HAL open science

Thésaurus d'agroécologie : une approche par les usages et avis d'experts

Veronique Batifol, Sophie Aubin, Laurent Hazard, Jean-Pierre Theau, Cathy Bouffartigue, Marie-Benoît Magrini

▶ **To cite this version:**

Veronique Batifol, Sophie Aubin, Laurent Hazard, Jean-Pierre Theau, Cathy Bouffartigue, et al.. Thésaurus d'agroécologie : une approche par les usages et avis d'experts. Cahiers Agricultures, 2020, 29, 10.1051/cagri/2020004 . hal-02874858

HAL Id: hal-02874858

https://hal.inrae.fr/hal-02874858

Submitted on 19 Jun 2020

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 4.0 International License

Cah. Agric. 2020, 29, 7
© V. Batifol et al., Hosted by EDP Sciences 2020
https://doi.org/10.1051/cagri/2020004

Cahiers Agricultures

Disponible en ligne :
www.cahiersagricultures.fr

Publication de données de recherche / Data Paper

Open Access

Thésaurus d'agroécologie : une approche par les usages et avis d'experts

Véronique Batifol[1,*], Sophie Aubin[2], Laurent Hazard[3], Jean-Pierre Theau[3], Cathy Bouffartigue[3] et Marie-Benoit Magrini[3]

[1] INRAE, Département-TERRA, 63122 Saint-Genès-Champanelle, France
[2] INRAE, UAR DipSO, 75338 Paris, France
[3] INRAE, UMR AGIR, 31326 Castanet-Tolosan, France

Résumé – En l'absence de thésaurus spécialisé dans le domaine de l'agroécologie, un groupe métier aux compétences complémentaires (experts scientifiques et spécialistes de l'information scientifique et technique) a construit un « thésaurus d'agroécologie ». Ce thésaurus est issu de la valorisation de l'ensemble des termes capitalisés par le dispositif de veille territoriale Agroécologie conduit à l'échelle de la région Midi-Pyrénées sur la période 2013–2017. L'ensemble des données constitutives de ce thésaurus est accessible sous Licence Ouverte et dans un format standard. Exposé sur différents portails de vocabulaires généralistes ou thématiques, ce thésaurus pourra être réutilisé à d'autres fins. Cet article décrit la méthodologie de constitution, le contenu et la structuration de ce thésaurus ainsi que son potentiel de réutilisation. L'originalité de ce travail est de reposer sur une expertise scientifique conduite à partir des termes d'usage des opérateurs du monde agricole et agro-alimentaire. La méthode proposée peut être réemployée pour construire des thésaurus sur d'autres domaines émergents.

Mots clés : agroécologie / thésaurus / data paper

Abstract – Agroecology thesaurus: An approach based on uses and expert advice. The lack for a French thesaurus dedicated to agroecology led a group of scientific experts and scientific and technical information specialists to build the "Agroecology thesaurus". Its content is based on the output of a territorial watch on Agroecology run in Midi-Pyrénées region between 2013 and 2017. The result, *i.e.* a structured list of terms, is made accessible under an open license and in a standard format. Its publication on various portals will allow its broad reuse to unforeseen purposes. In this article, we present the methodology used, the thesaurus content and structure as well as its potential for reuse. The original aspect and value of our work is that the group of scientists worked on terms actually used by various stakeholders in the agriculture and food sector. The proposed method can be reused to build thesauri on other emerging domains.

Keywords: agroecology / thesaurus / data paper

1 Introduction

Le contexte de la transition des systèmes de production vers l'agroécologie (AE) (Encadré 1) amène l'ensemble des acteurs du secteur agricole à reconcevoir leurs pratiques, à intégrer et développer de nouvelles connaissances dépassant largement les seules questions agricoles. Les termes d'usage de la profession, tout comme les concepts utilisés par la recherche, évoluent au fil de cette transition. Recenser ces nouveaux termes et s'interroger sur leur sens est un enjeu majeur pour accompagner ce processus de transition, afin de favoriser la diffusion des nouvelles connaissances et aider les acteurs à mieux comprendre les processus de changement en cours. L'usage des termes est porteur de sens pour les acteurs, et un thésaurus, en tant que recueil terminologique structuré, vise à les révéler (Roche, 2005).

*Auteur de correspondance : veronique.batifol@inrae.fr

This is an Open Access article distributed under the terms of the Creative Commons Attribution License CC-BY-NC (https://creativecommons.org/licenses/by-nc/4.0), which permits unrestricted use, distribution, and reproduction in any medium, except for commercial purposes, provided the original work is properly cited.

Encadré 1. ***: Définition d'agroécologie.***

Définition agroécologie : https://dicoagroecologie.fr/encyclopedie/agroecologie/

L'agroécologie vise à promouvoir des systèmes alimentaires viables respectueux des hommes et de leur environnement. Ces systèmes engagent des modes de productions agricoles et des filières valorisant les potentialités écologiques, économiques et sociales d'un territoire. Leur développement s'appuie sur des approches transdisciplinaires réunissant professionnels du monde agricole, scientifiques, acteurs des mouvements sociaux de l'agroécologie et des politiques publiques. [...]

L'évolution des technologies informatiques (puissance de calcul, objets connectés, e-infrastructures...) permettent d'aborder les questions scientifiques de manière de plus en plus transversale voire interdisciplinaire. C'est particulièrement le cas dans le domaine de l'agriculture où la multiplication et la diversité des sources d'information et de données rendent nécessaire la mise en place de systèmes intelligents pour y avoir accès, et être en capacité de les analyser pour en déduire de nouvelles connaissances utiles à la recherche et à l'innovation. Les politiques de recherche en faveur de la Science Ouverte telle celle décrite par l'Inra (2016) accompagnent ces évolutions en préconisant et en donnant les moyens de partager les publications scientifiques et les données dans un objectif de réutilisation. Les approches sémantiques sont un des éléments clés de cette démarche en permettant une représentation plus harmonisée et plus formelle des métadonnées et des données elles-mêmes. Le Web Sémantique (Laublet *et al.*, 2002) offre un cadre intéressant et des standards pour le développement et l'exploitation des ressources sémantiques.

Les communautés de recherche en agriculture, alimentation et environnement sont particulièrement actives dans la mise en œuvre des approches sémantiques et s'organisent pour mutualiser les ressources. En attestent les nombreuses initiatives comme AgroPortal (AgroPortal, 2019), RDA Agrisemantics (Agrisemantics, 2019), GACS (Baker *et al.*, 2016), Global Open Data for Agriculture and Nutrition (GODAN, 2019), Crop Ontology (Crop Ontology, 2019), etc. qui impliquent des acteurs académiques et industriels. Depuis deux ans, ces mêmes communautés s'attachent à mettre en œuvre les principes FAIR (Wilkinson *et al.*, 2016) pour rendre les données plus Faciles à trouver, plus Accessibles, plus Interopérables et par conséquent plus facilement Réutilisables (*i.e.* FAIR). Dans ce cadre, il est important de disposer de vocabulaires partagés qui permettent de s'accorder sur la manière de nommer les choses et de s'abstraire des barrières linguistiques.

Dans un contexte où aucune ressource terminologique francophone dédiée à l'agroécologie n'était disponible, le travail réalisé en 2018 a eu pour ambition de constituer un thésaurus spécialisé sur la thématique Agroécologie. La constitution et la structuration de ce thésaurus avaient pour objectif initial de faciliter :

- l'analyse automatisée de documents collectés dans le cadre d'une veille territoriale en Midi-Pyrénées ;
- l'identification de nouveaux termes pour alimenter ce même processus de veille.

Une fois le thésaurus constitué, il nous a paru utile de le diffuser.

Cet article de type *data paper* (Dedieu, 2014) rappelle le contexte autour des enjeux de la transition agroécologique et explique la façon dont peut se créer une sémantique dans une approche originale combinant expertise scientifique et usages. Il présente les données recensées (des termes), explique la manière dont elles ont été collectées et organisées dans le thésaurus (en fonction de leur sens et de l'application cible), puis mises à disposition pour de futures réutilisations.

2 Contexte scientifique

2.1 Transition agroécologique (TAE)

Accompagner la transition agroécologique vers de nouveaux systèmes de production plus durables et performants dans toutes leurs dimensions – économique, environnementale, et sociale – est l'un des défis auxquels répondent la recherche agronomique et l'enseignement agricole. Aujourd'hui, dans le contexte de la TAE, la recherche et l'enseignement sont amenés à repenser leurs paradigmes et postures. L'agroécologie, reconnue comme une nouvelle discipline scientifique et paradigme de production, voire de consommation (Lhoste, 2017), se construit (Francis *et al.*, 2003 ; Wezel *et al.*, 2009 ; Gliessman, 2015) et renvoie aujourd'hui à différents modèles d'agriculture (Plumecocq *et al.*, 2018). Son périmètre et les notions qui y participent tendent à s'élargir, mais sont encore peu clairs, parfois sujets à controverse, sources de malentendus. En l'absence de thésaurus spécialisé sur ce domaine d'actualité, il reste difficile de mettre en place certains outils informatiques, à commencer par une indexation efficace des documents et des données qui permettrait aux acteurs du domaine de les retrouver et de les exploiter plus facilement. Ainsi, recenser et organiser les termes associés à l'agroécologie (pour construire un thésaurus spécifique) et s'interroger sur leur sens est une forme d'accompagnement de la transition. En favorisant la diffusion des nouvelles connaissances, ce thésaurus aide les acteurs à mieux comprendre les processus de changement en cours et à construire progressivement un langage partagé.

Les données présentées dans cet article sont issues d'un travail d'expertise réalisé par des chercheurs et des spécialistes de l'information scientifique et technique de l'Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) dans le cadre du projet ATA-RI (Accompagnement de la transition agroécologique – Recherche ingénierie) au sein du programme de recherche *Pour et sur le développement régional* (PSDR4) ; (Encadré 2). Les lecteurs trouveront dans les articles mentionnés ci-après des éléments de contexte liés à l'agroécologie et en particulier au projet PSDR4 ATA-RI (Altieri, 1987 ; Hazard *et al.*, 2016).

Encadré 2. ***: Informations sur projet ATA-RI.***

Projet ATA-RI : Accompagnement de la transition agroécologique – Recherche Intervention (2016–2020)

Laboratoires de recherche : UMR AGIR (INRAE Toulouse – INP/ENSAT) et UMR LEREPS (IEP – ENFA)

Structures du développement territorial : Confédération de Roquefort, PNR Grands Causses, CA12, CA31, FRCIVAM Occitanie, AVEM, Unotec, EPL Saint Afrique, CER31, APABA, Bio82, SCOPELA, Association Bleu Blanc Cœur, Coopérative Les Bergers du Larzac, ACVA

Coordinateur : Laurent Hazard, INRAE, UMR Agir

Pour en savoir plus : https://www6.Inra.fr/psdr-midi-pyrenees/Zoom-sur/Le-projet-ATA-RI

L'idée de la construction de ce thésaurus dédié à l'agroécologie est née d'une volonté (i) de valoriser le vocabulaire capitalisé durant les deux premières années du projet de veille territoriale Agroécologie (Veille AE, 2018) conduit de 2013 à 2017 à l'échelle d'une des plus grandes régions agricoles de France, la région Midi-Pyrénées ; (ii) d'avoir une vue d'ensemble du corpus documentaire capitalisé par cette veille au moyen de technologies de fouille de textes automatisée. Ces approches nécessitent un vocabulaire spécialisé.

Cette veille, partenaire du projet de recherche ATA-RI a été mise en place pour produire des résultats pour les chercheurs, les décideurs publics et praticiens à l'échelle régionale. Elle devait permettre de :

- qualifier l'Agroécologie du point de vue des acteurs du territoire ;
- identifier des initiatives intéressantes notamment par leur mobilisation dans la formation ;
- tracer l'impact des travaux du projet ATA-RI.

Cette idée de valorisation des termes collectés qui émanait de l'envie des chargés de veille devait être consolidée par l'adhésion de scientifiques dont l'expertise par domaine de spécialité était indispensable pour valider le choix des termes ou concepts à intégrer à ce vocabulaire thématique.

3 Méthodologie

3.1 Approche globale : du dispositif de veille à la construction du thésaurus (Fig. 1)

Le dispositif de veille territoriale visait à rechercher, en utilisant des équations de recherche constituées de mots-clés (Encadré 3), à collecter des informations issues de sources internet. Ces sources variées ont été ciblées en fonction de leur pertinence et fiabilité : scientifiques, réglementaires, associatives, administratives, etc., issues de pages Web, de sites internet, de flux RSS, de blogs, de la presse générale et spécialisée. Une fois collectées, ces informations ont été sélectionnées par les chargées de veille en fonction de leur adéquation thématique liée aux termes en lien avec l'AE qu'elles contenaient. Une validation par avis d'experts a complété ce processus pour confirmer ou infirmer le choix des chargées de veille. Les informations validées ont été ensuite archivées en prévision de leur exploitation.

Encadré 3. ***: Exemple d'équation de recherche utilisée dans le projet de veille territoriale AE.***

Exemple d'équation de recherche utilisée dans le projet de veille AE pour rechercher les informations issues de sources ciblées : quelques mots clés associés aux formes d'agriculture agroécologique :

« agriculture à faible intrant » OR « agriculture économe en intrant » OR « agriculture à haute intensité écologique » OR « agriculture à haute valeur environnementale » OR « agriculture de conservation » OR « ACS » OR « agriculture de montagne » OR « agriculture durable » OR « agriculture écologiquement intensive » OR « agriculture raisonnée » OR « agri-écologie » OR « agriécologie » OR « agro-écologie » OR agroécologie OR agroécosystème OR « agro-écosystème » OR agrosystème OR agroenvironnement OR « agroforesterie » OR « agro-pastoralisme » OR « agro-sylvo-pastoralisme » OR écoagriculture OR « ecopastoralisme » OR « intensification écologique » OR « intensification durable » OR « intensification écologiquement durable » OR « pastoralisme »...

Mis en place sur la période de mai 2013 à mai 2017, le dispositif de veille (ADBS, 2008) a permis de capitaliser une première liste de plus de 300 termes qui définissent, à partir de l'observation et de l'analyse des pratiques de terrain, les différents champs liés au domaine de l'agroécologie. Ils constituent une ressource terminologique qui méritait d'être organisée et diffusée. Deux projets ont été conduits pour valoriser cette ressource brute :

- (i) L'un, objet de cet article, a rendu possible la structuration d'un thésaurus spécialisé pour valoriser l'ensemble du vocabulaire capitalisé par la veille AE. Complété par avis d'experts, ce thésaurus s'est enrichi de nouveaux termes ou concepts liés au domaine de l'agroécologie (256 concepts ajoutés). Sa structuration, détaillée ci-après, a permis d'utiliser le thésaurus pour annoter automatiquement des documents (Fig. 2). Cette approche globale et systémique du corpus enrichi de connaissances expertes a permis d'identifier des tendances et de mieux comprendre certaines interactions entre sujet et acteurs de l'agroécologie en Midi-Pyrénées. Pour permettre aux chercheurs de réaliser les analyses en toute autonomie, nous avons construit un prototype avec la plateforme Luxid® Information Analytics, outil de découverte de l'information et d'extraction de connaissance, à partir de textes (Fig. 1) ;
- (ii) L'autre projet, par une approche sériée sur des entrées terminologiques choisies par avis d'experts, a permis la construction du Dictionnaire d'Agroécologie (Batifol et Fauré, 2016). Cette ressource multimédia en langue française et anglaise propose des définitions validées par la recherche qui explicitent des termes de l'agroécologie. À visée pédagogique, le dictionnaire répond à un besoin large du citoyen, de l'agriculteur, des acteurs du développement, des enseignants, des scientifiques, ... de comprendre ce qui se cache derrière chaque terme aujourd'hui couramment utilisé par les différents acteurs associés à l'Agroécologie (et les médias qui s'en sont emparés vis-à-vis du grand public).

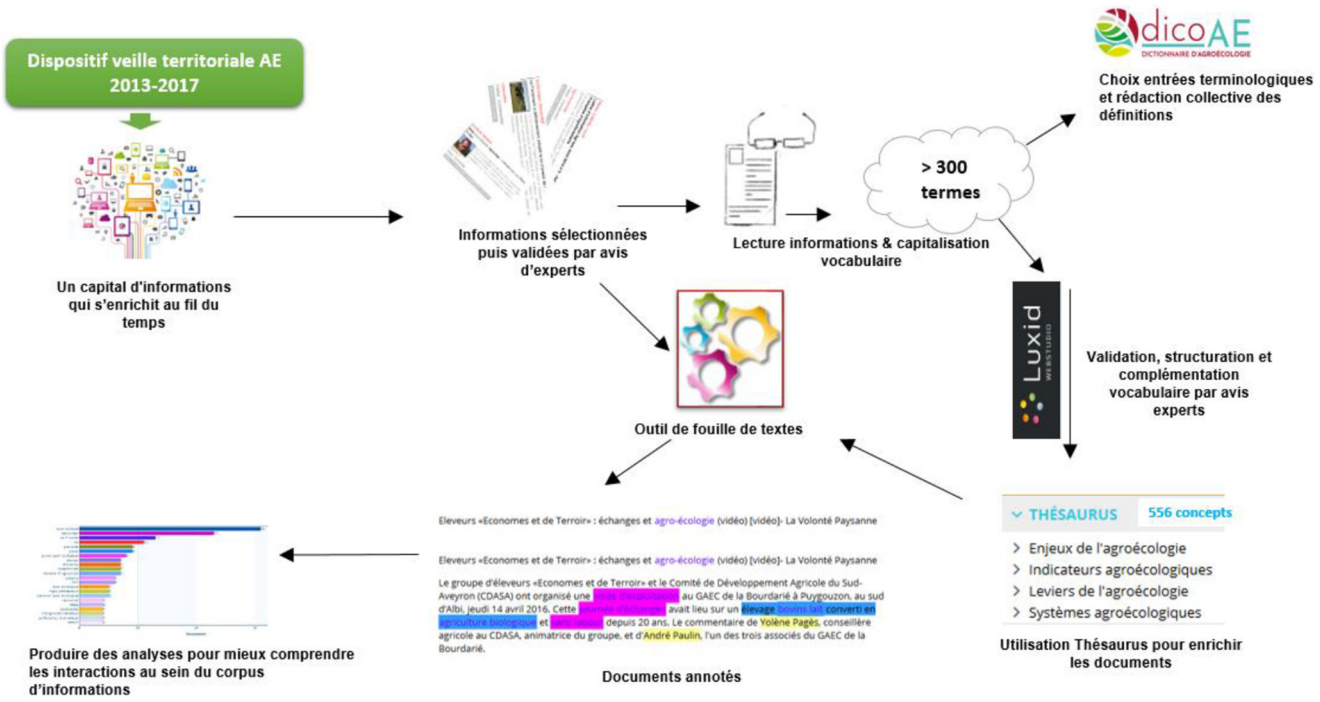

Fig. 1. Démarche d'ensemble : du projet de veille à la valorisation de la terminologie Agroécologie.
Fig. 1. *Overall approach: from the monitoring project to the valorization of Agroecology terminology.*

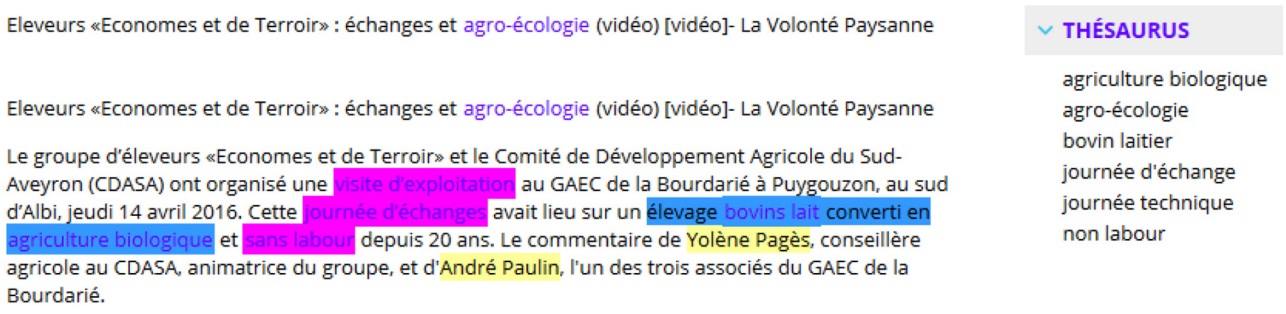

Fig. 2. Extraction de connaissances à partir des informations de la veille AE.
Fig. 2. *Extracting knowledge from AE watch.*

3.2 Construction du thésaurus

3.2.1 Capitalisation des données

Les termes présentés dans le thésaurus ont été capitalisés de quatre manières différentes et à des moments successifs, le coordinateur s'assurant de la cohérence tout au long du projet. Le projet de veille constitue la source primaire du thésaurus (termes des requêtes de recherche fournis par les experts). Dix premiers termes ont permis d'initier le projet de veille. La remontée d'informations validées au fil de l'eau a permis d'identifier à la lecture des documents de nouveaux termes ou concepts. Capitalisés puis organisés, ils ont été soumis pour validation aux experts scientifiques associés au projet.

Dans un second temps, nous avons exploité les résultats d'un outil d'extraction de termes, Luxid® Webstudio, appliqué à 1114 informations collectées lors du projet de veille. Cette étape nous a permis d'enrichir les entrées existantes avec des synonymes et d'intégrer de nouveaux concepts. Un travail collectif a permis de sélectionner, *via* l'interface de l'outil, les termes pertinents parmi environ 12 000 candidats calculés automatiquement et ordonnés selon la mesure TF-IDF (Term Frequency Inverse Document Frequency). À cette étape, une sélection de concepts du thésaurus French Crop Usage, pertinents pour notre sujet et jugés fiables, ont également été intégrés à l'ensemble. Ils concernent principalement les prairies et plantes fourragères.

Enfin, des définitions, synonymes et équivalents anglais disponibles dans les versions française et anglaise du dictionnaire d'agroécologie ont été intégrés au thésaurus.

3.2.2 Structuration sémantique

Pour construire le thésaurus, les nombreux termes capitalisés par la veille ont dû être organisés et consolidés par avis d'experts. Ce travail consistait à regrouper les termes

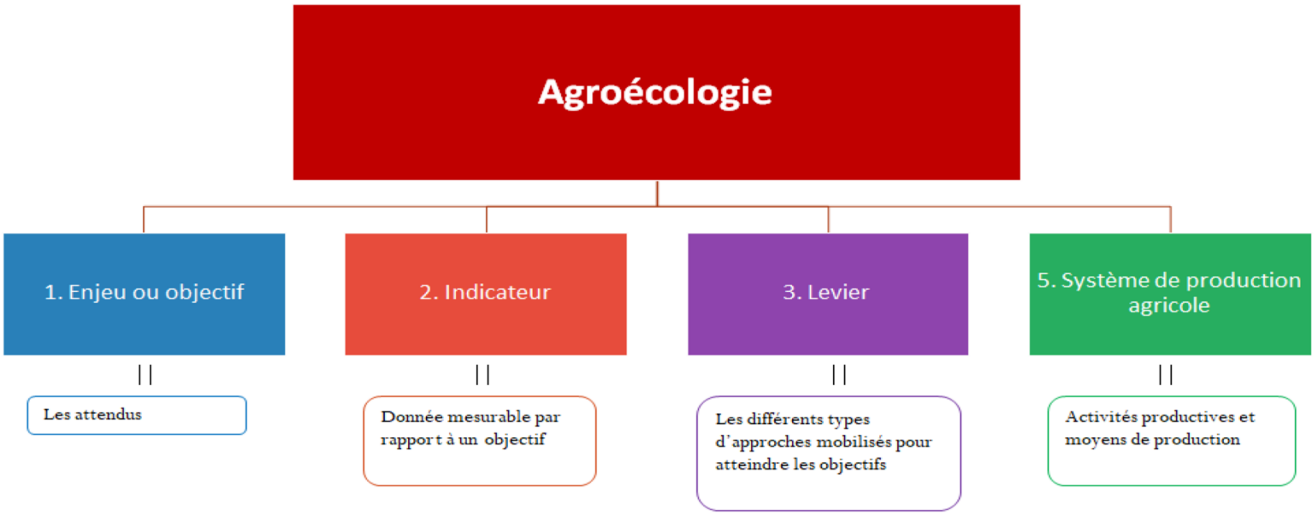

Fig. 3. Structure niveau 1 du thésaurus AE.
Fig. 3. *Level 1 structure of the AE thesaurus.*

synonymes et définir les relations hiérarchiques qu'entretiennent les concepts ainsi définis. Le résultat de ce travail est le thésaurus présenté ici.

Ce sont les questions posées dans le cadre de la veille qui ont guidé la façon de constituer et de structurer notre thésaurus (Fig. 3). En effet, les attentes relatives à l'identification des besoins des acteurs investis dans une démarche de transition agroécologique peuvent s'apparenter à des enjeux ou objectifs (leurs attendus) mesurables au travers d'indicateurs et identifiables par des leviers (leurs différentes approches) dans le contexte d'un système de production agricole donné.

À cela s'est ajouté un ensemble de questions exprimées par le porteur du projet ATA-RI qui ont guidé les analyses et nous ont confortés dans notre façon de concevoir la structure du thésaurus : par exemple, le souhait des chercheurs d'identifier les mots clés employés pour parler d'agroécologie et leur intérêt pour suivre leur évolution au fil du temps. Dans la figure 5, on peut voir par exemple les méthodes de lutte contre les bioagresseurs et maladies regroupées au sein d'une même hiérarchie. L'expertise ainsi capitalisée et représentée au sein du thésaurus est reflétée dans les résultats des outils d'analyse textuelle automatisée. *In fine*, l'annotation d'un grand nombre de documents issus de la veille AE avec le vocabulaire organisé dans le thésaurus a permis le calcul d'indicateurs pertinents en réponse au besoin du projet ATA-RI.

3.2.3 Animation du projet

Une étape clé de ce travail de constitution a été de donner envie à quelques scientifiques de s'investir sur le projet dans la durée. Le partenariat entre la veille AE et le projet de recherche ATA-RI a facilité cela. Une réunion organisée à l'initiative des chargés de veille pour présenter le projet de construction du thésaurus a permis de mobiliser quatre experts scientifiques aux domaines de compétences complémentaires (agronomie, économie, agroécologie) volontaires pour s'investir dans ce travail. Un groupe projet, auteur de cet article, a travaillé durant presque deux ans en synergie pour constituer ce thésaurus. Il a réuni des compétences diversifiées dont les complémentarités ont largement permis la réalisation de ce projet :

- gestion des connaissances pour l'identification, création, stockage et mise à disposition des connaissances ou données ;
- animation de groupes de travail pour faciliter l'activité de relecture en ouvrant des espaces communs d'échanges propices à la confrontation des points de vues et au maintien de la motivation du groupe ;
- expertises scientifiques disciplinaires indispensables à la validation et à l'organisation du vocabulaire dans le thésaurus.

Quatre réunions en visio-conférence, d'une durée de deux heures chacune, entre mai 2017 et novembre 2018 ont été suffisantes pour avancer sur la structuration et la validation du thésaurus avec les quatre experts scientifiques associés au projet. La réunion de lancement du projet a permis de valider notre proposition d'organisation du thésaurus (Fig. 3) et la manière d'envisager le travail de relecture des différents termes ou concepts dans le thésaurus. À savoir :

- la répartition du travail de relecture par binôme de relecteurs et par branches de niveau 1 choisies en fonction des capacités d'expertises disciplinaires des relecteurs (agronomie, économie, agroécologie) ;
- la prise en main de l'outil Luxid® Webstudio, avec remise d'un guide d'utilisation adapté, pour permettre à chaque expert de valider, refuser ou commenter un terme ;
- le pas de temps associé à chaque période de relecture.

Les réunions suivantes ont été organisées pour faire un point collectif sur l'avancée du travail de relecture, discuter des points de divergence (le choix des entrées terminologiques en lien avec le domaine concerné et leur positionnement dans la hiérarchie) et des difficultés rencontrées (compréhension de l'exercice,...). Chaque réunion a été préparée en amont de

Micro-thésaurus	Concepts	Termes
Enjeux	123	237
Indicateurs	53	83
Leviers	334	565
Systèmes de production	50	109

Fig. 4. Nombre de concepts et de termes les désignant pour chaque micro-thésaurus du thésaurus AE.
Fig. 4. *Number of concepts and terms designating them for each thesaurus micro thesaurus.*

façon à fournir aux experts un état des lieux précis de leur travail de relecture et des termes restants à discuter. La réunion du 18 novembre 2018 a permis de clôturer le projet, l'ensemble de vocabulaire ayant été validé après relecture intégrale par le groupe d'experts scientifiques.

4 Description du thésaurus

Les données présentées ici constituent un thésaurus spécialisé. Préalablement à la présentation des données du thésaurus, nous indiquons quelques définitions dans l'encadré 4.

Encadré 4. : Définitions.

Terme : mot ou groupe de mots servant à désigner une notion – Source : ISO 704:1987.

Concept : unité de connaissance créée par une combinaison unique de caractères – Source : norme ISO 1087.

Thésaurus : liste organisée de termes normalisés (descripteurs et non-descripteurs) servant à l'indexation des documents et des questions dans un système documentaire. Les descripteurs sont reliés par des relations sémantiques (génériques, associatives et d'équivalence) exprimées par des signes conventionnels. Les synonymes (non-descripteurs) sont reliés aux descripteurs par la seule relation d'équivalence. On peut distinguer les thésaurus en fonction du mode de regroupement des termes (thésaurus à facettes) ; de la variété linguistique des termes (mono- ou multilingue) ; des domaines de connaissances couverts (thésaurus spécialisé ou sectoriel, thésaurus encyclopédique) – Source : ADBS (Association des professionnels de l'information et de la documentation).

4.1 Contenu

Le thésaurus Agroécologie contient 556 concepts relatifs aux pratiques, aux enjeux et aux formes d'agriculture relevant de l'agroécologie (Fig. 4). La langue du thésaurus est le français. Chaque concept est désigné par un terme préférentiel (descripteur) et optionnellement un ou plusieurs synonymes. À ce jour, 134 concepts sont accompagnés d'une définition en français ou en anglais, issue du Dictionnaire d'Agroécologie et du thésaurus agricole multilingue de la FAO AGROVOC (Agrovoc, 2019). 31 concepts présentent un alignement (ou correspondance) exprimé par la propriété skos:exactMatch vers un concept du thésaurus French Crop Usage (Roussey *et al.*, 2017) et dont certains concepts ont été repris tels quels dans notre thésaurus.

Le thésaurus est structuré en 4 micro-thésaurus correspondant aux 4 dimensions présentées dans la partie Cadre conceptuel que nous avons choisies pour étudier l'agroécologie, à savoir « enjeux ou objectifs », « indicateurs », « leviers » et « systèmes de production agricole ». Au sein de chaque micro-thésaurus, les concepts peuvent être organisés au moyen de relations hiérarchiques (Fig. 5). Le niveau de profondeur maximale de la hiérarchie est 7.

4.2 Format

Le thésaurus est livré sous la forme d'un fichier au format RDF/XML structuré au moyen du modèle de représentation SKOS (Simple Knowledge Organisation System ou Système Simple d'Organisation des Connaissances), standard créé et maintenu par le consortium W3C (SKOS, 2009). SKOS est un langage de représentation de schémas de concepts, ce qui recouvre les langages documentaires tels que les thésaurus, classifications, listes de vedettes matière, taxonomies, folksonomies, etc. (Lénart, 2007).

Le fichier RDF/XML embarque également les métadonnées descriptives du thésaurus : titre, auteurs, éditeur, dates, format, type de ressource, etc. Nous avons sélectionné un ensemble d'éléments de métadonnées empruntés à des vocabulaires standards tels que Dublin Core (par ex. dc:creator), FOAF (par ex. foaf:homepage), DCAT (par ex. dcat:contactPoint), ou OMV (par ex. omv:hasFormalityLevel). Ces métadonnées sont interprétables par les applications informatiques, dont le portail de ressources sémantiques pour l'agriculture AgroPortal (AgroPortal, 2019). L'utilisation de métadonnées standards et partagées par notre communauté assure une meilleure visibilité de notre thésaurus.

Ces choix techniques, réalisés par le pôle numérique de la Direction pour la science ouverte (DipSO) INRAE, concourent à la mise en œuvre des principes FAIR qui visent à rendre les données de la recherche – dont les ressources sémantiques – plus faciles à trouver (Findable), plus accessibles (Accessible), plus interopérables (Interoperable) et plus facilement réutilisables (Reusable) (Produire des données FAIR, 2018).

4.3 Conditions d'accès aux données

4.3.1 Licence

Le thésaurus est mis à disposition sous Licence Ouverte (Licence Ouverte, 2019), équivalente à la Creative Commons Attribution (CC-BY 4.0). À condition de citer les auteurs, toute exploitation ou modification du thésaurus est donc autorisée.

4.3.2 Localisation

Afin de lui donner de la visibilité auprès de la communauté scientifique en agriculture, les données du thésaurus sont exposées et téléchargeables dans AgroPortal à cette adresse : http://agroportal.lirmm.fr/ontologies/THESAE. Ce portail thématique

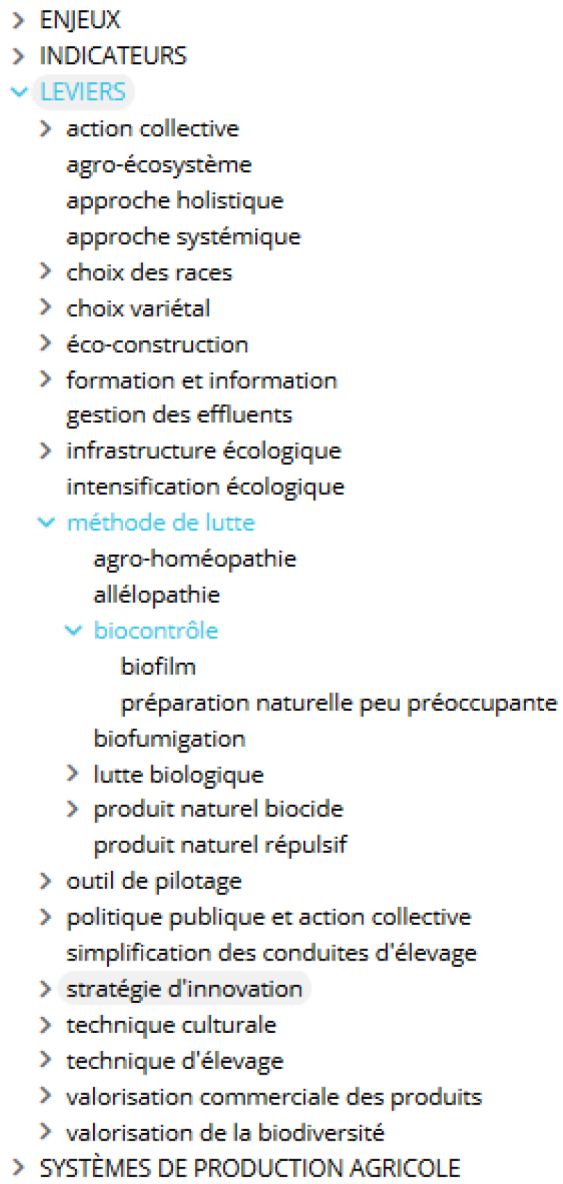

Fig. 5. Mise en évidence des quatre micro-thésaurus et des hiérarchies du thésaurus AE.
Fig. 5. *Highlighting the four micro thesauri and hierarchies of the AE thesaurus.*

permet aux visiteurs de découvrir des ontologies et des thésaurus représentés dans les standards du web sémantique (Jonquet *et al.*, 2018). Il offre des services permettant de tester ces ressources telle l'annotation de documents à la volée.

Le thésaurus est référencé dans le portail des données institutionnel Data INRAE (Data INRAE, 2019) qui permet (1) l'obtention automatique d'un Digital Object Identifier (DOI : https://doi.org/10.15454/1EKYVE) utilisable pour citer le thésaurus ; (2) sa description au moyen de métadonnées adaptées, l'espace LovInra du portail étant dédié aux ressources sémantiques ; (3) un stockage pérenne des métadonnées du thésaurus (au moins 10 ans).

Enfin, le thésaurus sera prochainement référencé dans le catalogue international et interdisciplinaire FAIRsharing (FAIRsharing, 2019) lors de la prochaine synchronisation entre Agroportal et ce catalogue. Cette exposition permettra de le faire connaître auprès d'un plus large public et lui confèrera un statut de ressource standard pour l'implémentation des principes FAIR rappelés précédemment.

5 Potentiel de réutilisation des données / précautions

Le thésaurus d'agroécologie a été construit dans un cadre et un objectif bien précis qui ont amené ses auteurs à faire des choix tant sur le contenu que sur la structuration. Il a cependant vocation à être mobilisé pour différents besoins tels que :

- l'indexation de documents, de données, et la recherche documentaire ;
- la constitution de requêtes de veille ou bibliométriques dans le domaine de l'agroécologie, domaine pour lequel peu de ressources terminologiques sont disponibles ;
- l'analyse de l'information qui mobilise des dictionnaires thématiques pour indexer le corpus puis analyser les contenus textuels ;
- le développement d'une ontologie du domaine.

L'organisation des concepts en quatre micro-thésaurus permet notamment une réutilisation plus facile, chaque partie pouvant être exploitée indépendamment des autres.

Il est important de noter que cette première version du thésaurus a vocation à évoluer avec de nouvelles expertises. Toutes les utilisations des données doivent faire mention de la source, il convient de les citer ainsi :

Batifol, Véronique ; Aubin, Sophie ; Hazard, Laurent ; Theau, Jean-Pierre ; Bouffartigue, Cathy ; Magrini, Marie-Benoit, 2019, « Thésaurus d'Agroécologie », https://doi.org/10.15454/1EKYVE, Portail Data Inra, V1

6 Conclusion

Fruit d'une collaboration étroite entre des équipes de recherche et des spécialistes de l'Information scientifique et technique, la ressource présentée dans cet article est un thésaurus spécialisé en agroécologie conçu pour annoter automatiquement des documents en français issus de sources variées collectés sur le web. S'il n'a pas de vocation universelle puisqu'il vise une application précise, nous avons toutefois souhaité en faire une ressource facilement réutilisable et adaptable pour d'autres besoins. Dans ce but, une structuration claire en quatre axes matérialisés par des micro-thésaurus a été mise en œuvre. Le thésaurus est présenté dans un format standard (SKOS), téléchargeable en ligne et sa licence (Licence Ouverte d'Etalab) permet de le réutiliser et de le modifier.

La conception et le développement d'un thésaurus font appel à des compétences variées de l'expertise scientifique à l'ingénierie des connaissances pour fournir des ressources scientifiquement fondées et réutilisables. Tout au long du travail de collecte, validation et structuration, l'animation mise en place a permis une mobilisation efficace et minimale des experts du domaine. Chacun d'eux a été mobilisé dans son domaine de compétence, là est très certainement la clé de la réussite d'un tel projet.

Financeurs du projet

Programme PSDR4, projet ATA-RI 2016–2020, financé par l'INRAE et la région Occitanie.

Le département Transformations de l'agriculture, des systèmes socio-écologiques et des systèmes alimentaires, du point de vue des acteurs et de leurs actions (TERRA, INRAE).

La Direction pour la science ouverte (DipSO, INRAE).

Remerciements. Ce projet a mobilisé, à différentes phases de travail, allant de la réflexion du projet jusqu'à sa mise en œuvre, l'expertise et le savoir-faire de collègues que nous souhaitons remercier. Sans leur intérêt, disponibilité, expertise et apports techniques, ce projet aurait eu plus de difficulté à aboutir. Parmi eux, nous remercions particulièrement : Marie-Colette Fauré, Sylvie Cocaud, Mathieu Andro, Jean-Pierre Sarthou et Olivier Thérond.

Références

ADBS. 2008. Pleins feux sur la veille : enjeux, pratiques et scénarios. *Documentaliste Sciences de l'Information* 45(4): 78.

Agrisemantics Working Group in the Research Data Alliance. [2019/09/23]. https://www.rd-alliance.org/groups/agrisemantics-wg.html.

AgroPortal. [2019/09/23]. http://agroportal.lirmm.fr/.

Agrovoc. [2019/09/23]. http://aims.fao.org/fr/agrovoc.

Altieri MA. 1987. Agroecology: The scientific basis of alternative agriculture. Westview Special Studies in Agriculture Science and Policy, 1st ed. Westview Press, 227 p. ISBN-13: 978-0813372846.

Baker T, Caracciolo C, Doroszenko A, Suominen O. 2016. GACS Core: Creation of a global agricultural concept scheme. In : Garoufallou E, Subirats Coll I, Stellato A, Greenberg J, eds. *Metadata and semantics research, 10th International Conference, MTSR 2016, November 22–25, 2016*, Göttingen (Germany), pp. 311–316. DOI: 10.1007/978-3-319-49157-8_27.

Batifol V, Fauré M-C. 2016. Dictionnaire d'Agroécologie [Vocabulaire]. https://prodinra.inra.fr/record/375065.

Crop Ontology for Agricultural Data. [2019/09/23]. http://www.cropontology.org/about.

Data INRAE. [2019/09/23]. https://data.inrae.fr/.

Dedieu L. 2014. Rédiger et publier un data paper dans une revue scientifique, en 5 points. Cirad, 7 p. http://url.cirad.fr/ist/data-paper.

FAIRsharing. [2019/09/23]. https://fairsharing.org/.

Francis C, Lieblein G, Gliessman S, Breland TA, Creamer N, Harwood R, *et al*. 2003. Agroecology: The ecology of food systems. *Journal of Sustainable Agriculture* 22(3): 99–118.

Gliessman SR. 2015. Agroecology: The ecology of sustainable food systems, 3rd ed. CRC press.

GODAN. A global partnership advocating for food security. [2019/09/23]. https://www.godan.info/.

Hazard L, Couix N, Magne M-A, Magrini M-B. 2016. Accompagnement de la transition agroécologique recherche ingénierie. Projet PSDR ATA-RI, Région LRMP, *Série Les 4 pages PSDR*. https://www.psdr-occitanie.fr/content/download/4940/47651/version/1/file/4pages_ATARI_20160616.pdf.

Inra. 2016. Charte pour le libre accès aux publications et aux données. Éditions Inra, 8 p. DOI: 10.15454/1.485854076583696E12 ; https://prodinra.inra.fr/record/365234.

Jonquet C, Toulet A, Arnaud E, Aubin S, Dzalé Yeumo E, Emonet V, *et al.* 2018. "AgroPortal: A vocabulary and ontology repository for agronomy". *Computers and Electronics in Agriculture* 144: 126–43.

Laublet P, Reynaud C, Charlet J. 2002. Sur quelques aspects du Web sémantique. *Assises du GDR I* 3: 59–78. http://lalic.paris-sorbonne.fr/PUBLICATIONS/2001-2002/laublet/I3V7.pdf.

Lénart M. 2007. SKOS, un langage de représentation de schémas de concepts. *Documentaliste-Sciences de l'Information* 44(1): 75–75. DOI: 10.3917/docsi.441.0075.

Lhoste V. 2017. La diffusion du concept d'agroécologie au sein des standards de produits alimentaires en France. Rapport de stage réalisé dans le cadre du Certificat de Césure Tutorée (CCT), 76 p. https://odr.inra.fr/intranet/carto_joomla/files/Rapport_IDAE_INRA_VLhoste.pdf.

Licence Ouverte / Open Licence. [2019/09/23]. https://www.etalab.gouv.fr/licence-ouverte-open-licence.

Plumecocq G, Debril T, Duru M, Magrini M-B, Sarthou J-P, Therond O. 2018. Caractérisation socio-économique des formes d'agriculture durable. *Economie Rurale* 1: 99–120.

Produire des données FAIR. [2018/08/09]. https://www6.inrae.fr/datapartage/Produire-des-donnees-FAIR.

Roche C. 2005. Terminologie et ontologie. *Langages* 157(1): 48–62. DOI: 10.3917/lang.157.0048.

Roussey C, Bernard S, Pinet F, Reboud X, Cellier V, Sivadon I *et al.* 2017. A methodology for the publication of agricultural alert bulletins as LOD. *Computers and Electronics in Agriculture* 142: 632–650. DOI: 10.1016/j.compag.2017.10.022.

SKOS Simple Knowledge Organization System Reference. [2009/08/18]. https://www.w3.org/TR/2009/REC-skos-reference-20090818/.

Veille territoriale Agroécologie Midi-Pyrénées. [2018/02/01]. https://www6.toulouse.inrae.fr/agir/Les-equipes/MAGELLAN/Realisations/Veille-Agroecologie.

Wezel A, Bellon S, Doré T, Francis C, Vallod D, David C. 2009. Agroecology as a science, a movement and a practice. *Agronomy for Sustainable Development* 29(4): 503–515.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, *et al.* 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. DOI: 10.1038/sdata.2016.18.

Citation de l'article : Batifol V, Aubin S, Hazard L, Theau J-P, Bouffartigue C, Magrini M-B. 2020. Thésaurus d'agroécologie : une approche par les usages et avis d'experts. *Cah. Agric.* 29: 7.