



Gestion Sémantique des Bulletins de Santé du Végétal dans le projet Vespa

Catherine ROUSSEY, Stephan BERNARD, François Pinet
Irstea Clermont Ferrand

Xavier Reboud, Vincent Cellier
INRA de Dijon



Atelier InOvive, Montpellier, 7 juin 2016

Pour mieux
affirmer
ses missions,
le Cemagref
devient Irstea



www.irstea.fr



Plan

Contexte:

- Projet Vespa
- Bulletin de Sante du Végétal

Archivage des BSV sur le LOD: annotation et vocabulaire associé

- Annotation spatial
- Annotation temporelle
- Annotation thématique

Conclusion et perspective



Projet Vespa

Valeur et optimisation des dispositifs d'épidémio-surveillance dans une stratégie durable de protection des cultures.

- Responsable Xavier Reboud INRA
- Financement « Pour et Sur le Plan Ecophyto 2018 » (PSPE 2012)
- Étude des dispositifs d'épidémio-surveillance
 - Volet économique: évaluation du coût / bénéfice d'un réseau de surveillance
 - Volet mathématique: évaluation de la taille d'un réseau et de la fréquence d'observation
 - Volet archive: numériser les alerte agricoles et fouiller cette archive



Projet Vespa: volet Archive

Responsable Vincent Cellier

Le corpus

- Avertissements agricoles sur les périodes 1943-2009
 - Numérisation réalisé par la BNF
 - OCR et correction réalisés par la société Jouve
 - Mise en ligne Internet Archive INRA et Gallica de la BNF
- Bulletins de Santé du Végétal (BSV) 2009 jusqu'à maintenant
 - Disponible sur le web sur le site des DRAAF
 - pdf

Outil de recherche dans les résultats de fouilles du corpus:

<http://pestobserver.eu>



Archive des BSV sur le LOD : Iristea

Bulletin de Sante du Végétal

Bulletin d'alerte français

des observations sur le développement des cultures et les risques d'attaques de leurs agresseurs

BSV par Région française

BSV dédié à un type de culture (céréale, petit fruit)

BSV accessible sur le web au format pdf

Différents auteurs,

Différents style de présentation

Différents contenus

Comment stocker les annotations des bulletins avec les techno web sémantique.

Construire un jeu de données sur les observations des cultures Annotations Spatio-temporelles



Bulletin de Sante du Végétal

Filière agricole

La région

La date de parution

Le numéro

BULLETIN DE SANTE DU VEGETAL

MIDI-PYRENEES

Grandes Cultures n°28 8 juillet 2010

A retenir

MAÏS :

- Pyrale :**
 - 1ère génération : vol très étalé avec un pic enregistré entre le 13 juin et le 1er juillet selon les secteurs.
 - 2ème génération : pic de vol prévu au cours de la 1ère décennie d'août.
- Sésamie :**
 - 2ème génération : le début du vol devrait intervenir entre le 20 et le 25 juillet, avec un pic prévu entre le 26 juillet et le 03 août.

Maïs

- Stades phénologiques et état des cultures**
Les températures de ces derniers jours ont fortement accéléré le développement du maïs. Aujourd'hui, le stade moyen est situé entre 14 feuilles à sortie panicule, la floraison est proche pour la majorité des parcelles. Les derniers semis arrivent à 10 feuilles.
- Sésamie**
Le vol de première génération est terminé. L'observation de pieds de ponte se poursuit. Toutefois ces attaques ne dépassent jamais 1 à 2 % de pieds touchés dans la parcelle. Les piégeages sont peu nombreux pour l'instant.
D'après le modèle, le pic de vol de deuxième génération devrait se situer entre le 26 juillet et le 03 août.
- Évaluation du risque :**
Le risque faible en première génération se confirme. Toutefois, si le climat reste sec et chaud, les attaques de deuxième génération pourraient être significatives dans les secteurs les plus touchés en 2009.
- Pyrale**
L'étalement du vol de première génération se confirme. Les piégeages restent significatifs sur la majorité des secteurs. Le pic de vol est dépassé dans l'ensemble de la région; du 13 juin au 1er juillet selon les secteurs.
Cependant, les conditions optimales pour les pontes et leur survie se situaient autour du 15 juin, ce qui devrait donner un maximum de vol de la deuxième génération lors de la première décennie d'août.
Le début de vol devrait être enregistré au 20 juillet dans les secteurs les plus chauds et au 28 juillet dans les secteurs les plus frais.

Directeur de publication : Jean-Louis CAZAUENON
Président de la Chambre régionale d'agriculture de Midi-Pyrénées
en 2010 - 10 rue des Vignerons
31000 Toulouse - France
05 61 21 11 11 - Fax 05 61 21 11 12
Site Web : www.cra-midi-pyrenees.fr

BULLETIN DE SANTE DU VEGETAL - GRANDES CULTURES N° 28 du 08 JUILLET 2010 - Page 1/2

ARVALIS Institut du végétal

CETIOM Centre d'études techniques de l'oléiculture et de la biomasse

Freder Fédération Française des Régions Agricoles

MAIRIS Midi-Pyrénées Association pour l'Information et la Recherche en Agriculture

MAIRIS Midi-Pyrénées Association pour l'Information et la Recherche en Agriculture

Les auteurs

Annotation : Objectifs

la culture

MAÏS

• Stade phénologique et état de la culture

Après un ralentissement du développement lié à la diminution des sommes de températures, l'offre climatique est de nouveau au rendez-vous.

Les parcelles les plus avancées atteignent 16 feuilles. Le stade moyen sur la région est de 12 feuilles.

Les pluies de ces dernières semaines ont permis le suivi des dernières parcelles, notamment dans l'Aveyron. Ces parcelles présentent des stades autour de 4 feuilles.

Le stade de développement de la culture

L'agresseur

• Sésamie

La venue première génération est quasiment terminée, quelle que soit la zone.

Peu de parcelles sont touchées par des pieds de ponte. Elles se situent essentiellement sur l'ouest de la région, en particulier dans l'ouest du Gers. Le nombre de larves vivantes observées au niveau de ces pieds de ponte, est faible. Cela résulte probablement de l'humidité ambiante de ces dernières décades, peu favorable à la ponte.

Évaluation du risque : Il est trop tôt pour prévoir l'importance du deuxième vol.

Le niveau de risque

• Pyrale

A part sur le sud du Tarn et Garonne, le piégeage est en diminution significative, avec quelques individus par piège. Cela peut s'expliquer à la fois par un vol en décroissance, ainsi que par la météo de ces derniers jours peu favorable à une activité des papillons.

Dans les situations les plus froides de Midi-Pyrénées, le vol a tendance à s'étaler du fait de la diminution des températures des semaines précédentes.



Archivage des BSV

BSV de 27 régions = 19448 BSV

<u>2009</u>	<u>2010</u>	<u>2011</u>	<u>2012</u>	<u>2013</u>	<u>2014</u>	<u>2015</u>	<u>2016</u>
801	1791	2934	3174	3536	3838	3280	55

Evolution annuelle du nb de BSV pour la région centre Val de Loire

<u>2009</u>	<u>2010</u>	<u>2011</u>	<u>2012</u>	<u>2013</u>	<u>2014</u>	<u>2015</u>	<u>2016</u>
134	234	255	244	270	288	285	14

Sites web des régions
proposant des BSV

Base des BSV organisée par
région / année / rubrique

9

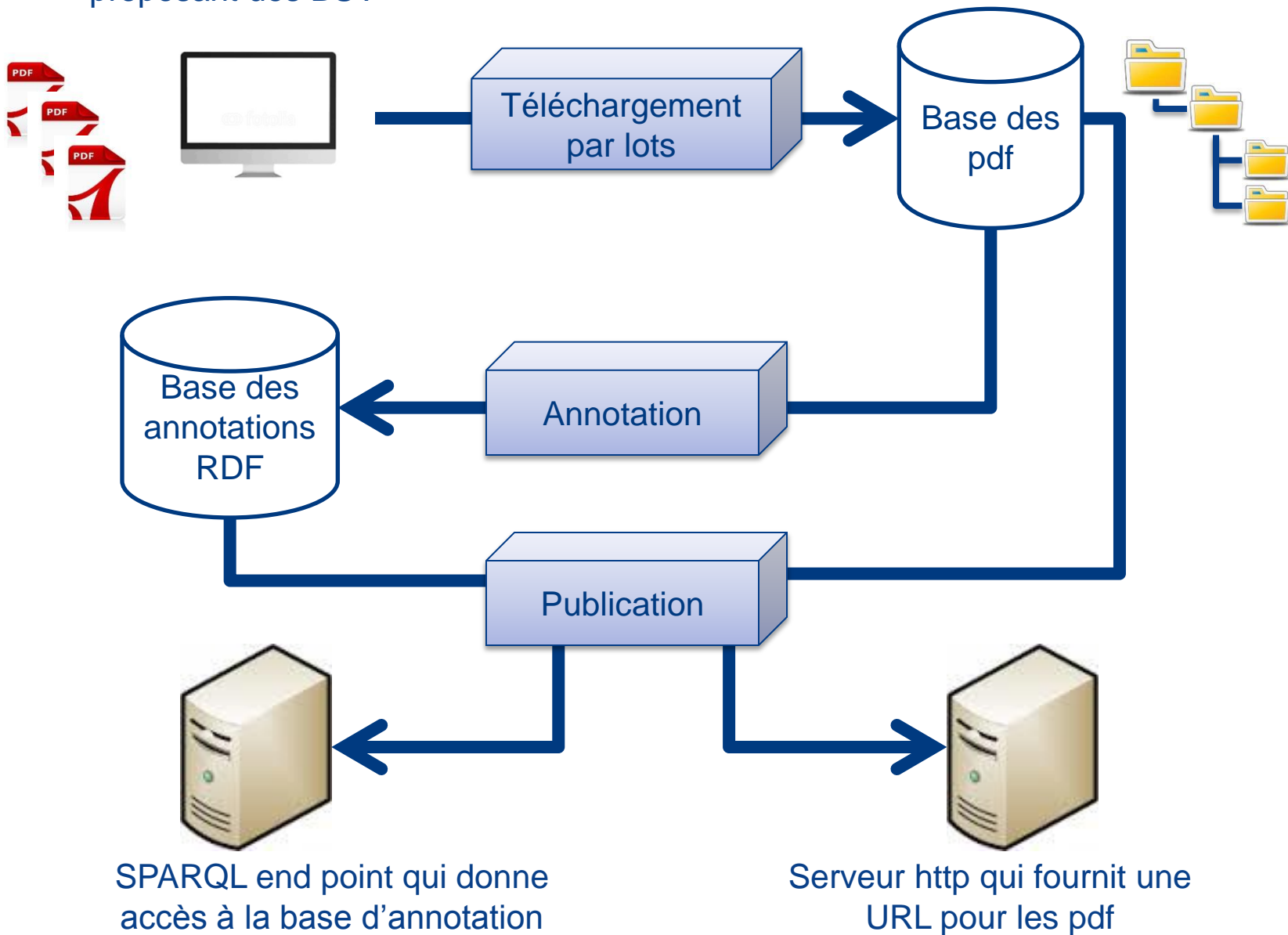




Schéma d'annotation

CONTENU EXTRAIT DES SITES WEB DES DRAAF

Annotation spatiale
les régions

Annotation thématique
les grandes catégories de culture : GC, fourrage
aucune cohérence entre les régions

Annotation temporelle
la date de publication du bulletin



Vocabulaire des régions

Jeux de données de l'IGN sans les DOM-TOM

Jeux de données de l'INSEE le plus complet

Evolution des régions due à la réforme territoriale

Chaque Région

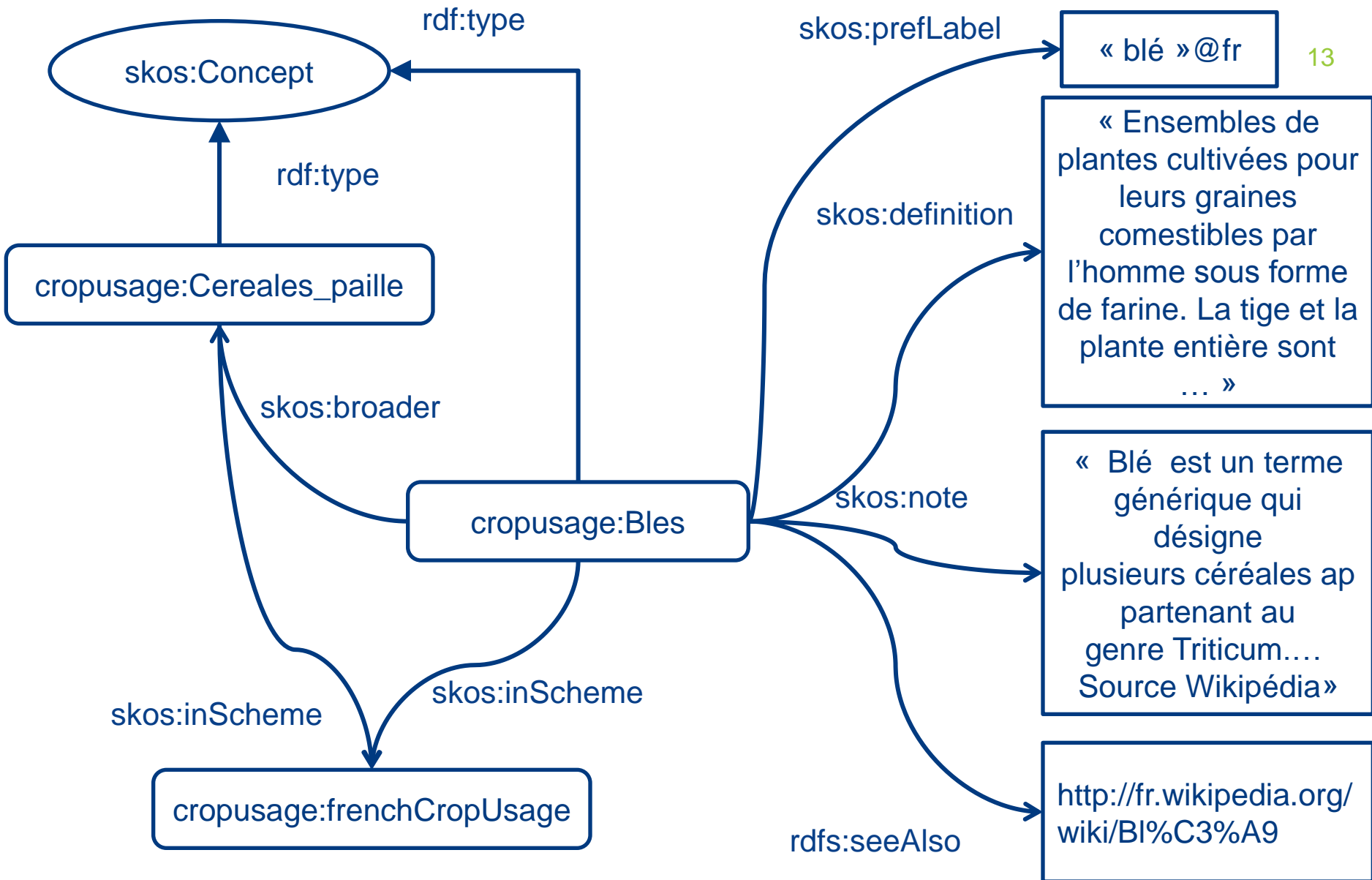
- est une instance de la classe `irstea:Region`
équivalente à `insee:Region` et `ign:Region`
- à un label `rdfs:label` pour faciliter l'interrogation
- lien d'équivalence vers `insee`, `ign`, `dbpedia`
- les nouvelles regions lien de provenance avec les anciennes régions (`prov:wasDerivedFrom`)



Thesaurus des usages des cultures en France

UNE RESSOURCE POUR CLASSER LES BSV

- Les sources utilisées: le Larousse Agricole, le registre parcellaire, la statistique agricole annuelle, Wikipedia,
- Catégorise les cultures par leur usage (alimentation humaine, animale, industrie) et les modes de cultures (GC) en FRANCE
- Une hiérarchie de rôles!
- 272 catégories hiérarchisées
- Peut évoluer en fonction des besoins (contact avec Sophie Aubin de l'Inra)
- Accessible en licence libre au format skos





Validation du thesaurus

Validation structurelle à l'aide de requêtes Sparql en utilisant corese de l'équipe Wimmics

test si chaque concept à

- un label préféré unique

- une définition qui justifie la place du concept dans la hierarchie

- une note indiquant une source utilisée

- un parent (skos:broader)

- un lien inScheme correcte

Test si à tout lien skos:broader correspond un lien skos:narrower

Validation par l'usage



Cas des plantes qui ont plusieurs usages

Concept ambigu: même nom de culture donné à plusieurs plantes qui ont des usages différents

Utilisation d'une multi-hiérarchie!

Chaque usage à un ensemble de plantes dédiées

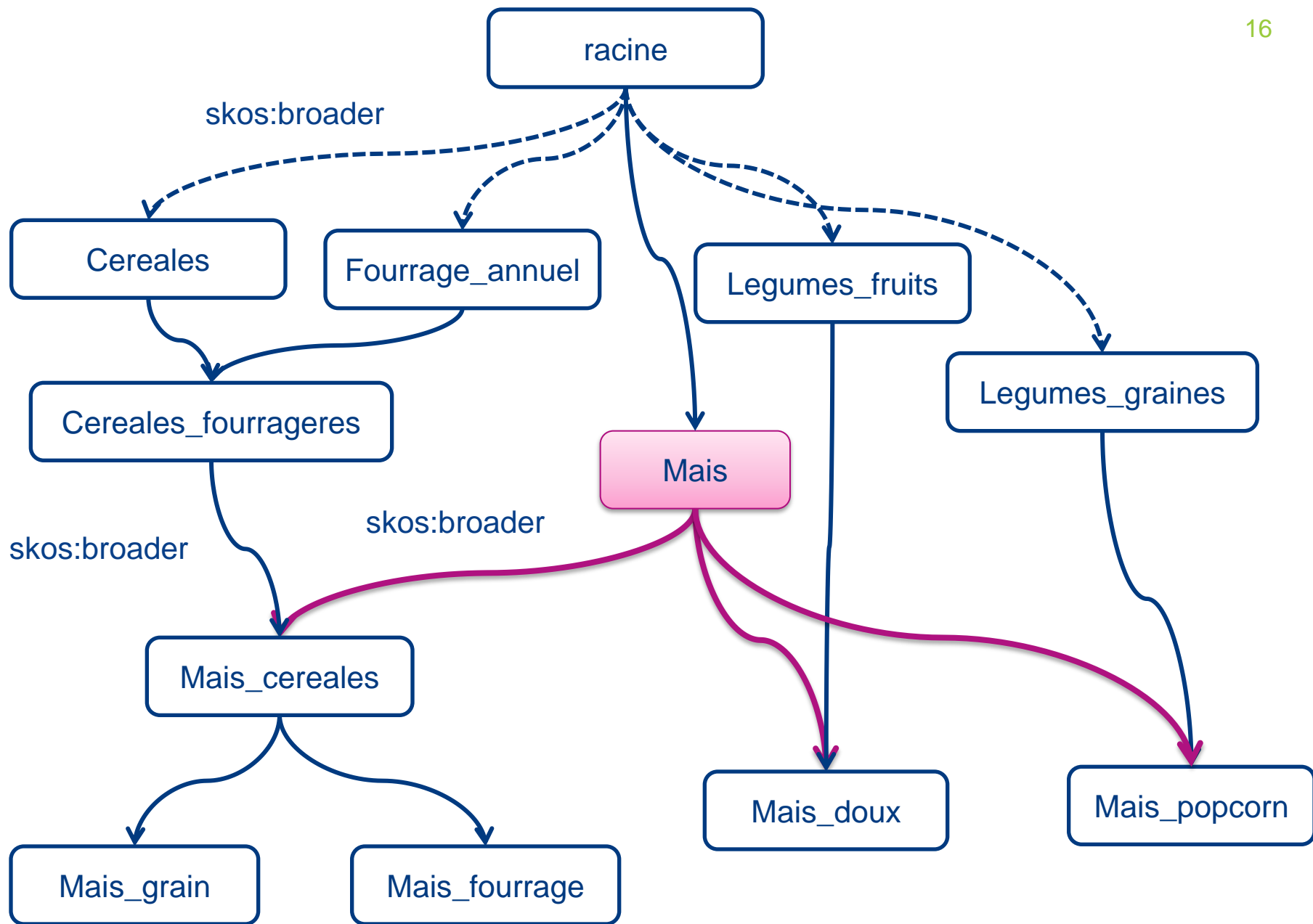
mais: `mais_cereales` (céréales fourragères), `mais_doux` (légume fruit), `mais_popcorn` (légume graine)

Le parent de `mais` est l'ancêtre le plus spécifique commun à chaque fils

Une même plante à plusieurs usages:

`Cereales_fourragere`: `Cereales` (alimentation humaine), `Fourrage_annuel` (alimentation animale)

Penser à désambiguïser et trouver le bon usage



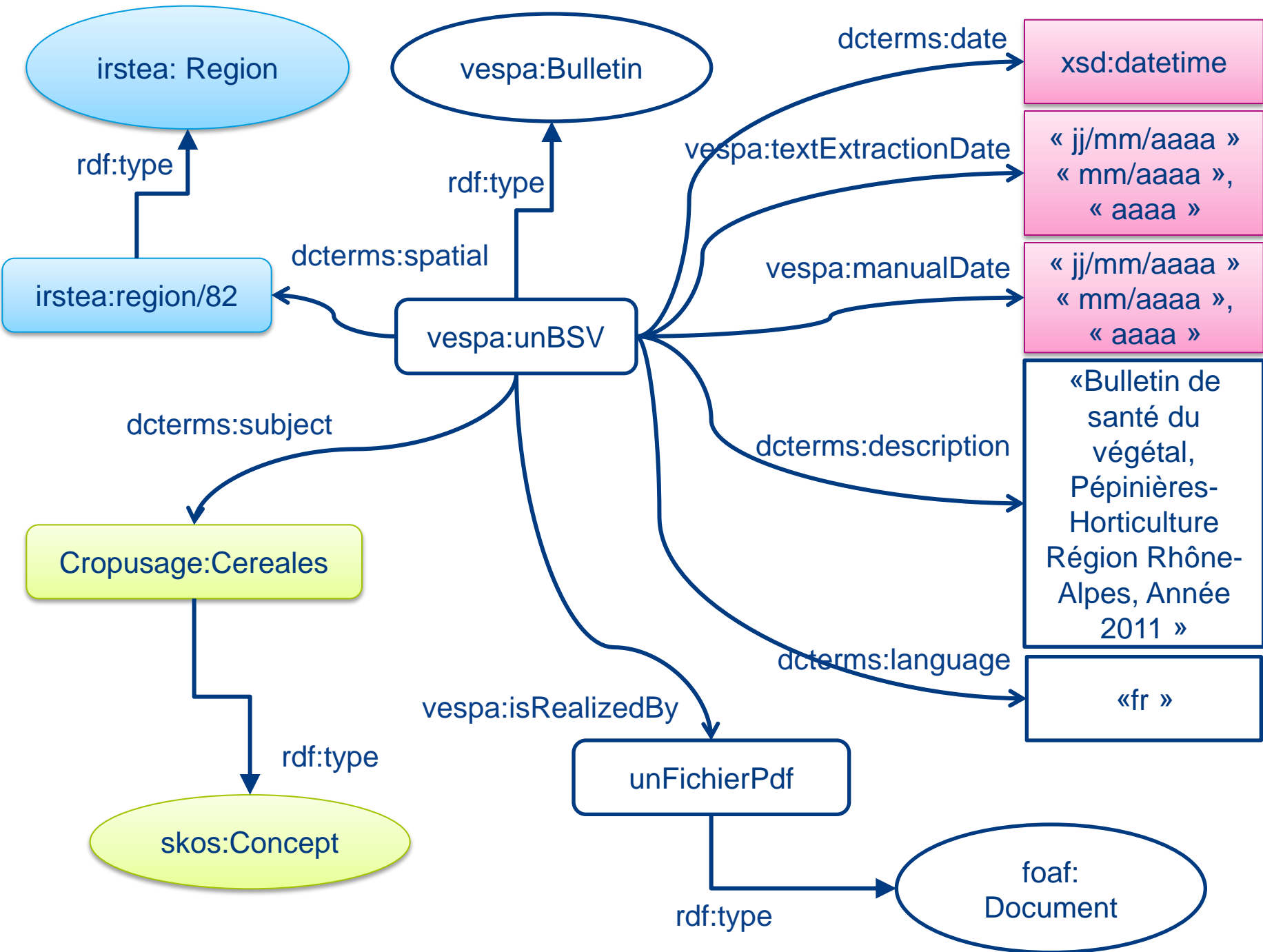
Extraction de la date de publication

Fusion des résultats des 4 processus d'extraction de date

- 1) nom des fichiers
- 2) métadonnées des pdf
- 3) processus gate: jour mois année les plus fréquentes
- 4) Pestobserver: 1^{ère} date

Méthode	Nb BSV	%	ok	P	score
Nom de fichier	263	53%	242	92%	27
Gate	454	91%	384	85%	25
Métadonnées	491	98%	353	72%	22
PestObserver	411	82%	354	86%	26
Fusion	500	100%	451	90%	

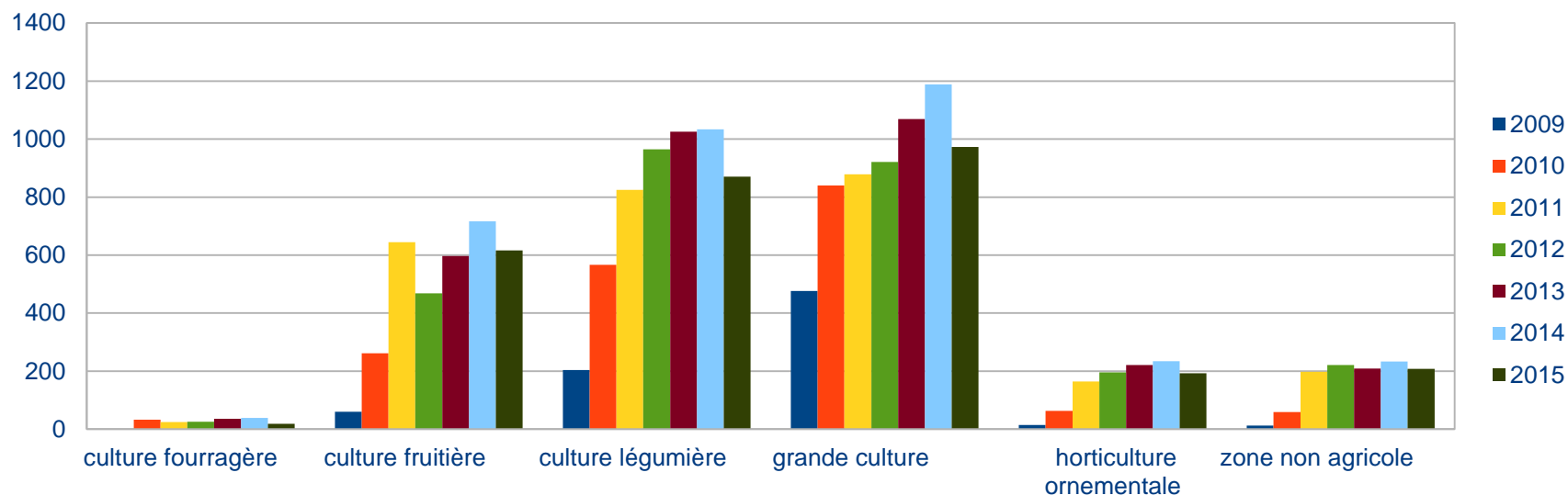
- Sauvegarde des résultats des processus et du score de la fusion
- Annotation manuelle



Repartitions des BSV

PAR ANNÉE ET PAR GRANDE CATÉGORIE DE CULTURE

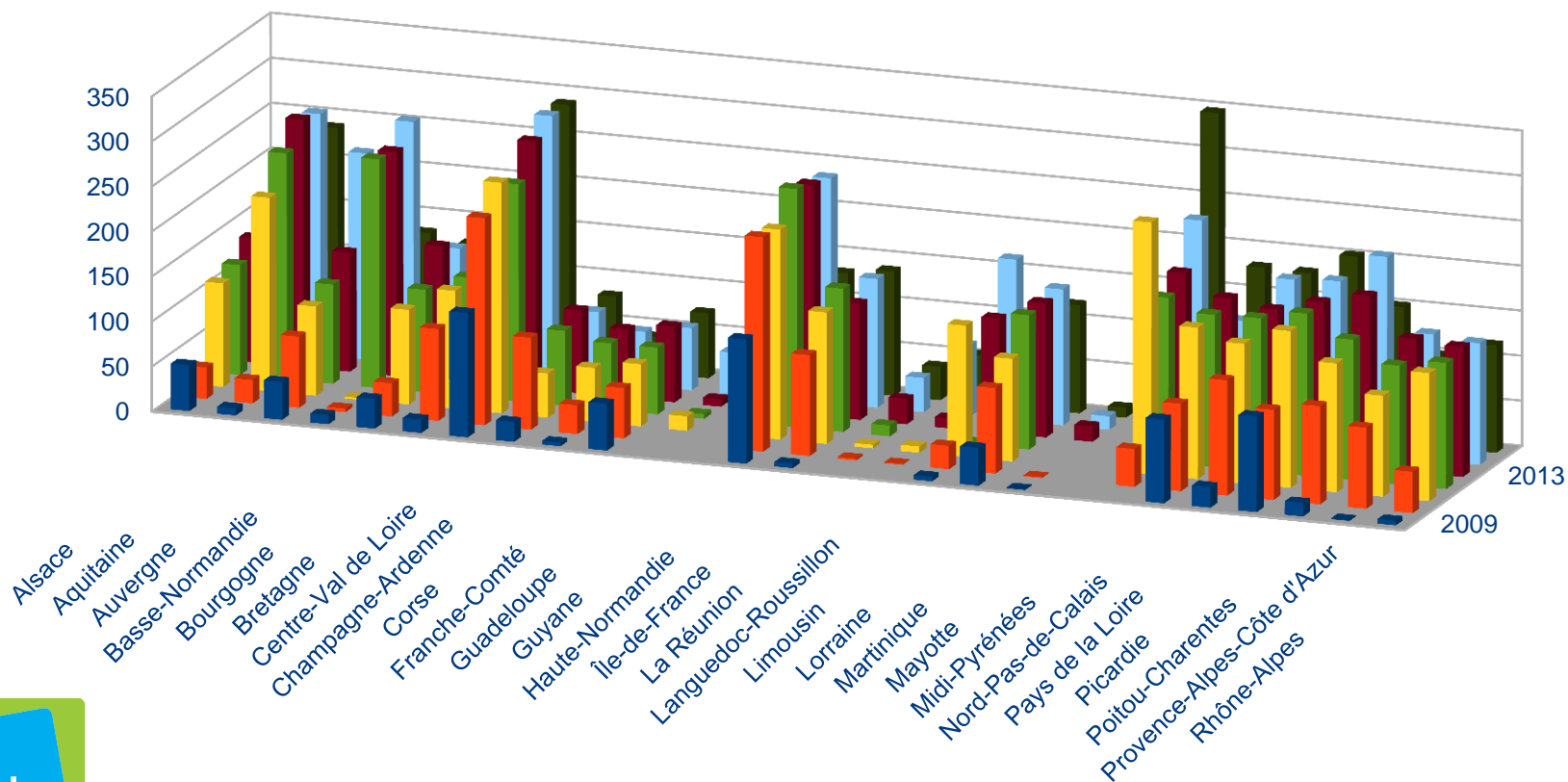
Nombre de bulletins par culture et par an



Repartitions des BSV

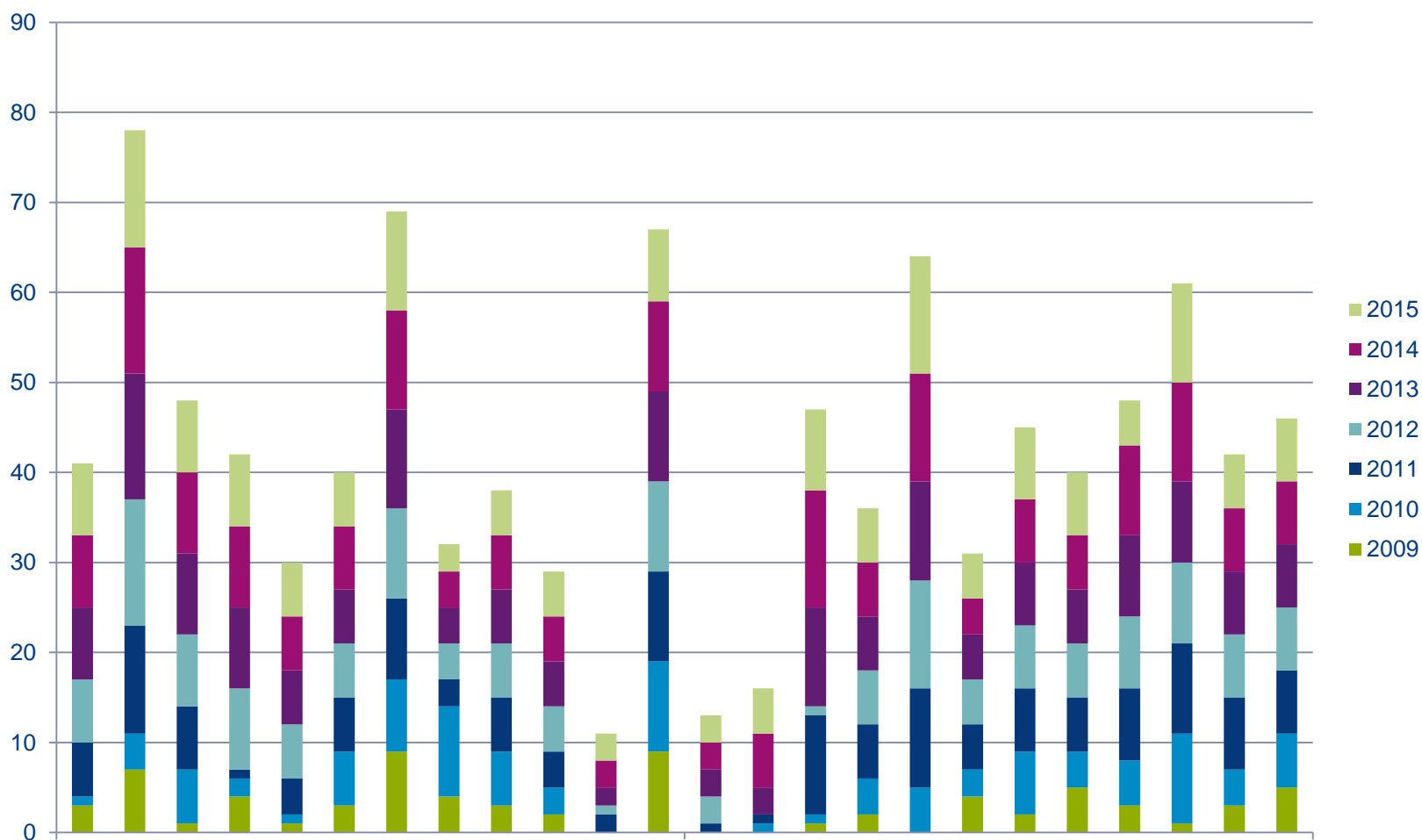
PAR ANNÉE PAR RÉGION

Nombre de BSV par région et par an



Répartition des catégories

PAR ANNÉE PAR RÉGION





Conclusion et perspectives sur l'archive

Une archive de BSV disponible sur le LOD
ontology.irstea.fr/bsv/

interface Snorql du SPARQL endpoint
ontology.irstea.fr/bsv/snorql

Un thesaurus d'usage des cultures en libre accès
bientôt disponible sur agroportal du lirmm

Extraction des cultures contenues dans le texte des BSV
test de plusieurs annotateurs
outil libre namazu annotation automatique
annotation manuelle
réutilisation de l'open annotation data model du W3C