



## Data Article

# A lexicon about innovation in research and innovation projects for agrifood system transformation and sustainability transitions

Chloé Alexandre<sup>a,b</sup>, Sarah Valentin<sup>b,c,\*</sup>, Aleksandra Dolinska<sup>d</sup>,  
Renaud Guillonnet<sup>d</sup>, Thierry Helmer<sup>b,e</sup>, Priscila Henriquez<sup>d</sup>,  
Margarida Lima De Faria<sup>f</sup>, Syndhia Mathe<sup>a,g,h</sup>, Joaquin Navas<sup>d</sup>,  
Marlene Roefs<sup>i</sup>, Bernard Triomphe<sup>a,b</sup>, Mathieu Roche<sup>b,c</sup>,  
Aurélie Toillier<sup>a,b</sup>

<sup>a</sup> UMR INNOVATION, Univ Montpellier, French Agricultural Research Centre for International Development (CIRAD), INRAE, Institut Agro, Montpellier, France

<sup>b</sup> CIRAD, F-34398 Montpellier, France

<sup>c</sup> TETIS, Univ. Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Maison de la Télédétection, 500, rue J.F. Breton, 34090, Montpellier, France

<sup>d</sup> Agrinatura / DeSIRA-Lift

<sup>e</sup> CIRAD, DSI, 34398 Montpellier, France

<sup>f</sup> ISA, University of Lisbon

<sup>g</sup> CIRAD, UMR Innovation, Accra, Ghana

<sup>h</sup> Department of Agricultural Extension, School of Agriculture, College of Basic & Applied Sciences (CBAS), University of Ghana, Accra, Ghana

<sup>i</sup> Wageningen University and Research

## ARTICLE INFO

## Article history:

Received 13 January 2026

Revised 24 March 2026

Accepted 16 April 2026

Available online 21 April 2026

Dataset link: [Innovation lexicon \(Original data\)](#)

## ABSTRACT

Between 2019 and 2025, the European Commission funded 82 Research & Innovation projects in the agricultural sector across the globe with the main goal to support agrifood system transformation and sustainability transitions. To analyze the type of innovations developed and promoted by these projects, we used a text-mining approach on a corpus of various projects' documents (including project descriptions,

\* Corresponding author.

E-mail addresses: [chloe.alexandre@cirad.fr](mailto:chloe.alexandre@cirad.fr) (C. Alexandre), [sarah.valentin@cirad.fr](mailto:sarah.valentin@cirad.fr) (S. Valentin), [ola.dolinska@protonmail.com](mailto:ola.dolinska@protonmail.com) (A. Dolinska), [rguillonnet@outlook.fr](mailto:rguillonnet@outlook.fr) (R. Guillonnet), [thierry.helmer@cirad.fr](mailto:thierry.helmer@cirad.fr) (T. Helmer), [prishenriquezv@gmail.com](mailto:prishenriquezv@gmail.com) (P. Henriquez), [margaridalf@isa.ulisboa.pt](mailto:margaridalf@isa.ulisboa.pt) (M. Lima De Faria), [syndhia.mathe@cirad.fr](mailto:syndhia.mathe@cirad.fr) (S. Mathe), [jnavas@alumni.uoguelph.ca](mailto:jnavas@alumni.uoguelph.ca) (J. Navas), [marlene.roefs@wur.nl](mailto:marlene.roefs@wur.nl) (M. Roefs), [bernard.triomphe@cirad.fr](mailto:bernard.triomphe@cirad.fr) (B. Triomphe), [mathieu.roche@cirad.fr](mailto:mathieu.roche@cirad.fr) (M. Roche), [aurelie.toillier@cirad.fr](mailto:aurelie.toillier@cirad.fr) (A. Toillier).

<https://doi.org/10.1016/j.dib.2026.112792>

2352-3409/© 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

Keywords:  
Lexicon  
Semantic resources  
Text-mining  
Innovation  
Agrifood system transformation  
Sustainability transitions

project reports, descriptions of project events, publications, and websites), produced at different stages of their implementation. Since these Research & Innovation projects addressed agricultural innovation challenges from different angles (i.e. different innovation purposes, strategies, stakeholders, processes and outcomes), we decided to create a comprehensive lexicon on innovation. This lexicon covers six entry points to analyze multi-dimensional innovation: innovation actors, triggers, processes, outputs, purposes and phases. This data paper presents the lexicon and the methodology used to create it. The lexicon is composed of 717 keywords in English, divided into 44 concepts and associated with the 6 entry points presented above. It has been built iteratively, through 3 workshops with experts in agricultural innovation. The lexicon is intended to be used as a semantic layer in knowledge management tools, for instance to identify the most cited keywords and the most frequent keyword associations according to the type of documents, the localization of projects and the phase of implementation. These results contribute to building typologies of Research & Innovation projects to better understand the contribution of large project portfolios funded by the European Commission to innovation and impact.

© 2026 The Authors. Published by Elsevier Inc.  
This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>)

Specifications Table

|                          |                                                                                                                                                                                                                                                                                                                                                           |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Subject                  | Social Sciences; Computer and Information Science; Planning and Development;                                                                                                                                                                                                                                                                              |
| Specific subject area    | Lexicon in English dealing with innovation.                                                                                                                                                                                                                                                                                                               |
| Type of data             | Table (textual data)                                                                                                                                                                                                                                                                                                                                      |
| Data collection          | Data are manually acquired by combining 4 types of resources (primary source): LEAP4FNSSA lexicon (lexicon dealing with food security, previously published by [1]), terms proposed by two experts in innovation (Chloé Alexandre and Aurélie Toillier), terms proposed by Core Team members and terms proposed by a Large Language Model (i.e. ChatGPT). |
|                          | The LIFT-innovation lexicon is obtained with 3 iterative phases based on surveys and brainstorming with experts.                                                                                                                                                                                                                                          |
| Data source location     | The data are hosted on the CIRAD Dataverse. The data were built in the context of the DeSIRA-LIFT Initiative funded by the European Commission.                                                                                                                                                                                                           |
| Data accessibility       | Repository name: CIRAD Dataverse<br>Data identification number: <a href="https://doi.org/10.18167/DVN1/VMHTYL">https://doi.org/10.18167/DVN1/VMHTYL</a> Direct URL to data: <a href="https://dataverse.cirad.fr/dataset.xhtml?persistentId=doi:10.18167/DVN1/VMHTYL">https://dataverse.cirad.fr/dataset.xhtml?persistentId=doi:10.18167/DVN1/VMHTYL</a>   |
| Related research article | None.                                                                                                                                                                                                                                                                                                                                                     |

1. Value of the Data

- This dataset can be used to develop similar lexicons on innovation to analyze Research and Innovation (R&I) project portfolios and their contributions to transformative changes and sustainability transitions, or to analyze textual data dealing with innovation in similar contexts, in the fields of research for development, in the sector of agriculture and agrifood systems.
- This dataset is valuable for computer scientists to enrich thesauri and ontologies related to innovation. It can also be used for database indexing, for instance, by proposing these keywords as metadata.

- This dataset is useful to funders and their reporting activities for accountability purposes, to use the lexicon to apply meta-analysis of their project portfolios, and/or to group projects for monitoring and evaluation needs.
- This dataset is useful for Research & Innovation project managers and project designers to analyze and take stock from other R&I projects to generate transformative changes.
- This lexicon is intended to be used in an automatic knowledge extraction pipeline.

## 2. Background

Between 2019 and 2025, the DeSIRA (Development Smart Innovation through Research in Agriculture) Initiative of the European Commission (EC) funded 82 Research and Innovation (R&I) projects to support agrifood system transformation and sustainability transitions in the Global South. These projects received support from DeSIRA-LIFT (Leveraging the DeSIRA Initiative for the Transformation of Agri-Food Systems), which provides tailored services to strengthen project capacities for impact [2]. Innovation is indeed needed to address the interconnected challenges of food and nutrition security, ecosystem integrity, climate, and social justice [3]. Following the Food and Agriculture Organization of the United Nations Strategic Framework, we define innovation as a process whereby individuals, communities, or organizations bring new or existing products, services, processes, or organizational forms into social and/or economic use, enhancing effectiveness, competitiveness, resilience, or sustainability.

Given the diversity of ways in which innovation is described, a standardized, multi-dimensional analysis is required to (1) identify how projects approach innovation and (2) track potential evolutions during implementation. While text-mining tools such as Cortext [4] or TXM [5] enable co-occurrence analysis, lexical statistics, or thematic mapping, they operate directly on corpora without controlled vocabularies. By contrast, the KEOPS tool (*Knowledge ExtractOr Pipeline System*) [6] relies on a structured vocabulary to guide corpus analysis [7]. Previous studies have applied text analysis to innovation [8,9,20], but without producing a reusable lexicon. To address this gap, we are developing a specialized vocabulary.

## 3. Data Description

Several initiatives have developed lexicons and glossaries for the domains of agriculture (e.g. AGROVOC [10]) or innovation [11], but none of them are specifically dedicated to innovation in agriculture. Existing work on innovation uses definitions of innovation that do not apply to agrifood system transformation and sustainability transitions in the Global South and do not aim to understand innovation approaches in their diverse dimensions (e.g., technological, social, organizational, policy, etc.). For the analysis of textual documents generated by the DeSIRA projects, we built a lexicon, in English, based on a two-level hierarchy: axis of analysis and concepts (see Fig. 1). The higher level, hereafter referred to as “axis”, corresponds to the fundamental entry points used to analyze innovation (i.e. Innovation actors, Innovation triggers, Innovation process, Innovation purpose, Innovation outputs and Innovation phases). Axes contain sets of keywords that are grouped into thematic categories, referred to as “concepts”.

The six axes of the lexicon include:

- Innovation actors: This axis refers to individuals, groups of individuals and organizations involved in or benefiting from the innovation process;
- Innovation triggers: This axis refers to challenges that R&I projects seek to address, as well as emotions that may drive stakeholders to get involved in an innovation process. Innovation decisions are indeed not driven solely by economic or technical considerations but are also influenced by value-driven emotions and identity-related motivations. For instance, farmers' decisions to adopt or resist innovations have been shown to be closely linked to emotions tied to professional identity, such as pride in one's occupation, attachment to established practices, and the desire for autonomy [12]. Emotions have also been shown to play a mul-

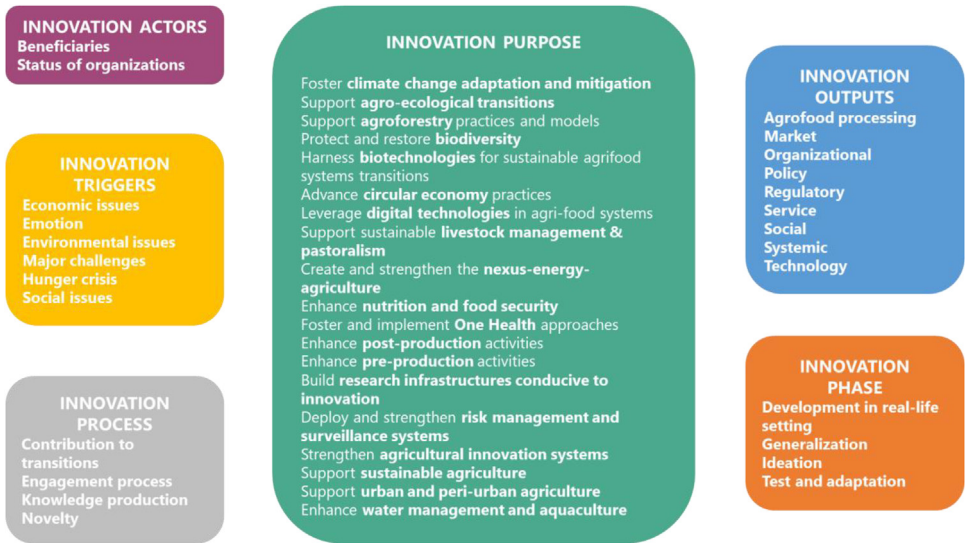


**Fig. 1.** The six Innovation axes of our lexicon on innovation for agrifood system transformation and sustainability transitions, and their corresponding concepts.

tidimensional role in innovation adoption, shaping motivation, risk perception, and engagement in learning processes [13]. In addition, new technologies may be rejected if perceived as incompatible with local norms, values, or moral frameworks, thus constituting a barrier to innovation [14].

- Innovation process: This axis includes keywords referring to modalities through which actors are engaged in the project (e.g. through farmer field schools, learning alliance, living labs, multisector platforms, etc.) and through which new knowledge is produced (e.g. by considering academic and/or non-academic knowledge, by favoring or not expert metrics, etc.)
- Innovation purpose: This axis refers to a specific area or field where innovative approaches are needed to contribute to a fair transition of agrifood systems (e.g. adaptation to climate change, agro-ecology, sustainable agriculture, livestock management, strengthening of agricultural innovation systems, etc.). These innovation purposes were identified based on the targets by the EC, which we adapted to better fit with the innovation purposes of R&I projects.
- Innovation outputs: This axis refers to the new goods, services, or products developed by the actors engaged in the innovation process. They include, among others, new agrifood processes, new markets, new organizational processes, new policy, new service, new technology, etc.
- Innovation phase: This axis refers to the stage of development of the innovation products (from ideation to test and adaptation, development in a real-life setting, and lastly, generalization).

When combined, these 6 axes allow for a multi-dimensional cross-analysis of the different ways R&I projects interpret and implement innovation in agrifood systems.

Each of the 44 concepts contains between 23 and 335 keywords, with a total number of 717 keywords (Table 1).

The dataset contains one sheet per Innovation axis of analysis. In each sheet, the concepts and corresponding keywords are provided in two columns. In a third column, we concatenated the keyword and its plural form, as well as, when provided by experts, synonyms for the keyword.

**Table 1**  
Number of concepts and keywords per axis.

| Axis                | Number of concepts | Number of keywords |
|---------------------|--------------------|--------------------|
| Innovation actors   | 2                  | 26                 |
| Innovation triggers | 6                  | 92                 |
| Innovation process  | 4                  | 88                 |
| Innovation purpose  | 19                 | 335                |
| Innovation outputs  | 9                  | 152                |
| Innovation phase    | 4                  | 23                 |

4. Experimental Design, Materials and Methods

1) Phases of construction of the lexicon

Our lexicon about innovation for agrifood system transformation and sustainability transitions in the Global South was developed in three phases (see Fig. 2). It was first developed by a Core Team of 3 researchers, specialists in analyzing and supporting agricultural innovation processes in the Global South (phase 1). The first version of the lexicon was then supplemented by a larger team of 11 experts from the DeSIRA-LIFT team, comprising researchers in different innovation-related fields (innovation evaluation, innovation studies, innovation systems, innovation policies) and practitioners with in-depth knowledge of DeSIRA projects (phases 2 and 3).

The overall process was backed up by a team in charge of developing the Knowledge Management System of the DeSIRA Initiative (the KEOPS team).

- Phase 1: first version of the lexicon  
Three researchers on agricultural innovation processes (CA, AT, SM) designed the lexicon's overall architecture. The KEOPS team (TH, MR, SV) provided guidance at this stage to ensure

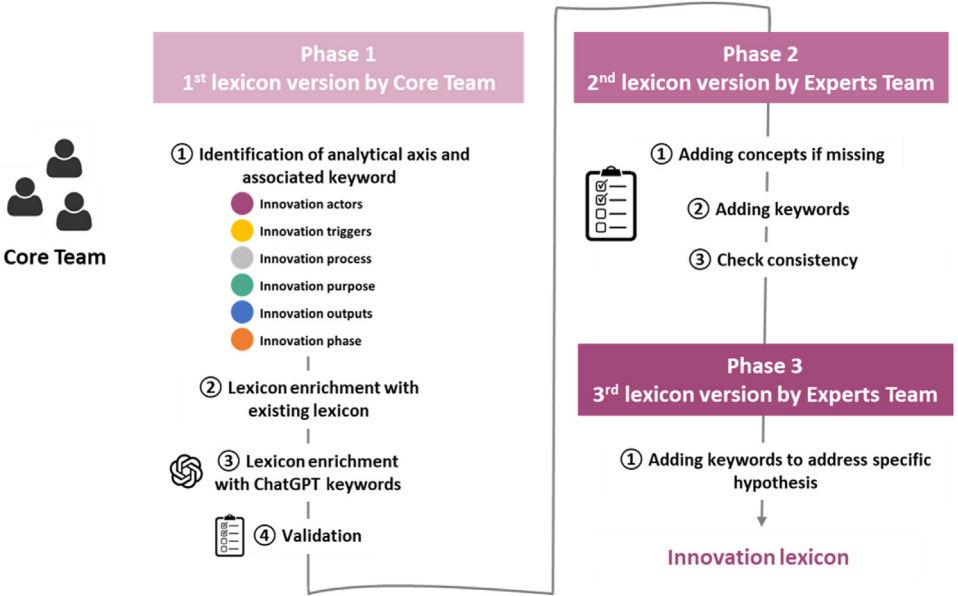


**Fig. 2.** Overview of the main steps involved in constructing the Innovation Lexicon.

that the constructed lexicon was compliant with text-mining analysis requirements and operable within knowledge-management systems. Keywords were then identified by combining 3 types of resources:

- The researchers first identified the 6 analytical axes and associated keywords, based on previous scientific work on agricultural innovation [15–17].
- Those keywords were enriched with 159 additional keywords picked out from the closest existing lexicon from another EC-funded Initiative, the LEAP4FNSSA lexicon, dedicated to food security [1].
- To complement the initial list of keywords identified by the researchers, additional keywords were generated using ChatGPT-4o mini. For each concept, the researchers asked the model to generate a list of 20 keywords. The prompt relied on a set of scientific articles previously identified in the literature review, as well as links to websites describing the DESIRA initiative and DESIRA-funded projects, which were provided as contextual input to ChatGPT. The prompt used to generate these keywords was the following (illustrated here for the concept “*Agrofood processing*”): “Based on the following set of scientific articles and websites, identify 20 keywords that best capture the main concept ‘Agrofood processing’. Focus on terms that are specific, relevant, and commonly used in academic and professional literature on this topic. Present the keywords as a simple numbered list, without explanations.”

The researchers then compared the list they had initially identified with those generated by ChatGPT. The lists overlapped for the concepts belonging to Axis 1 (Innovation actors), Axis 2 (Innovation triggers), Axis 3 (Innovation process), and Axis 6 (Innovation phase). In contrast, the keywords suggested by ChatGPT differed more substantially for Axis 4 (Innovation purpose) and Axis 5 (Innovation outputs). The verification of these keywords took place during the second phase of the lexicon construction process, that we detail below.

- Phase 2: First consolidation by the DeSIRA-LIFT experts’ team (second version of the lexicon)

This first version was reviewed individually by 11 experts from the DeSIRA-LIFT Team. The instruction(s) were to: a) check if concepts are missing; b) check if all concepts are comprehensively described by enough keywords; c) identify and mitigate potential biases associated with the use of AI for keywords gestion; and d) check consistency and/or duplication across concepts and keywords. Experts were also asked to select their “preferred” concept, like the approach adopted in AGROVOC [10].

One main bias of using AI lies in the fact that the generated keywords may reflect only the paradigms of innovation represented in the articles and websites provided as input to the model. To mitigate this bias, the experts reviewed the AI-generated keywords and enriched the lexicon by adding keywords reflecting a broader diversity of innovation paradigms, beyond dominant approaches focusing on techno-centred and top-down innovation. In particular, they introduced keywords referring to bottom-up, systemic, and inclusive approaches to innovation. These included keywords related to grassroots innovation and the inclusion of local, tacit, and traditional knowledge systems.

Overall, 147 keywords proposed by ChatGPT were added to seven concepts in the Innovation purpose axis (agroecology; agroforestry; biotechnologies; energy–agriculture nexus; One Health; risk management; urban and peri-urban agriculture), and 74 were added to six concepts in the Innovation Outputs axis (market innovation; organizational innovation; policy innovation; regulatory innovation; social innovation; systemic innovation).

Additionally, 78 keywords were added based on experts’ suggestions, including:

- 4 keywords to the axis “Innovation actors” (e.g. women groups, rural youth, etc.);
- 12 to “Innovation triggers” (e.g. soil degradation, biodiversity loss, monoculture, war, migration, etc.);
- 8 to “Innovation process” (e.g. tacit knowledge, capacity building, farmer field school, etc.)
- 28 to “Innovation output” (e.g. equality, cross-sector collaboration, ecosystem innovation, etc.)

## ROM report FAIR Sahel.docx

19 jui. 2024 - 13 sept. 2024 - ROM report FAIR Sahel.docx

MOTS-CLÉS >  
CLASSES >

Voir moins

## Project Synopsis

Description of the context with a focus on the problematic to be addressed by the action. Throughout the Sahel region, agricultural systems are characterized by high vulnerability to extreme climate variability, a high exposure of family producers to risks of various natures (climatic, economic, health, even security), a low level of livelihood diversification, and agricultural value chains that are unfavorable to producers. In these conditions, innovative practices aimed at both sustainable intensification and greater resilience of agriculture have great potential.

The main challenge is to promote a transition towards new forms of production compatible with more efficient use of resources and inputs mobilized, long-term preservation of natural resources (water, soil, biodiversity), as well as pathways intensification that is more agroecological and less impactful on health and the environment.

## Themes

Major challenges

CHALLENGE CRISIS ISSUE LACK PROBLEM

Emotions &amp; behaviours related to innovation

ACHIEVEMENT RECOGNITION

Economic issues

VULNERABILITY

Social issues

RESILIENCE

Novelty

INNOVATIVE NEW

Knowledge production

DATA INFORMATION LOCAL KNOWLEDGE MEASURE

Fig. 3. Example of KEOPS output: processed textual content (left) and lexicon extracted keywords (right).

- 18 to “Innovation purpose” (e.g. food availability, healthy diets, drying, storage, biodynamic agriculture, etc.)
- 8 to “Innovation phase” (e.g. creation, ideation, model, benchmark, etc.).

• Phase 3: Second consolidation by the team of DeSIRA-LIFT experts (third version of the lexicon)

During an online workshop, the DeSIRA-LIFT experts and the KEOPS team discussed the potential results of the text mining based on the second version of the lexicon and the research questions that could be explored thanks to the text-mining analysis with KEOPS. They formulated hypotheses to be tested with KEOPS on the nature of the innovations developed by R&I projects and the interrelationships between different concepts. 58 additional keywords relating to these hypotheses were added to the lexicon. Experts also provided feedback on concepts and keywords that needed to be refined, including “innovation triggers”, “major challenges”, and “contextual factors” that play an important role in projects’ activities and impact pathways.

## 2) Application and coverage evaluation

The lexicon is aimed to be used as an input in a knowledge management tool dedicated to textual data, KEOPS (Knowledge ExtractOr Pipeline System) [6]. Knowledge management tools aim to support knowledge discovery, which is defined as “the nontrivial extraction of implicit, previously unknown, and potentially useful information from given data” [18]. KEOPS is designed to facilitate the process of collecting, processing, analyzing and visualizing information extracted from textual data. The approach is guided by the semantic information provided through lexicons and used to analyze the texts [7,1]. The innovation lexicon was used as an input to KEOPS in addition to the documents associated with DESIRA-LIFT projects (Fig. 3). The interface allows to filter projects depending on their indexation with the lexicon (using the Keywords, Concepts or Innovation axis as levels).

To evaluate how effectively the lexicon represents the DESIRA-LIFT projects, we used KEOPS indexation to measure the frequency of each Innovation axis (identified through its associated keywords) across project documents (Fig. 4). Results show that all projects include keywords from at least two Innovation axes, and a majority (56.8%) contain keywords from all six axes defined in the lexicon. The “Innovation process” axis is the most widely represented, occurring in 100% of projects, while “Innovation triggers” is the least represented axis.

Fig. 5 illustrates the most frequent keywords, based on term frequency (TF) calculated from KEOPS-indexed project documents. Term frequency is calculated for each document by dividing the number of occurrences of a keyword by the total number of words in the document. The resulting TF values are then summed across all documents. The five most frequent terms are shown for each Innovation axis; for instance, the most frequent keywords from “Innovation phase” are *development*, *scaling*, *design*, *diffusion* and *dissemination*.

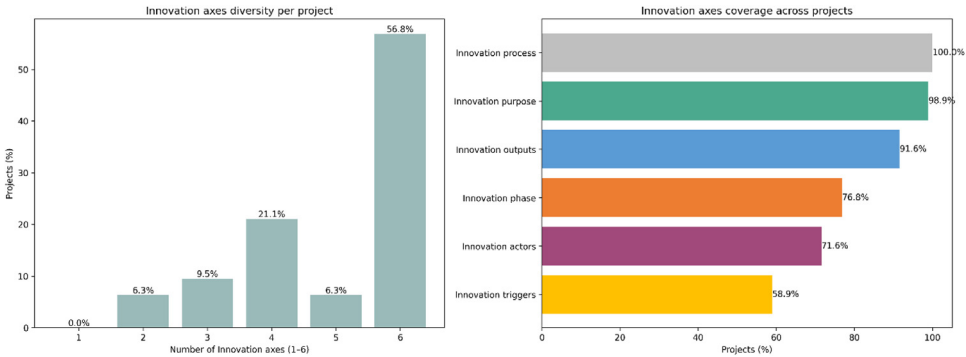


**Fig. 4.** Lexicon coverage: distribution of projects by diversity of Innovation axes (left) and frequency of projects containing keywords for each Innovation axis (right).

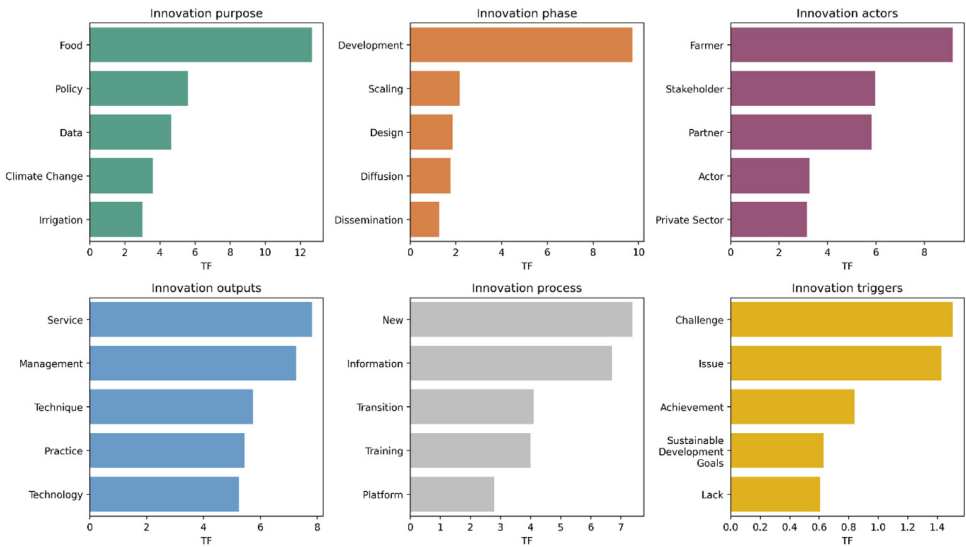


**Fig. 5.** Top five keywords per Innovation axis based on aggregated term frequency (TF).

**Limitations**

We sought to build the most comprehensive possible lexicon on innovation for agrifood system transformation to capture the various approaches by which EC-funded projects interpret and implement innovation in countries. The lexicon has been tested but has not yet been fully exploited. It is therefore not yet possible to draw a thorough critical analysis of its full potential.

However, we can suggest possible pathways for improvement. A first limitation lies in the fact that R&I project implementers were not directly involved in constructing the lexicon. Although the DeSIRA-LIFT experts have in-depth knowledge of these R&I projects, project implementers can help to enrich the keywords describing the different concepts.

Some concepts could also be enriched thanks to existing lexicons in sub-areas that we identified, such as the “emotion” concept from the axis “Innovation triggers”. Indeed, the emotional register is particularly well studied from a lexical perspective, and numerous lexicons exist to support the identification and classification of emotional expressions [21].

To facilitate analyses at the Axis or Concept level (e.g., analysis of keyword frequencies along these dimensions), we have ensured that the same keyword does not appear in two or more concepts. However, some concepts may overlap (e.g., agroecology and adaptation to climate change, or One Health and livestock management), which makes it difficult to associate a keyword solely with one or the other. Generally, it has proven more challenging to exclusively associate keywords with concepts related to cross-cutting or systemic areas of innovation, such as agroecology or adaptation to climate change.

Eventually, we will be able to adapt this lexicon according to semantic web standards (e.g. SKOS), to be integrated in open access repositories of thesauri and ontologies like AgroPortal [19].

Ethics Statement

The authors have read and follow the [ethical requirements](#) for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Credit Author Statement

| Tasks                      | Definition                                                                                                                                                                                                                              | Authors                                |
|----------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------|
| Conceptualization          | Ideas; formulation or evolution of overarching research goals and aims                                                                                                                                                                  | AT, CA, SV, MR, TH                     |
| Methodology                | Development or design of methodology; creation of models                                                                                                                                                                                | CA, AT, SM, SV, MR, TH                 |
| Software                   | Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components                                                                      | TH                                     |
| Investigation              | Literature review on text-mining approaches to define innovation; and on scientific articles characterizing innovation processes and products; identification of the lexicon overall architecture (vocabularies, concepts and keywords) | CA, AT, SV,<br>DeSIRA-LIFT Expert Team |
| Resources                  | Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools                                                                                 | AT, DeSIRA-LIFT Expert Team            |
| Writing - Original Draft   | Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation)                                                                                             | CA, SV                                 |
| Writing - Review & Editing | Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre-or postpublication stages                               | CA, SV, AT, MR                         |
| Visualization              | Preparation, creation and/or presentation of the published work, specifically visualization/ data presentation                                                                                                                          | SV                                     |
| Supervision                | Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team                                                                                                | AT, CA                                 |
| Project administration     | Management and coordination responsibility for the research activity planning and execution                                                                                                                                             | AT, CA                                 |
| Funding acquisition        | Acquisition of the financial support for the project leading to this publication                                                                                                                                                        | AT                                     |

CA: Chloé Alexandre  
AT: Aurélie Toillier  
SV: Sarah Valentin

MR: Mathieu Roche

TH: Thierry Helmer

SM: Syndhia Mathé

DeSIRA Expert Team: Aleksandra Dolinska, Renaud Guillonnet, Priscila Henriquez, Margarida Lima De Faria, Joaquin Navas, Ricardo Ramirez, Marlene Roefs, Bernard Triomphe.

## Data Availability

Innovation lexicon (Original data) (Dataverse).

## Acknowledgements

This research was funded through the DeSIRA-LIFT Initiative.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

- [1] M. Roche, A. Lindsten, T. Lundén, T. Helmer, LEAP4FNSSA lexicon: towards a new dataset of keywords dealing with food security, *Data Br.* (2022) 45108680, doi:[10.1016/j.dib.2022.108680](https://doi.org/10.1016/j.dib.2022.108680).
- [2] A. Toillier, R. Guillonnet, S. Mathé, P. Henriquez, R. Ramirez, B. Triomphe, M. Roefs, C. Lecomte, A. Dolinska, B. Douthwaite, J. Navas, M. Lima de Faria, Actionable Learning Proving and Improving R&I Impacts in Agrifood Systems Through Project-Based Approaches, CIRAD, Paris, 2025 32DeSIRA-LIFT Synthesis Brief) ISBN 978-2-87614-863-5, doi:[10.19182/agritrop/00234](https://doi.org/10.19182/agritrop/00234).
- [3] Guy Faure, Yuna Chiffolleau, Frédéric Goulet, Ludovic Temple, Jean-Marc Touzard, Innovation and Development in Agricultural and Food Systems, Éditions Quae, 2018, doi:[10.35690/978-2-7592-2960-4](https://doi.org/10.35690/978-2-7592-2960-4).
- [4] P. Breucker, J. Cointet, A. Hannud Abdo, G. Orsal, C. de Quatrebarbes, T. Duong, C. Martinez, J.P. Ospina Delgado, L.D. Medina Zuluaga , D.F. Gómez Peña, T.A. Sánchez Castaño, J. Marques da Costa, H. Laglil, L. Villard, M. Barbier. (2016). CorTexT Manager (Version v2). URL: <https://docs.cortext.net>.
- [5] S. Heiden, The TXM platform: building open-source textual analysis software compatible with the TEI encoding scheme, in: 24th Pacific Asia Conference on Language, Information and Computation, Sendai, Japon, 2010, pp. 389–398. [http://halshs.archives-ouvertes.fr/docs/00/54/97/64/PDF/paclic24\\_sheiden.pdf](http://halshs.archives-ouvertes.fr/docs/00/54/97/64/PDF/paclic24_sheiden.pdf). Retrieved from.
- [6] P. Martin, T. Helmer, J. Rabatel, M. Roche, KEOPS: knowledge ExtractOr Pipeline System, in: S. Cherfi, A. Perini, S. Nurcan (Eds.), *Research Challenges in Information Science*, Springer International Publishing, Cham, 2021, pp. 561–567, doi:[10.1007/978-3-030-75018-3\\_36](https://doi.org/10.1007/978-3-030-75018-3_36).
- [7] C. Rakotomalala, J.-M. Paillat, F. Feder, A. Avadí, L. Thuriès, M.-L. Vermeire, J.-M. Médoc, T. Wassenaar, C. Hotte-lart, L. Kieffer, E. Ndjie, M. Picart, J. Tchamgoue, A. Tulle, L. Valade, A. Boyer, M.-C. Duchamp, M. Roche, A lexicon obtained and validated by a data-driven approach for organic residues valorization in emerging and developing countries, *Front. Artif. Intell.* 8 (2025) 1557137, doi:[10.3389/frai.2025.1557137](https://doi.org/10.3389/frai.2025.1557137).
- [8] A. Gök, A. Waterworth, P. Shapira, Use of web mining in studying innovation, *Scientometrics* 102 (2015) 653–671, doi:[10.1007/s11192-014-1434-0](https://doi.org/10.1007/s11192-014-1434-0).
- [9] I. Kuzminov, P. Bakhtin, E. Khabirova, M. Kotsemir, A. Lavrynenko, Mapping the radical innovations in food industry: a text mining study, *Bioeng. eJournal* (2018) <https://api.semanticscholar.org/CorpusID:55473974>.
- [10] I. Subirats-Coll, K. Kolshus, A. Turbati, A. Stellato, E. Mietzsch, D. Martini, M. Zeng, AGROVOC: the linked data concept hub for food and agriculture, *Comput. Electron. Agric.* 196 (2022) 105965, doi:[10.1016/j.compag.2020.105965](https://doi.org/10.1016/j.compag.2020.105965).
- [11] Ultimate Innovation Glossary: All the Words. *Innovations 4 EU* (2023). <https://innovations4.eu/english/innovation-glossary/ultimate-innovation-glossary-all-the-words/>. Accessed on 01/01/2025.
- [12] A. Rieple, S. Snijders, The role of emotions in the choice to adopt, or resist, innovations by Irish dairy farmers, *J. Bus. Res.* 85 (2018) 23–31, doi:[10.1016/j.jbusres.2017.11.039](https://doi.org/10.1016/j.jbusres.2017.11.039).
- [13] C. Valor, P. Antonetti, B. Crisafulli, Emotions and consumers' adoption of innovations: an integrative review and research agenda, *Technol. Forecast. Soc. Change* 179 (2022) 121609, doi:[10.1016/j.techfore.2022.121609](https://doi.org/10.1016/j.techfore.2022.121609).
- [14] G.N. Curry, S. Nake, G. Koczberski, M. Oswald, S. Rafflegeau, J. Lummani, E. Peter, R. Nailina, Disruptive innovation in agriculture: socio-cultural factors in technology adoption in the developing world, *J. Rural. Stud.* 88 (2021) 422–431, doi:[10.1016/j.jrurstud.2021.07.022](https://doi.org/10.1016/j.jrurstud.2021.07.022).

- [15] A. Toillier, G. Faure, E. Chia, Designing and organizing support for collective innovation in agriculture, in: G. Faure, Y. Chiffolleau, F. Goulet, L. Temple, J.-M. Touzard (Eds.), *Innovation and Development in Agricultural and Food Systems* Eds., Quæ, 2018.
- [16] A. Toillier, R. Guillonnet, M. Bucciarelli, R. Hawkins, Developing capacities for agricultural innovation systems, *Agrinatura* FAO (2020), doi:[10.4060/cb1251en](https://doi.org/10.4060/cb1251en).
- [17] A. Toillier, P. Kola, The role of inter-organizational management practices in strengthening inexperienced innovation communities: exploratory study in Burkina Faso, *Innovations* 62 (2) (2020) 191–220 No.
- [18] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus, Knowledge discovery in databases: an overview, *AI. Mag.* 13 (1992) 57–70, doi:[10.1609/aimag.v13i3.1011](https://doi.org/10.1609/aimag.v13i3.1011).
- [19] C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé Yeumo, V. Emonet, J. Graybeal, M.-A. Laporte, M.A. Musen, V. Pesce, P. Larmande, AgroPortal: a vocabulary and ontology repository for agronomy, *Comput. Electron. Agric.* 144 (2018) 126–143, doi:[10.1016/j.compag.2017.10.012](https://doi.org/10.1016/j.compag.2017.10.012).
- [20] P.J.H. Daas, S. van der Doef, Detecting innovative companies via their website, *Stat. J. IAOS.* 36 (2020) 1239–1251, doi:[10.3233/SJI-200627](https://doi.org/10.3233/SJI-200627).
- [21] S.M. Mohammad, P.D. Turney, Crowdsourcing a word–EMOTION association Lexicon, *Comput. Intell.* 29 (2013) 436–465, doi:[10.1111/j.1467-8640.2012.00460.x](https://doi.org/10.1111/j.1467-8640.2012.00460.x).