

Instanciation des substances :

Description des fichiers sources :

1. le fichier “*substance_active_v3_Windows-1252.csv*”

Fichier source	"substance_active_v3_Windows-1252.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
Nom substance active	Bromoxynil	Chaine de caractère
Numéro CAS	1689-99-2 (bromoxynil octanoate) 1689-84-5 56634-95-8 (bromoxynil heptanoate)	liste de Chaine de caractère
Etat d'autorisation	NON_INSCRITE	Chaine de caractère
Variant	bromoxynil bromoxynil octanoate bromoxynil butyrate bromoxil (octanoate, heptanoate) bromoxynil phenol bromoxynil sel d'amine bromoxynil heptanoate bromoxynil sel de potassium	Liste de Chaine de caractère
(produit) titulaire	CHEMINOVA AGRO FRANCE SAS	Chaine de caractère
(produit) type commercial	Produit de référence	Chaine de caractère

2. le fichier “*famille_chimique.csv*”

Fichier source	"famille_chimique.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
mom_substance	Desmedipham	Chaine de caractère
famille_chimique	Carbamate	Chaine de caractère

Description des fichiers produits et script d'instanciation :

Fichier sources	"substance_active_v3_Windows-1252.csv" et "famille_chimique.csv"			
Fichiers produits	Nom du fichier	Noms de ces colonnes		La règle Celfie utilisé
	substance_active_simple.xlsx	A	familles_chimique	Individual: @B*(mm:replace(":", " ")) types: Substance annotations: rdfs:label @B* facts: etat @C* , familleChimique @A*
		B	Nom_substance_active	
		C	Etat_d_autorisation	
	substances_active_variants.xlsx	A	Nom_substance_active	Individual: @A*(mm:replace(":", " ")) facts: variant @B*
		B	Variant	
	substances_active_numero_cas.xlsx	A	Nom_substance_active	Individual: @A*(mm:replace(":", " ")) facts: numeroCas @B*
		B	Numero_CAS	

Validation de l'instanciation :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT ?p (count(?p) as ?c_p)
WHERE {
  ?s a ephy:Substance;
  ?p ?v
  filter(?v != owl:NamedIndividual)
}
GROUP BY ?p
```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observations
Nom_substance_active	1274	1477	173 substances utilisé dans les produits mais non définie
Etat_d_autorisation	1274	1274	
familles_chimique	546	546	
Variant	1703	1702	3 valeurs en double
Numero_CAS	622	621	1 valeur en double

Instanciation des produits par type :

Description des fichiers sources :

1. le fichier "produits_v3_utf8.csv"

Fichier source	"produits_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
(produit) type produit	PPP	Chaine de caractère
(produit) numero AMM	8800006	Chaine de caractère
(produit) nom produit	DIMATE BF 400	String
(produit) seconds noms commerciaux	DANADIM PROGRESS DANADIM SUPER ROGOR PLUS	Liste de Chaine de caractère
(produit) titulaire	CHEMINOVA AGRO FRANCE SAS	Chaine de caractère
(produit) type commercial	Produit de référence	Chaine de caractère
(produit) gamme usage	Professionnel	Chaine de caractère

(produit) mentions autorisees	Mention abeille Utilisable en agriculture biologique Liste biocontrôle	Liste de Chaîne de caractère
(produit) restrictions usage	Restriction néonicotinoïdes Restriction jardins amateurs	Liste de Chaîne de caractère
(produit) restrictions usage libelle	A compter du 01/09/2018, l'utilisation de ce produit contenant une substance de la famille des néonicotinoïdes est interdite pour les usages qui ne font pas l'objet d'une dérogation par arrêté ministériel, conformément à l'article L. 253-8 du Code rural et de la pêche maritime A compter du 01/01/2019, la mise sur le marché, la délivrance, l'utilisation et la détention de ce produit sont réservées aux utilisateurs professionnels, conformément à l'article L. 253-7-III du Code rural et de la pêche maritime	Liste de Chaîne de caractère
(produit) Substances actives	diméthoate (Dimethoate) 400.0 g/L diméthoate (Dimethoate) 400.0 g/L	Liste de Chaîne de caractère
(produit) fonctions	Insecticide Nématicide Fongicide	Liste de Chaîne de caractère
(produit) formulations	Huile Suspension concentrée huileuse	Liste de Chaîne de caractère
(produit) Etat d'autorisation	RETIRE	Chaîne de caractère
(produit) Date de retrait du produit	01/02/2016	Date
(produit) Date de première autorisation	01/02/1988	Date
(produit) Numéro AMM du produit de référence	2140139 2120193	Liste de Nombre
(produit) Nom du produit de référence	GLYPHYT GLISTER	Liste de Chaîne de caractère

2. le fichier "produits_phrases_de_risque_v3_utf8.csv"

Fichier source	"produits_phrases_de_risque_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
numero AMM	8800006	Chaine de caractère
nom produit	DIMATE BF 400	Chaine de caractère
Libellé court phrase de risque	EUH208	Chaine de caractère
Libellé long phrase de risque	Contient (nom de la substance sensibilisante). Peut produire une réaction allergique	Chaine de caractère

3. le fichier "produits_condition_emploi_v3_utf8.csv"

Fichier source	"produits_condition_emploi_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
type produit	PPP	Chaine de caractère
numero AMM	8800006	Chaine de caractère
nom produit	DIMATE BF 400	Chaine de caractère
catégorie de condition d'emploi	Etiquetage	Chaine de caractère
condition d'emploi libelle	Condition: - SP 1 : Ne pas polluer l'eau avec le produit ou son emballage. Ne pas nettoyer le matériel d'application près des eaux de surface. à éviter la contamination via les systèmes d'évacuation des eaux à partir des cours de ferme ou des routes.	Chaine de caractère

4. le fichier "produits_classe_et_mention_danger_v3_utf8.csv"

Fichier source	"produits_classe_et_mention_danger_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
numero AMM	8800006	Chaine de caractère
nom produit	DIMATE BF 400	Chaine de caractère
Libellé court	Sens. Resp. 1	Chaine de caractère
Libellé long	Historique - Sensibilisation cutanée, catégorie 1	Chaine de caractère

Description des fichiers produits et scripts d'instanciation :

Fichier sources	"produits_v3_utf8.csv"			
Fichiers produits	Nom du fichier	Noms de ces colonnes		La règle Celfie utilisé
	produits_partie_simple.xls	A	(produit) type produit	Individual: @B* Types: @A* Annotations: skos:notation @B* , rdfs:label @C* Facts: titulaire @D*, typeCommercial @E*, gammeUsage @F*, etat @G*, dateReraiteProduit @H* (xsd:date), datePremiereAutorisation @I* (xsd:date)
		B	(produit) numero AMM	
		C	(produit) nom produit	
		D	(produit) titulaire	
		E	(produit) type commercial	
		F	(produit) gamme usage	
		G	(produit) Etat d'autorisation	
		H	(produit) Date de retrait du produit	
		I	(produit) Date de première autorisation	

	produits_fonctions.xls	A	(produit) numero AMM	Individual: @A* facts: fonction @B*
		B	(produit) fonctions	
	produits_formulations.xls	A	(produit) numero AMM	Individual: @A* facts: typeFormulation @B*
		B	(produit) formulations	
	produits_mentions_autoriser.xls	A	(produit) numero AMM	Individual: @A* facts: mentionAutoriser @B*
		B	(produit) mentions autorisees	
	produits_restrictions.xls	A	(produit) numero AMM	Individual: @B* types: Restriction annotations: rdfs:label @B* skos:definition @C* Individual: @A* facts: restriction @B*
		B	(produit) restrictions usage	
		C	(produit) restrictions usage libelle	
	produits_seconds_noms_commerciaux.xls	A	(produit) numero AMM	Individual: @A* facts: secondNomCommercial @B*
		B	(produit) seconds noms commerciaux	
	produits__produits_references.xls	A	(produit) numero AMM	Individual: @A* facts: produitReference@C*(Class)
		B	(produit) Numéro AMM du produit de référence	
		C	(produit) Nom du produit de référence	
	produits_substances_actives.xlsx	A	_produit__Substances _actives	Individual: @E* types: Dose annotations: rdfs:label @D*(mm:prepend(@C*)) facts: doseValeur @C*(xsd:float) , unite @D* Individual: @F* types: SubstanceActive
		B	_produit__Substances	
		C	_produit__teneur_sa	
		D	_produit__unite	
		E	_produit__dose_encod	

			ed	annotations: rdfs:label @A* facts: substance @B*(Class mm:replace(":", " ") , teneurSaPure @E*(Class) Individual: @G* facts: compositionIntegrale @F*(Class)
		F	_produit_Substances _actives_encoded	
		G	_produit_numero_A MM	
Fichier source	"produits_phrases_de_risque_v3_utf8.csv"			
Fichiers produits	produits_phase_risque.xlsx	A	numero_AMM	Individual: @C*(mm:replace(":", " ")) types: PhraseRisque annotations: rdfs:label @C* , rdfs:label @B* Individual: @A* facts: phraseRisque @C*(Class mm:replace(":", " "))
		B	Libelle_court_phrase_ de_risque	
		C	Libelle_long_phrase_d e_risque	
Fichier source	"produits_classe_et_mention_danger_v3_utf8.csv"			
Fichiers produits	produits_classe_danger.x lsx	A	Libelle_court	Individual: @B*(mm:replace(":", " ")) types: ClasseDanger annotations: rdfs:label @B* , rdfs:label @A* Individual: @C* facts: classeDanger @B*(Class mm:replace(":", " "))
		B	Libelle_long	
		C	numero_AMM	
Fichier source	"produits_condition_emploi_v3_utf8.csv"			
Fichiers produits	produits_condition_empl oi.xlsx	A	numero_AMM	Individual: @C*(mm:hashEncode mm:replace(":", " ")) Types: ConditionEmploiProduit annotations: rdfs:label @C* facts: conditionEmploiCategorie @B*
		B	categorie_de_conditio n_d_emploi	
		C	condition_d_emploi_li belle	

				Individual: @A* facts: conditionEmploiProduit @C* (Class mm:hashEncode mm:replace(":", " "))
--	--	--	--	--

Validation de l'instanciation :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

```

```

SELECT ?p (count(?p) as ?c_p)
WHERE {
  ?s rdfs:type/rdfs:subClassOf ephy:Intrant.
  ?s ?p ?v
  filter(?p!=owl:NamedIndividual)
}
GROUP BY ?p

```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observation
_produit__type_produit	14093	14093	
_produit__numero_AMM	14093	14093	
_produit__nom_produit	14093	14093	
_produit__titulaire	14093	14093	
_produit__type_commercial	14092	14092	
_produit__gamme_usage	3071	3071	
_produit__Etat_d_autorisation	14093	14093	
_produit__Date_de_retrait_du	11734	11734	

_produit			
_produit__Date_de_premiere_autorisation	13768	13768	
Libelle_court_phrase_de_risque	7262	7252	10 valeur en double
Libelle_long_phrase_de_risque			
_produit__restrictions_usage	153	153	
_produit__restrictions_usage_libelle			
_produit__fonctions	8282	8260	22 valeurs en double
_produit__Nom_du_produit_de_reference	4276	4127	138 valeurs en double + 11 valeurs erroné
_produit__Numero_AMM_du_produit_de_reference			
_produit__formulations	13577	13576	1 valeur en double
_produit__seconds_noms_commerciaux	2424	2424	
melange1	9	18	
melange2	9		
categorie_de_condition_d_emploi	19914	19903	11 valeurs double
condition_d_emploi_libelle			
_produit__mentions_autorisees	2209	2209	

Instanciation des écart sur composition et revendication des MFSC et produit mixte :

Description des fichiers sources :

1. le fichier "mfsc_et_mixte_composition_v3_utf8.csv"

Fichier source	"mfsc_et_mixte_composition_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
type produit	MFSC	Chaine de caractère
numero AMM	1150022	Chaine de caractère
nom produit	ALTERNAZOTE	Chaine de caractère
Composition	Soufre exprimé en équivalent SO3 total (Min: 17.5 %, Max: 17.5 %) Oxyde de calcium (CaO) total (Min: 9.5 %, Max: 9.5 %) Matière sèche (Min: 44.0 %, Max: 44.0 %)	Liste de Chaine de caractère
Revendication	Tension superficielle à 20 °C (Min: 33.0 mN/m, Max: 33.0 mN/m) Masse volumique à 20° C (Min: 1.0015 g/cm3, Max: 1.0015 g/cm3)	Liste de Chaine de caractère
Dénomination de classe	Engrais apportant du calcium Engrais apportant du soufre	Liste de Chaine de caractère, non utilisé car c'est la même contenu que la colonne "(produit) formulations" du fichier "produits_v3_utf8.csv"

Description des fichiers produits et script d'instanciation :

Fichier sources	"mfsc_et_mixte_composition_v3_utf8.csv"			
Fichiers produits	Nom du fichier	Noms de ces colonnes		La règle Celfie utilisé
	mfsc_mixte_composition.xlsx	A	composition_complet	Individual: @B* types: Dose annotations: rdfs:label @D* facts: doseValeur @D*(xsd:float ["(^[\s]+)"] , unite @D*(["([\s]+)"])
		B	min_encoded	
		C	max_encoded	
		D	min	
		E	max	

		F	composition_complet_encoded	<p>Individual: @C*</p> <p>types: Dose</p> <p>annotations: rdfs:label @E*</p> <p>facts: doseValeur @E*(xsd:float</p> <p>["(^[\s]+)"] , unite @E*([</p> <p>([\s]+)"])</p> <p>Individual: @F*:</p> <p>Types: EcartAdmissible</p> <p>annotations: rdfs:label @A*</p> <p>facts: valeurRetenueMax @C* ,</p> <p>valeurRetenueMin</p> <p>@B* ,</p> <p>composantSimple</p> <p>@H*</p> <p>individual: @G*</p> <p>facts:</p> <p>ecartAdmissibleElementCompositon</p> <p>@F*</p>
		G	numero_AMM	
		H	Composition	
	mfsc_mixte_revendication.xlsx	A	revendication_complet	<p>Individual: @E*</p> <p>types: Dose</p> <p>annotations: rdfs:label @G*</p> <p>facts: doseValeur @G*(xsd:float</p> <p>["(^[\s]+)"] , unite @G*([</p> <p>([\s]+)"])</p> <p>Individual: @D*</p> <p>types: Dose</p> <p>annotations: rdfs:label @F*</p> <p>facts: doseValeur @F*(xsd:float</p> <p>["(^[\s]+)"] , unite @F*([</p> <p>([\s]+)"])</p> <p>Individual: @B*</p> <p>Types: EcartAdmissible</p> <p>annotations: rdfs:label @A*</p> <p>facts: valeurRetenueMax @E* ,</p> <p>valeurRetenueMin</p> <p>@D* ,</p> <p>composantSimple</p>
		B	revendication_complet_encoded	
		C	revendication	
		D	min_encoded	
		E	max_encoded	
		F	min	
		G	max	
		H	numero_AMM	

				<p>@C*</p> <p>individual: @H*</p> <p>facts:</p> <p>ecartAdmissibleRevendicationComplementaire @B*</p>
--	--	--	--	---

Validation de l'instanciation :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT ?p1 ?c_p1 ?p2 ?c_p2
WHERE {
  {select ?p1 (count(?p1) as ?c_p1) where {?s1 ephy:ecartAdmissibleElementComposition
?v1 . ?v1 ?p1 ?v11 . filter(?v11!=owl:NamedIndividual)} GROUP BY ?p1}
  UNION
  {select ?p2 (count(?p2) as ?c_p2) where {?s1
ephy:ecartAdmissibleRevendicationComplementaire ?v1 . ?v1 ?p2 ?v22 .
filter(?v22!=owl:NamedIndividual)} GROUP BY ?p2}
}

```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observations
numero_amm	1781	1781	
composition		1781	
max		1778	3 valeurs vide
min		1781	
num_amm_revendication	272	271	une valeur double
revendication			

max_revendication	42	42	
min_revendication		42	

Instanciación des cultures préconisé :

Description des fichiers sources :

1. le fichier "mfsc_et_mixte_usage_v3_utf8.csv"

Fichier source	"mfsc_et_mixte_usage_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
type produit	MFSC	Chaine de caractère
numero AMM	1150022	Chaine de caractère
nom produit	ALTERNAZOTE	Chaine de caractère
type culture libelle	Arbres et arbustes d'ornement	Chaine de caractère
dose min par apport	2.0	nombre flottant
dose min par apport unite	L/ha	Chaine de caractère
dose max par apport	1.0	nombre flottant
dose max par apport unite	kg/m3	Chaine de caractère
stade cultural min (BBCH)	81	nombre entier
stade cultural max (BBCH)	81	nombre entier
etat usage	Autoriser	Chaine de caractère
saison application min	Au semis	Chaine de caractère
saison application max	Mars	Chaine de caractère
saison application min commentaire	Une application par an au moment de la préparation du sol avec enfouissement en interculture.	Chaine de caractère
saison application max commentaire	stade montaison	Chaine de caractère
date decision	19/10/2016	date
culture commentaire	Gazons/prairies (Support de culture)	Chaine de caractère

2. le fichier “mfsc_et mixte_nombre_apport_minmax.csv”

Fichier source	"mfsc_et_mixte_nombre_apport_minmax.csvv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
Pas de nom mais représente nombre apport min	2	Nombre entier
Pas de nom mais représente nombre apport max	3	Nombre entier

Description des fichiers produits et script d'instanciation :

Fichier sources	"mfsc_et_mixte_usage_v3_utf8.csv" et "mfsc_et_mixte_nombre_apport_minmax.csvv"			
Fichiers produits	Nom du fichier	Noms de ces colonnes		La règle Celfie utilisé
	culture_preonise.xlsx	A	nombre_apport_min	individual: @C* types: CulturePreconise annotations: rdfs:label @D* , rdfs:comment @K* facts:typeCulture @F* , etat @I* , stadeCulturalMin @G*(xsd:integer) , stadeCulturalMax @H*(xsd:integer) , dateDecision @J*(xsd:date), nombreApportMax @B*(xsd:integer), nombreApportMin @A*(xsd:integer)
		B	nombre_apport_max	
		C	uri_encoded	
		D	uri	
		E	numero_AMM	
		F	type_culture_libelle	
		G	stade_cultural_min__BBCH_	
		H	stade_cultural_max__BBCH_	
		I	etat_usage	
		J	date_decision	
		K	culture_commentaire	
	culture_preonise_doses.xlsx	A	uri_encoded	individual: @C* types: Dose

		B	dose_max_encoded	annotations: rdfs:label @E* facts: doseValeur @F* , unite @G* individual: @B* types: Dose annotations: rdfs:label @D* facts: doseValeur @H* , unite @I* individual: @A* facts: doseParApportMax @B* , doseParApportMin @C*
		C	dose_min_encoded	
		D	dose_max	
		E	dose_min	
		F	dose_min_par_apport	
		G	dose_min_par_apport_unite	
		H	dose_max_par_apport	
		I	dose_max_par_apport_unite	
	culture_preonise_saison.xlsx	A	saison_min_encoded	individual: @A* types: SaisonApplication annotations: rdfs:label @D* , rdfs:comment @F* individual: @B* types: SaisonApplication annotations: rdfs:label @E* , rdfs:comment @G* individual: @C* facts: saisonApplicationMin @A* , saisonApplicationMax @B*
		B	saison_max_encoded	
		C	uri_encoded	
		D	saison_application_min	
		E	saison_application_max	
		F	saison_application_min_commentaire	
		G	saison_application_max_commentaire	

Validation de l'instanciation :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```

PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

```

```

SELECT ?p count(?p)
WHERE {
    ?c a ephy:CulturePreconise.

```



```

    ?c ?p ?v
    filter(?v != owl:NamedIndividual)
}
group by ?p

```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observations
numero_amm	2484	2484	
type_culture_libelle	2089	2089	
etat_usage	2484	2484	
date_decision	2127	2127	
stade_cultural_min__BBC H_	47	47	
stade_cultural_max__BB CH_	44	44	
nombre_apport_min	2009	2009	
nombre_apport_max	2253	2253	
culture_commentaire	1877	1877	
uri	2484	2484	
saison_application_min	2063	2063	
saison_application_max	287	287	
dose_max	2399	2399	
dose_min	2123	2123	

Instanciation des usages

Description des fichiers sources :

1. le fichier "produits_usages_v3_utf8.csv"

Fichier source	"produits_usages_v3_utf8.csv"	
Nom des colonnes	Exemple de valeur	Types des colonnes
numero AMM	1150022	Chaine de caractère
nom produit	ALTERNAZOTE	Chaine de caractère
identifiant usage	Artichaut*Trt Part.Aer.*Pucerons	Chaine de caractère
date_decision	26/09/2012	Date
stade cultural min (BBCH)	81	nombre entier
stade cultural max (BBCH)	81	nombre entier
dose retenue	0.6	nombre entier
dose retenue unite	L/ha	Chaine de caractère
etat usage	Autoriser	Chaine de caractère
delai avant recolte jour	30	nombre entier
delai avant recolte bbch	3	nombre entier
nombre max d'application	3	nombre entier
date fin distribution	31/03/2013	date
date fin utilisation	31/03/2014	date
condition emploi	Efficacité montrée sur les taupins (Agriotes sp.). Application dans le sillon de plantation.	Chaine de caractère
ZNT aquatique (en m)	20.0	nombre réel
ZNT arthropodes non cibles (en m)	50.0	nombre réel
ZNT plantes non cibles (en m)	5.0	nombre réel
mentions autorisées	Mention abeille	Chaine de caractère
intervalle minimum entre applications (jour)	7.0	nombre réel

Description des fichiers produits et script d'instanciation :

Fichier sources	"produits_usages_v3_utf8.csv"			
Fichiers produits	Nom du fichier	Noms de ces colonnes		La règle Celfie utilisé
	usages_identifiant.xlsx	A	identifiant_usage	individual: @A* types: UsageCatalogue annotations: rdfs:label @A* facts: porteeUsage @B* , methodeApplication @C* , groupeOrganismeNuisible @D*
		B	portee_usage	
		C	methode_application	
		D	groupe_organisme_nuisible	
	usages_simple.xlsx	A	uri	individual: @A* types: Usage facts: identifiantUsage @C*(Class) delaiAvantRecolteJour @H* (xsd:integer) delaiAvantRecolteBbch @I* (xsd:integer) stadeCulturalMin @F* (xsd:integer) dateDecision @E* dateFinUtilisation @L* (xsd:date) conditionEmploi @M* nombreApplicationMax @J* (xsd:integer) intervalleMinApplication @N* (xsd:integer) stadeCulturalMax @B* (xsd:integer) etat @G* dateFinDistribution @K* (xsd:date) individual: @D* facts: usage @A*
		B	stade_cultural_max__BBC H_	
		C	identifiant_usage	
		D	numero_AMM	
		E	date_decision	
		F	stade_cultural_min__BBC H_	
		G	etat_usage	
		H	stade_cultural_max__BBC H_	
		I	delai_avant_recolte_jour	
		J	delai_avant_recolte_bbch	
		K	date_fin_distribution	
		L	date_fin_utilisation	
		M	condition_emploi	
		N	intervalle_minimum_entre_applications__jour_	
	usages_mentions_autoriser.xlsx	A	uri	individual: @B* facts: mentionAutoriser @A*

		B	mentions_autorisees	
	usages_doses.xlsx	A	dose_encoded	individual: @A* types: Dose annotations: rdfs:label @B* facts: doseValeur @D* , unite @E*
		B	dose	
		C	uri	
		D	dose_retenue	individual: @C* facts: doseRetenue @A*
		E	dose_retenue_unite	

Validation de l'instanciation des identifiants d'usage :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT ?p count(?p)
WHERE {
    ?c a ephy:UsageCatalogue.
    ?c ?p ?v
    filter(?v != owl:NamedIndividual)
}
group by ?p
```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observations
identifiant_usage	1540	1540	
portee_usage	1540	1540	
methode_application	1540	1540	
groupe_organisme_nuisible	1441	1441	

Validation de l'instanciation des usages :

La validation sera faite , en regardant le nombre d'éléments respectivement dans les fichiers sources (plus exactement ceux prétraiter avec talend) et dans l'ontologie

Pour trouver le nombre d'éléments de chaque colonne d'un fichier nous avons utilisé Talend et côté ontologie on a utilisé la requête suivante :

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ephy: <http://www.d2kab.org/ontologies/ephy#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
```

```
SELECT ?p count(?p)
WHERE {
    ?c a ephy:Usage.
    ?c ?p ?v
    filter(?v != owl:NamedIndividual)
}
group by ?p
```

Les colonnes du fichier	Dans les fichiers sources	Dans l'ontologie	Observations
numero_AMM	74061	74061	
identifiant_usage	74061	74061	
etat_usage	74061	74061	
date_decision	57364	57364	
stade_cultural_min__BB CH_	5483	5483	
stade_cultural_max__BB CH_	5901	5901	
nombre_max_d_applicati on	20677	20677	
delai_avant_recolte_jour	13579	13579	
delai_avant_recolte_bbc h	143	143	
condition_emploi	11828	11828	

uri	74061	74061	
intervalle_minimum_entree_applications__jour_	3546	3546	
date_fin_distribution	20844	20844	
date_fin_utilisation	21423	21423	
mentions_autorisees	638	638	
ZNT_aquatique__en_m_	12560	12560	
ZNT_arthropodes_non_cibles__en_m_	3482	3482	
ZNT_plantes_non_cibles__en_m_	3159	3159	
dose_encoded	65039	65039	