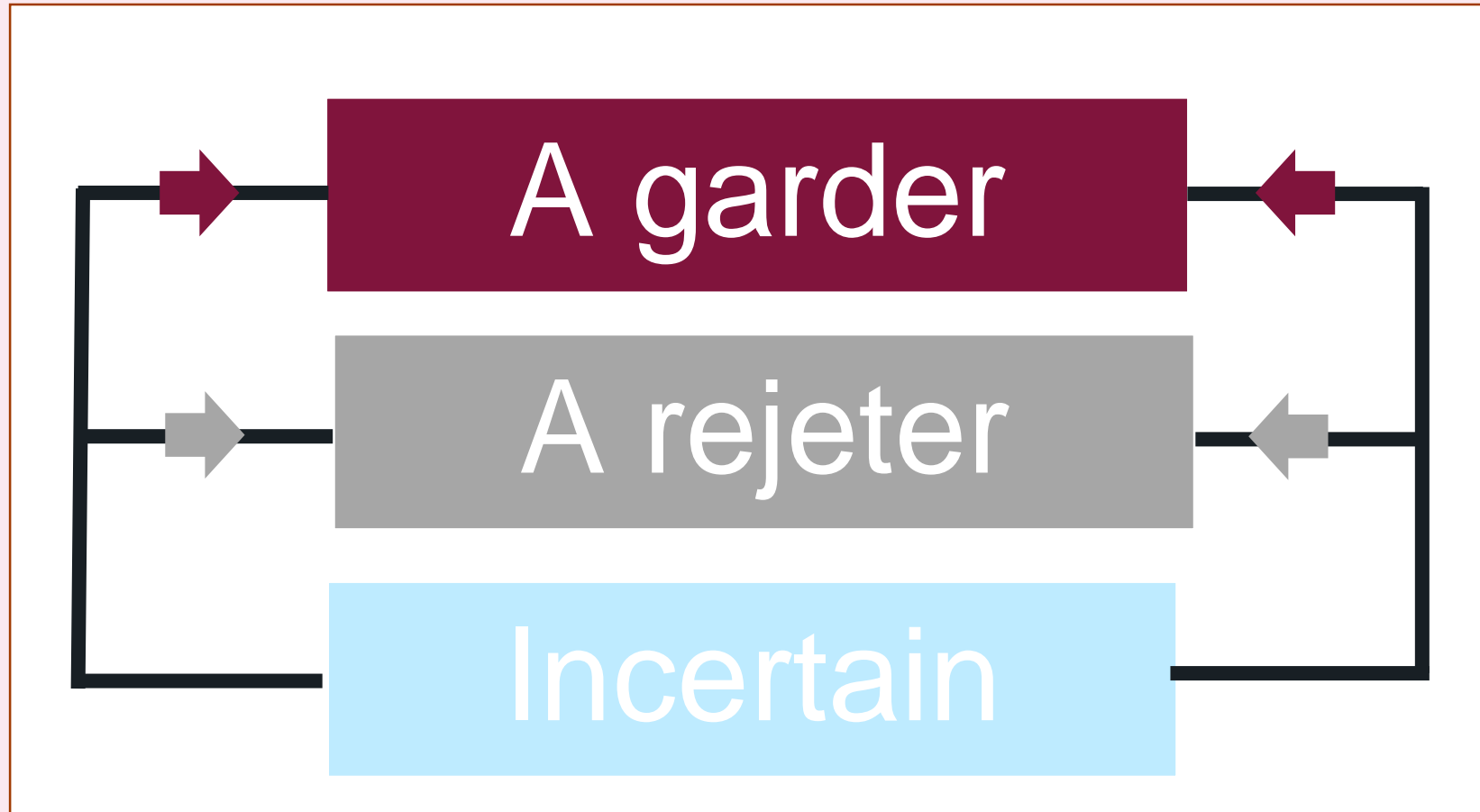
A gardner

A rejeter

Incertain



Tous les experts



Tous les experts

Liste finale de 2000
termes

600 définitions

Présence de Paleosaurus dans l'écosystème de la science ouverte

- Thésaurus disponible sur plateforme Loterre <https://skosmos.loterre.fr/QX8/fr/>



A propos CNRS | INIST

Paléoclimatologie (thésaurus)

Langue des données français

Chercher

Accueil / Explorer / Naviguer / Vocabulaires / Paléoclimatologie (thésaurus)

Aide à la recherche | español English

Liste Hiérarchie

A À Á B C D E É F G H I J K L
M N O P Q R S T U V W X Y Z

Aalénien
abondance naturelle
abondance relative
Acadien
accident climatique → anomalie climatique
accrétion
accumulation
accumulation en masse
acide
acide aminé
acide chlorhydrique
acide humique
acide muriatique → acide chlorhydrique
acide n-alkanoïque
acide nitrique
acide organique
acidification
acidité
action anthropique → effet anthropogénique
action anthropogène → effet anthropogénique
action anthropogénique → effet anthropogénique
action de l'homme → effet anthropogénique
action humaine → effet anthropogénique
activité anthropique → activité anthropogénique
activité anthropogène → activité anthropogénique
activité anthropogénique
activité cyclonique
activité humaine → activité anthropogénique
activité microbienne
activité solaire
ADN
ADN mitochondrial
ADN nucléaire
adsorption
advection
Adélaïdien
affleurement
Aftonien
Agnotozoïque → Protérozoïque
aimantation
aimantation rémanente
aimantation rémanente anhystérique
aire biotique → biome
Akchagylen

Description du vocabulaire

Titre	Paléoclimatologie (thésaurus) Paléoclimatologie Paleoclimatología (tesauro)
Description	Fruit d'une collaboration entre paléoclimatologues et experts de l'information scientifique le thésaurus de paléoclimatologie décrit et structure pour la première fois environ 2000 termes et concepts de la discipline, en français et en anglais. Destiné à offrir à la communauté scientifique un vocabulaire de référence, il facilite le partage, sur le web, des données et des productions scientifiques en paléoclimatologie.
Créateur(s)	<ul style="list-style-type: none">• LSCE : Laboratoire des Sciences du Climat et de l'Environnement• Géosciences Montpellier• DiBISO : Direction des Bibliothèques, de l'Information et de la Science Ouverte• Institut de l'information scientifique et technique (Inist) - CNRS/UAR76• GEOPS : Laboratoire de Géosciences Paris Saclay
Langue(s)	http://lexvo.org/id/iso639-3/eng http://lexvo.org/id/iso639-3/fra
Version	1.3
Date de création	vendredi 20 août 2021 00:00:00
Date de dernière modification	vendredi 9 juin 2023 00:00:00
Nom d'attribution	Institut de l'information scientifique et technique (Inist) - CNRS/UAR76
URL d'attribution	http://www.inist.fr

Présence de Paleosaurus dans l'écosystème de la science ouverte

- Poster de présentation sur HAL <https://universite-paris-saclay.hal.science/hal-03895146/document>

Présenté à l'occasion de la 7^e édition des journées « Climat et Impacts » qui s'est tenue du 23 au 25 novembre 2022 à l'université Paris-Saclay.

Un nouveau thésaurus de paléoclimatologie : création et perspectives d'utilisation

Banzet, Gilles [1], Bassinot, Franck [2], Bretel, Henri [3], Daux, Valérie [2], Hamiaux, Ludovic [1], Hatté, Christine [2], Kissel, Catherine [2], Mercier, Cédric [3], Penchenat, Tiphaine [2], Rousseau, Denis-Didier [4], Sempéré, Julien [3], Sepulcre, Sophie [5], Suhan, Stela [3], Vedovotto, Nathalie [1].
1. CNRS, INIST
2. Université Paris-Saclay, CNRS, CEA, UVSQ, LSCE
3. Université Paris-Saclay, DIBISO
4. Université de Montpellier, Géosciences Montpellier
5. Université Paris-Saclay, CNRS, GEOPS

Réalisation du poster : Stela Suhan



Introduction

Fruit d'une collaboration entre paléoclimatologues et experts de l'information scientifique et technique, ce thésaurus de paléoclimatologie décrit et structure pour la première fois environ 2000 concepts de la discipline, en français et en anglais.

Destiné à offrir à la communauté scientifique un vocabulaire de référence, il facilite le partage, sur le web, des données et des productions scientifiques en paléoclimatologie.



Etapes de création

1	Constitution d'un corpus de références bibliographiques et d'articles à l'aide du WOS, de Scopus et d'ISTEX	2	Extraction terminologique via TermSuite : production d'une liste de 7000 termes	3	Examen et choix de termes préférentiels : correction et validation des concepts
S1 2020 → S2 2020 → 2021			↓		
S2 2022 ← S1 2022 ← 2021			↓		
6	Structuration finale et publication du thésaurus sur la plateforme Loterre	5	Enrichissement du thésaurus avec des définitions	4	Conversion au format SKOS et premier essai de structuration

Perspectives d'utilisation

- ❖ Accessible, consultable et téléchargeable librement en ligne : <https://skosmos.loterre.fr/QX8> ;
- ❖ Conçu pour contribuer à une uniformité des termes utilisés et régler les problèmes de polysémie et d'homonymie
- ❖ Référentiel pour relier plus facilement les termes utilisés aux données qui leurs sont attribuées ;
- ❖ Outil de travail pour mieux explorer et partager les connaissances dans le cadre du web sémantique.



SCAN ME

Inist cnrs université PARIS-SACLAY LSCE Géosciences Montpellier GEOPS GÉOSCIENCES PARIS SACLAY

Présence de Paleosaurus dans l'écosystème de la science ouverte

- Article en cours de finalisation :
« Paleosaurus, un thesaurus en paléoclimatologie: création et perspectives d'utilisation »

PALEOSAURUS, UN THESAURUS EN PALÉOCLIMATOLOGIE : CRÉATION ET PERSPECTIVES

D'UTILISATION

Banzet, Gilles [1], Bassinot, Franck [2], Bretel, Henri [3], Daux, Valérie [2], Hamiaux, Ludovic [1], Hatté, Christine [2], Kissel, Catherine [2], Mercier, Cédric [3], Penchenat, Tiphaine [2], Rousseau, Denis-Didier [4], Sempéré, Julien [3], Sepulcre, Sophie [5], Suhan, Stela [3], Vedovotto, Nathalie [1].

1. CNRS, Inist
2. Université Paris-Saclay, CNRS, CEA, UVSQ, LSCE
3. Université Paris-Saclay, DiBISO
4. Université de Montpellier, Géosciences Montpellier
5. Université Paris-Saclay, CNRS, GEOPS

Résumé

Fruit d'une collaboration entre paléoclimatologues et experts de l'information scientifique et technique, le thesaurus de paléoclimatologie, dénommé Paleosaurus, décrit et structure pour la première fois environ 2000 concepts de la discipline, en français et en anglais.

Destiné à offrir à la communauté scientifique un vocabulaire de référence, l'ambition du thesaurus vise à faciliter le partage, sur le web, des données et des productions scientifiques en paléoclimatologie. Ses caractéristiques lexicales et normatives sont conformes aux standards du web sémantique. Il permet ainsi d'aligner les concepts de paléoclimatologie sur les systèmes internationaux d'organisation des connaissances.

Dans cet article, nous présentons les étapes de la création de Paleosaurus, les outils logiciels employés, ainsi que ses utilisations potentielles.

- Vidéo de présentation en cours d'élaboration

« Le Trésaurus de paléoclimatologie PALEOSAURUS
Qu'est-ce et pourquoi faire »



PALEOSAURUS

