INRAE DipSO

›Thésaurus INRAE

un outil central pour l'interopérabilité sémantique des systèmes d'information INRAE

Equipe technique : François-Xavier Sennesal, Sonia Bravo**, Sophie Aubin**
Comité éditorial : Agnès Girard, Anne-Sophie Martel, Christine Molé, Colette Cadiou, Emilie Bernard, Eric Cahuzac, Jeanine Martelli, Kareen Louembé, Magalie Weber, Marie-Pierre Raymond, Mohamad Taha, Olivier Dupré, Pascale Hénaut, Roselyne Tâche, Sabine Arbeille, Sonia Bravo, Sophie Aubin, Sylvie Blin, Valérie Pagneux, Véronique Decognet.

* L’interopérabilité sémantique

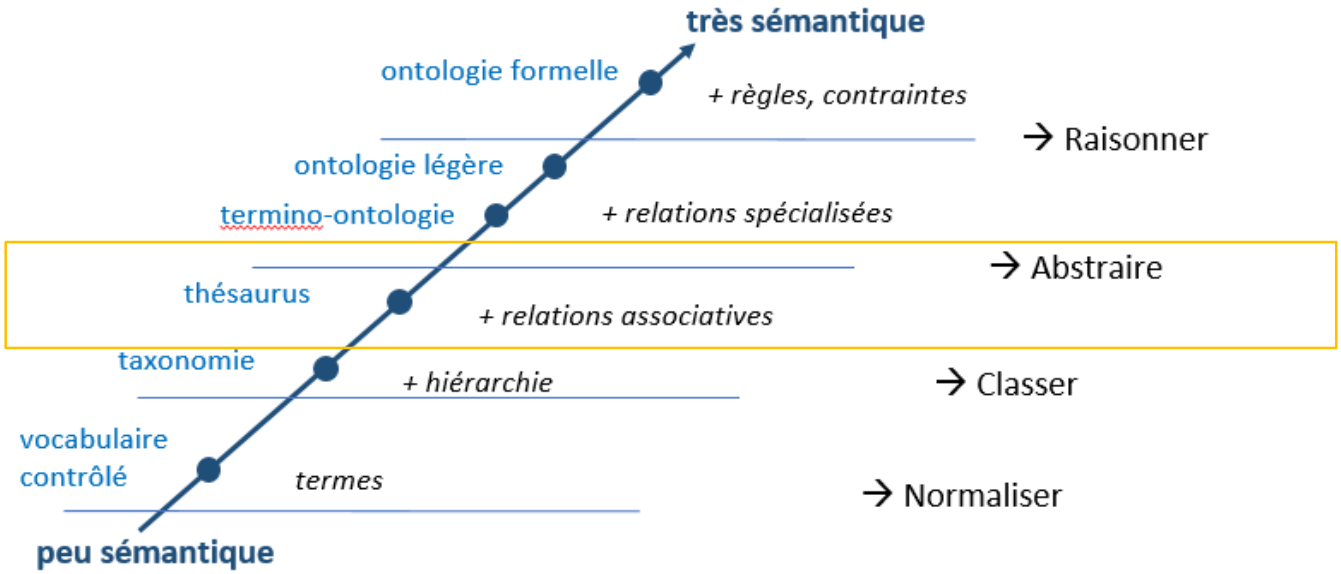
“La conduite des systèmes d’information doit être appréhendée à quatre niveaux : politique, organisationnel, sémantique et technique. L’interopérabilité, propriété essentielle des systèmes d’information, demande des réponses sur chacun de ces niveaux. ***Elle demande une stratégie sur les choix des référentiels utilisés, une organisation réactive pour assurer leur maintenance, une diffusion fiable adaptée aux nouvelles architectures des systèmes d’information****.”*

“L’interopérabilité [sémantique] permet le partage du sens par l’utilisation valide de systèmes terminologiques et l’usage combiné de données provenant de différents systèmes ou sous-systèmes.“

Source : La gouvernance de l’interopérabilité sémantique est au coeur du développement des systèmes d’information en santé - Rapport public, Marius Fieschi, 2009

* Qu'est-ce qu'un thésaurus ?

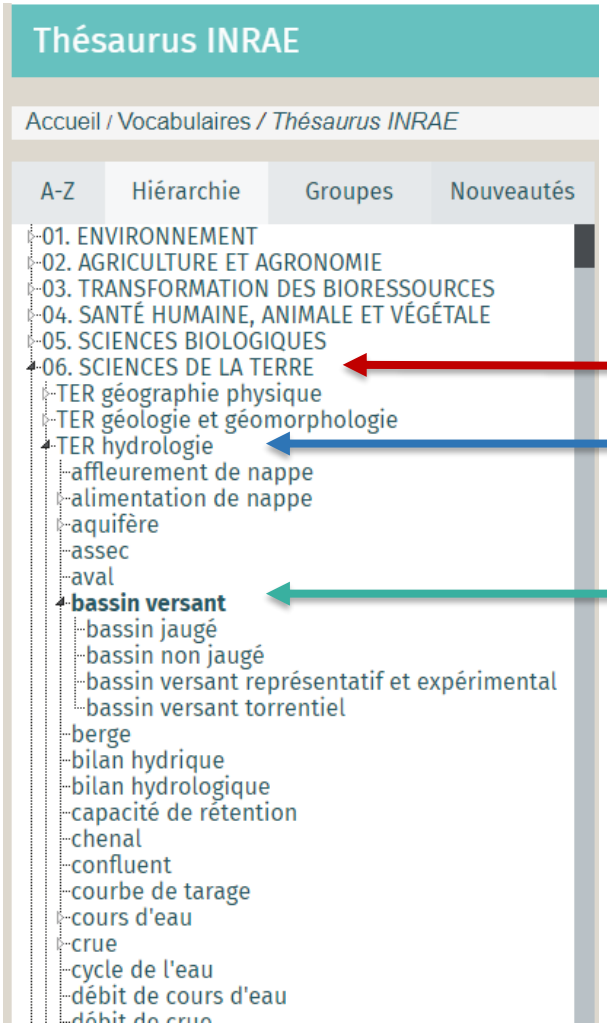
Un thésaurus est une liste structurée de concepts, représentés linguistiquement par des termes, et reliés par des relations sémantiques.



En tant que langage documentaire, un thésaurus propose un système d'indexation précis et structuré.

Thésaurus INRAE

Version 1.1 (6 sept 2021)



12 domaines

62 microthésaurus

15 475 concepts (1 concept = 1 notion)

21 724 termes FR

13 006 termes EN

+

- des définitions textuelles
- des alignements avec d'autres vocabulaires

* Le thésaurus INRAE : les relations

Le thésaurus présente deux types de relations distinctes :

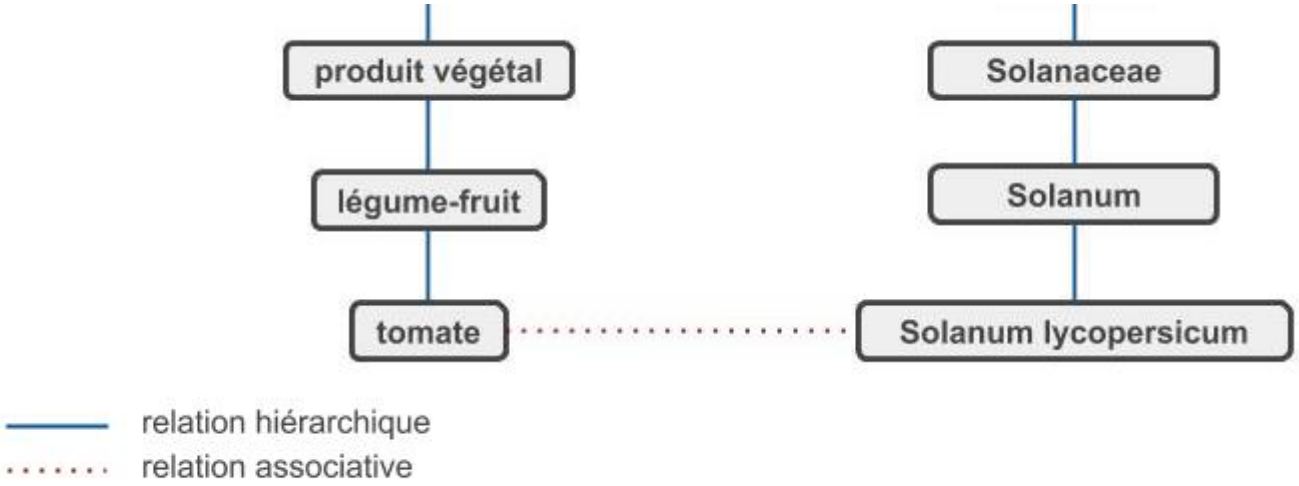
- Des **relations hiérarchiques** entre un concept générique et un spécifique qui se traduit par :

c.spécifique est un type de *c.générique*

tomate est un type *de légume-fruit*

- Des **relations associatives** entre certains concepts pouvant être liés sémantiquement:

Solanum lycopersicum est lié à *tomate*



* Pourquoi un thésaurus INRAE ?

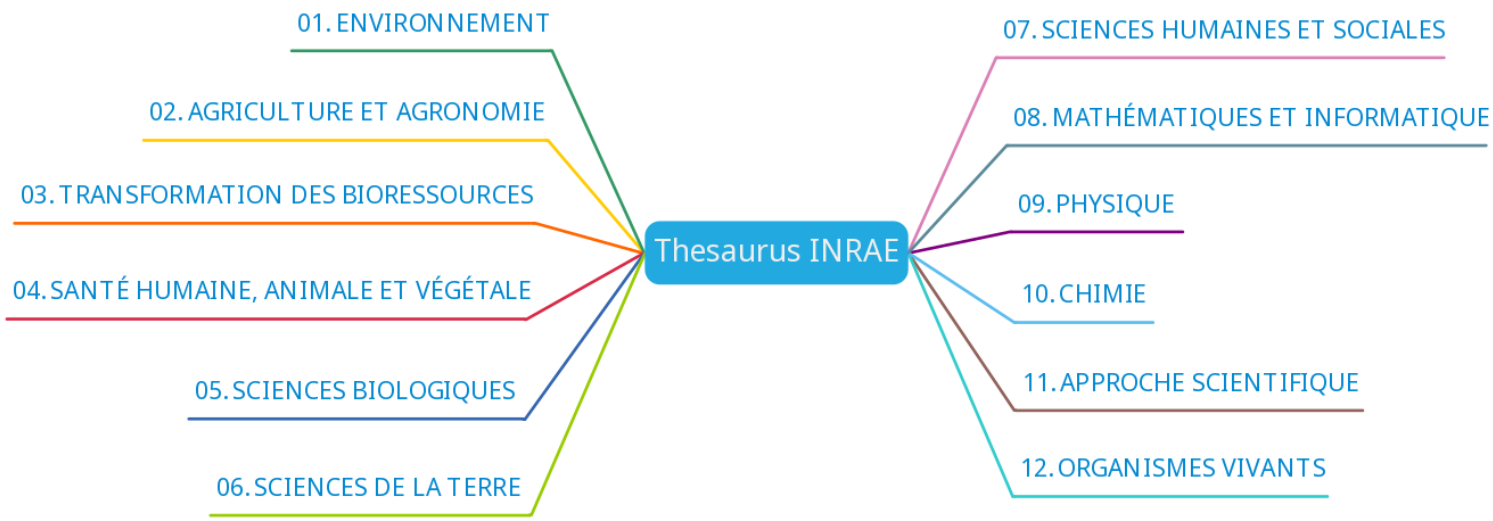
Gouvernance : un référentiel propre à l'institut qui s'adapte aux besoins internes

Qualité : base terminologique avec des définitions, des traductions en anglais et des alignements validés par un comité éditorial

Pérennité et continuité de service : un service assuré par la DipSO avec les moyens informatiques de la DSI

Complémentarité avec les ressources spécialisées (internes et externes INRAE)

Représentativité des thématiques de recherche



* Principes généraux et cadre technique

Science Ouverte, principes FAIR & Web sémantique

F	• identifiants uniques et pérennes : DOI pour la ressource et URI déréférencés pour les concepts ; • référencement sur Data INRAE, publication sur Agroportal prévue • métadonnées riches et standards

* Principes généraux et cadre technique

Science Ouverte, principes FAIR & Web sémantique

F	• identifiants uniques et pérennes : DOI pour la ressource et URI déréférencés pour les concepts ; • référencement sur Data INRAE, publication sur Agroportal prévue • métadonnées riches et standards
A	• IHM sur le web pour navigation et interrogation • API skosmos publique; web services PADRE (DSI) pour les applications internes (soon); • dépôt dans le git forge MIA

* Principes généraux et cadre technique

Science Ouverte, principes FAIR & Web sémantique

F	• identifiants uniques et pérennes : DOI pour la ressource et URI déréférencés pour les concepts ; • référencement sur Data INRAE, publication sur Agroportal prévue • métadonnées riches et standards
A	• IHM sur le web pour navigation et interrogation • API skosmos publique; web services PADRE (DSI) pour les applications internes (soon); • dépôt dans le git forge MIA
I	• représentation selon le standard SKOS (W3C) ; SKOS-XL disponible • différents formats d'export possibles (csv, rdf/xml, turtle,...) • alignements vers d'autres thésaurus et ontologies (à ce jour, GEMET et AGROVOC)

* Principes généraux et cadre technique

Science Ouverte, principes FAIR & Web sémantique

F	• identifiants uniques et pérennes : DOI pour la ressource et URI déréférencés pour les concepts ; • référencement sur Data INRAE, publication sur Agroportal prévue • métadonnées riches et standards
A	• IHM sur le web pour navigation et interrogation • API skosmos publique; web services PADRE (DSI) pour les applications internes (soon); • dépôt dans le git forge MIA
I	• représentation selon le standard SKOS (W3C) ; SKOS-XL disponible • différents formats d'export possibles (csv, rdf/xml, turtle,...) • alignements vers d'autres thésaurus et ontologies (à ce jour, GEMET et AGROVOC)
R	• licence ouverte (Etalab) / CC-BY • documentation en ligne

* Organisation & Outils

Un comité Thésaurus INRAE composé de 20 agents INRAE, se partageant la responsabilité de l'évolution des différents domaines via le rôle de référent.

Le comité peut être contacté par deux biais : une boîte mail partagée **thesaurusINRAE@inrae.fr** et le portail de consultation skosmos qui propose un formulaire de contact

<u>Outils web sémantique</u>

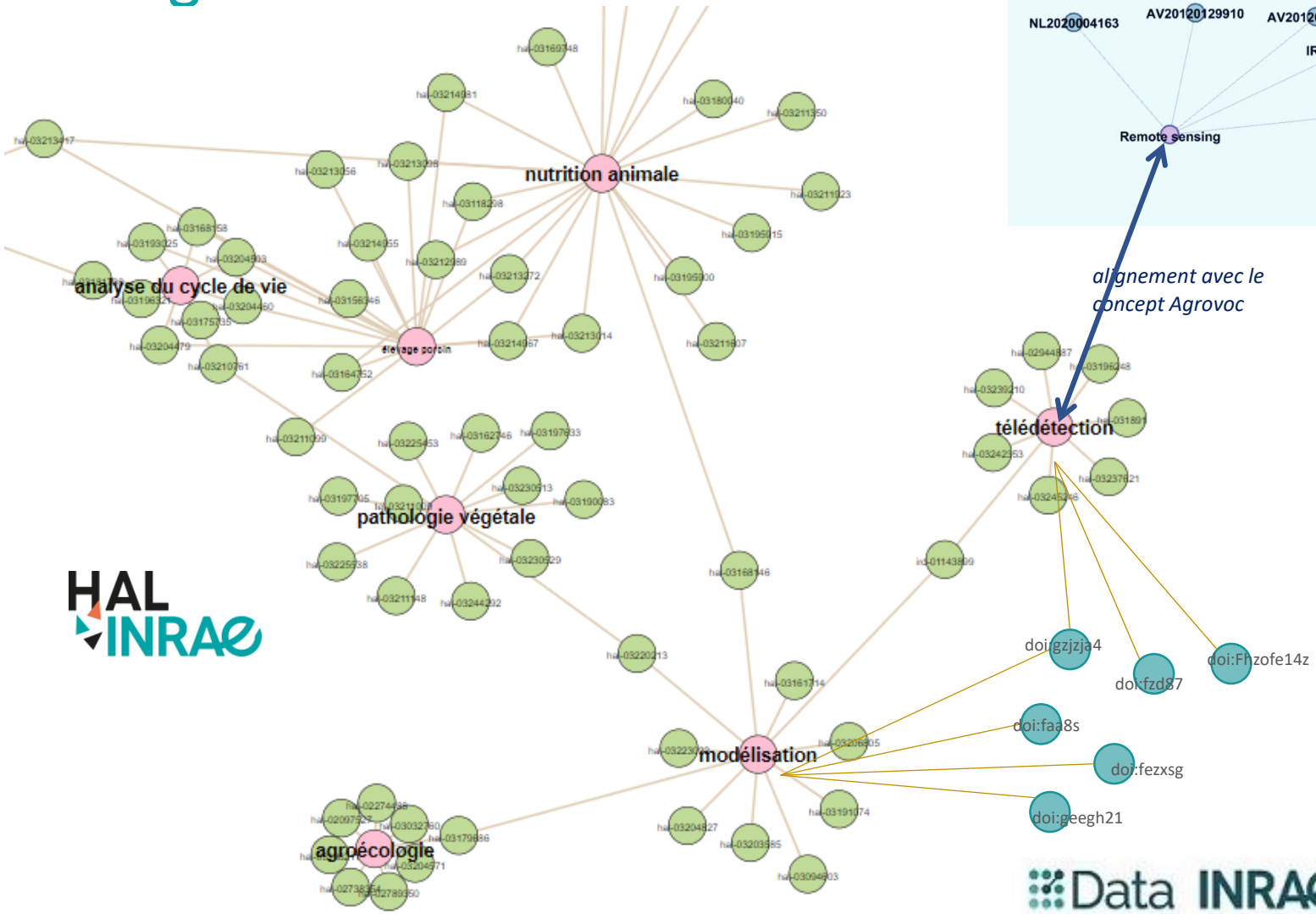
- L'administration du thésaurus est faite dans l'éditeur Vocbench 3.
- IHM SKOSMOS : https://consultation.vocabulaires-ouverts.inrae.fr/thesaurus-inrae/fr/
- Triple store DipSO, ex : https://opendata.inrae.fr/thesaurusINRAE/page/c_13730

API REST swagger : https://vocabulaires.irstea.fr/api/#/

* Usages

Base bibliographique AGRIS

alignement avec le concept Agrovoc

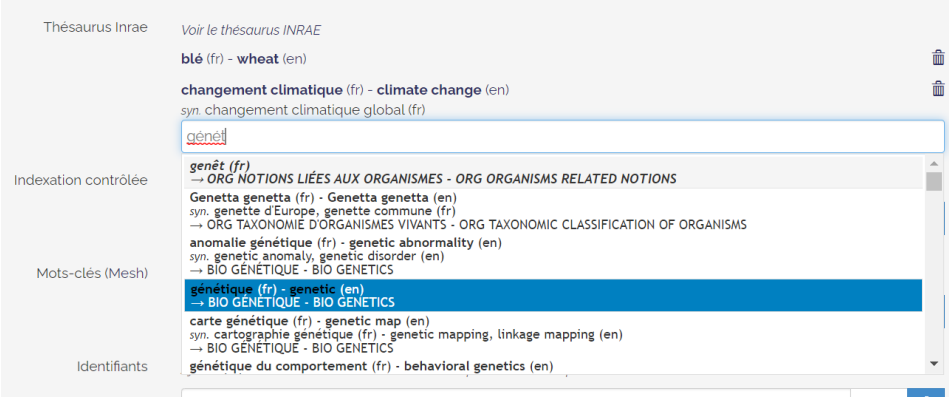


HAL INRAE

Data INRAE

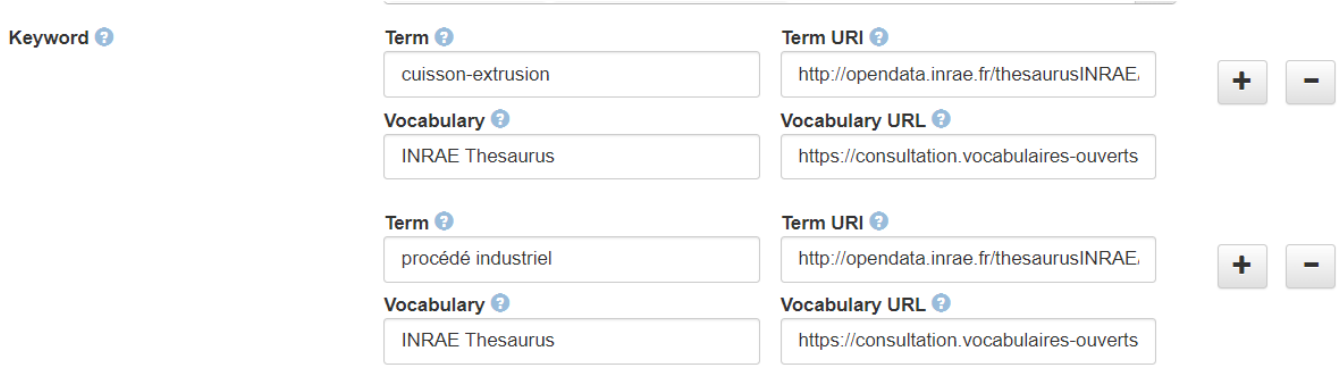
* Intégration dans les SI INRAE

- HAL INRAE : intégré depuis mars 2021 [vidéo de démonstration (1’27’’)]

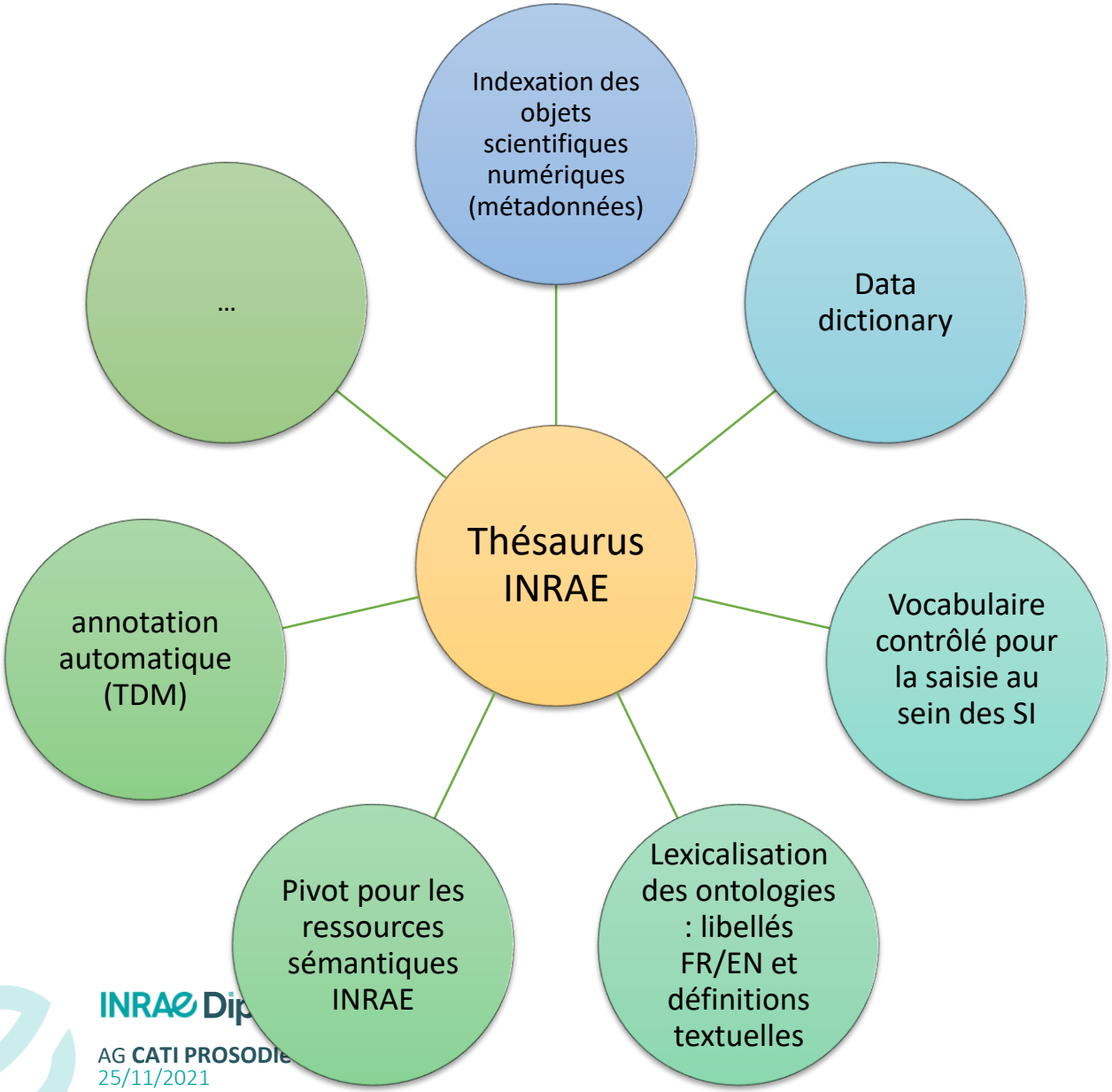


Export RDF envoyé au CCSD à chaque livraison

- Data INRAE : recommandé



* Chantiers en cours et à démarrer

- **Ajout des concepts et définition Terminologie Open au thésaurus et création d'un groupe de concepts dédiés**
- Migration des outils sur les serveurs INRAE
- Intégration à la plateforme PADRE de la DSI pour exposition du thésaurus aux SI INRAE (en cours)
- Ajout d'un champ d'indexation au référentiel « Structures » (en discussion)
- Formation
- Documentation

Quelle intégration du thésaurus INRAE aux actions de