



# Publication de données d'observation dans le Web de données

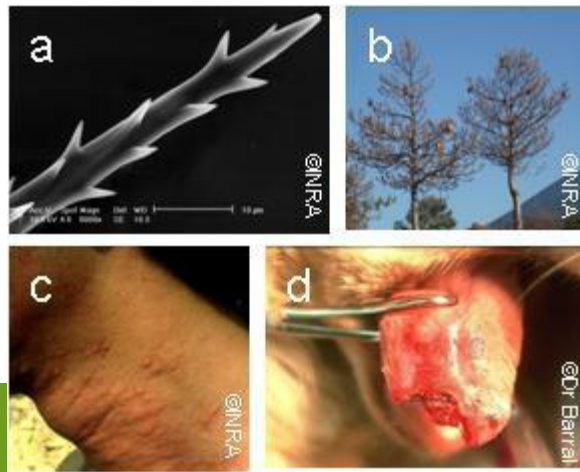


# Le projet scientifique – ALTERPRO

INRA, Plante&Cit , ONEMA, Plan Ecophyto 2018

## ❖ Probl matique

- ❖ Les chenilles sont responsables de la d foliation des pins (b) et des c dres
- ❖ Les poils des chenilles (a) sont tr s allerg nes (c, d)
- ❖ Les chenilles prolif rent sur le territoire depuis des ann es, aid es par le r chauffement climatique



# Le projet scientifique – ALTERPRO

INRA, Plante&Cit , ONEMA, Plan Ecophyto 2018

## ❖ Objectifs

- ❖ Contrôler les populations de Processionnaire du Pin
  - ❖ à un niveau tolérable permettant de protéger hommes, plantes et animaux
  - ❖ avec des moyens écologiques : piéger les papillons mâles avec des pièges à phéromones pour limiter la reproduction
- ❖ Comparer l'efficacité des pièges et des phéromones disponibles sur le marché

# Protocole de l'expérimentation

- 1 Pose des pièges à phéromones sur les sites test
- 2 Comptage des papillons piégés sur les sites test
- 3 Comptage des nids de chenilles sur les sites test et témoin

→ Période de piégeage des papillons

→ Période de dénombrement des nids de chenilles



# Les données de l'expérimentation

- ❖ **Données descriptives des sites** : géo-localisation, configuration paysagère, nombre d'arbres, hauteur des arbres, superficie, historique des traitements (chimiques ou non) contre la Processionnaire du Pin, etc.
- ❖ **Données descriptives de l'expérimentation** : types de piège et types de phéromone utilisés, nombre de pièges, dates de pose et de démontage des pièges
- ❖ **Résultats de l'expérimentation** : nombre de papillons piégés, nombre d'arbres relevés sur sites test et témoin, nombre de nids décomptés sur sites test et témoin

# Pourquoi une démarche LOD

- ❖ L'INRA s'est doté d'une Politique de partage des données
- ❖ Besoin interne en compétences sur les technologies du Web sémantique
- ❖ Les données de l'expérimentation présentent plusieurs intérêts
  - ❖ Pour d'autres chercheurs
  - ❖ Pour les gestionnaires d'espaces verts
  - ❖ Pour le grand public
- ❖ En l'état, les données étaient difficiles à comprendre et à réutiliser sans aide humaine
  - ❖ Sens des colonnes/lignes/cellules, cellules multi-valuées, unités de mesure hétérogènes, ambiguïtés, etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Communes	Saint-Fargeau-Ponthierry	Obernai	Angers							Saint-Martin-de-Ré	Marenne
2	Description Sites Témoins	Type de configuration	alignement	pas de site témoin déterminé	îlot boisé							îlot boisé	îlot boisé
3		Description des sites	parc de loisir		Echangeur de la Baumette							Parcelle forestière	à proximité la piscine Marenne
4		Géolocalisation	48°31'56.85"N, 2°33'47.42"E		47°27'53.61"N, 0°34'14.82"O							46°15'3.24"N, 1°31'0.58"O	45°49'28.21"N, 1°8'19.54"E
5		Surface	56 m		25100 m <sup>2</sup>							0,13 ha	10 ha
6		Historique du site	non communiqué		non communiqué							Comptage des nids par l'ONF depuis 2003 (données transmises)	non communiqué
7		Nombre de pins et/ou de cèdres	20		62							116	non communiqué
8		Nombre de sites tests	1		1	7							1
9	Tests	Type de configuration	alignement	îlot boisé	alignement	îlot boisé	alignement	alignement	alignement	îlot boisé	alignement	îlot boisé	îlot boisé
10		Description des sites	parc de loisir	jardin privé	Roseraie I bord de route	Roseraie II Parc	Roseraie III butte antibruit	Roseraie IV quartier résidentiel	Arboretum V Falun d'Orgemont	Arboretum VI cours d'école	Arboretum VII Plaine de jeux	parc	bosquet
11		Géolocalisation	48°31'59.30"N, 2°33'29.47"E	48°27'34.36"N, 7°30'18.04"E	47°26'55.62"N, 0°34'21.40"O	47°26'48.09"N, 0°33'51.29"O	47°26'33.64"N, 0°33'28.63"O	47°26'38.59"N, 0°33'14.78"O	47°26'43.17"N, 0°32'29.08"O	47°27'10.16"N, 0°32'47.32"O	47°27'45.99"N, 0°31'39.55"O	46°10'54.61"N, 1°22'26.48"O	45°49'14.14"N, 1°8'10.14"E
12		Surface	500 m	< 15 ares	3000 m <sup>2</sup>	17939 m <sup>2</sup>	8650 m <sup>2</sup>	8650 m <sup>2</sup>	3250 m <sup>2</sup>	5270 m <sup>2</sup>	1077 m <sup>2</sup>	4,6 ha	1 ha

A	B	C	D	E	F	G	H	I	J	K	L
	Communes	Saint-Fargeau-Ponthierry	Obernai	Angers							Saint-Ma de-Ré
	Nombre de pins et/ou de cèdres	96 pins	20 pins	97 pins 3 cèdres	18 pins	66 pins	22 pins	11 pins 3 cèdres	6 pins	31 pins	261 pins
	Hauteur des arbres	15 à 20 mètres	non relevé	10 à 15 mètres	10 à 25 mètres (avec un seul à 5 mètres)	non relevé	3 mètres à 20 mètres	5 mètres à 15 mètres	15 mètres à 20 mètres	4 mètres à 15 mètres	2 mètres à 10 mètres
Pièges à phéromones été 2011	Relevés des nids réalisés à l'hiver 2011	non réalisés	non réalisés	non réalisés							non réalisés s témoir
	Nombre de pièges installés	10	15	14 a) 1 b)	9	12 a) 1 b)	12 a) 2 b)	8 a) 1 b)	6 a) 1 b)	8	18
	Type de pièges	Procerex	Mastrap	a) Mastrap b) Nufarm	Mastrap	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	a) Mastrap b) Nufarm	Mastrap	Mastrap
	Type de phéromones	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex	Procerex
	Date de la pose des pièges	25 juillet (sem 30)	01 juin (sem 22)	01 juillet (sem 26)			01 juillet (sem 26) et 06 juillet (sem 27)		31 mai (sem 22) et 19 juillet (sem 29)	19 juillet (sem 29)	19 juillet (sem 29)
Premiers vols de papillons constatés	non communiqué	Sem 26 et 27	sem 25							non rele	



# Pourquoi une démarche LOD

- ❖ Lier les données de l'expérimentation avec d'autres données devient plus simple dans une démarche Web sémantique
  - ❖ Chaque donnée publiée est accessible par programme : requête Sparql, déréférencement, API
  - ❖ Les données publiées sont documentées
  - ❖ Les données sont décrites avec des vocabulaires connus et/ou documentés



# Les étapes de la publication

# Comprendre et qualifier les données

- ❖ Récupérer toute documentation utile sur les données
  - ❖ Schéma (XML, base de données relationnelle), rapports, etc.
- ❖ Se procurer un échantillon des données
- ❖ Désambiguïser, harmoniser, préciser

# Modéliser

- ❖ Lister les objets d'intérêt
  - ❖ Objets matériels (piège, phéromone, chenille, etc.), ou immatériels/concepts (piégeage, campagne d'observation, etc.)
- ❖ Décrire les objets : attributs, relations, classes
  - ❖ L'utilisation de **vocabulaires standards** accroît le potentiel **d'interopérabilité** et de **réutilisation** des données => LOV (Linked Open Vocabularies)
- ❖ Établir des liens vers des ressources externes
  - ❖ Privilégier des sources de référence (Insee pour les communes) ou provenant de partenaires => qualité, pérennité

# Créer la base de connaissances

- ❖ Identifier les données (URI: Uniform Resource Identifier)
  - ❖ Disposer d'une politique d'URI garantissant unicité (sans ambiguïté), pérennité (nom de domaine maîtrisé)
  - ❖ Préférer le protocole HTTP
- ❖ Transformer les données au format RDF
  - ❖ Par programme (XLWrap, XLST, etc.)
  - ❖ Avec des outils (TopBraid, GoogleRefine, etc.)
- ❖ Mettre les données dans un entrepôt RDF (Sesame, Jena, etc.)

# Faire savoir

- ❖ Catalogue de Linked data

- ❖ <http://datahub.io/>, Lov : <http://lov.okfn.org/dataset/lov/>

- ❖ Moteur de recherche sémantique

- ❖ Sindice : <http://sindice.com/>

- ❖ Portail ad hoc

# Linked Open Data ALTERPRO

LE PROJET

RESULTATS

FAQ

CONTACT

## Les concepts manipulés et leur sémantique

**CampagneObservation** : concept qui englobe toutes les données recueillies lors de la mise en œuvre du protocole ALTERPRO sur une année : pose des pièges en été, dénombrement des papillons piégés sur les sites tests en automne, dénombrement des nids de chenilles sur les sites tests et témoins en hiver.

**Étude** : ce concept englobe les données d'observation sur un site test et le site témoin correspondant. Nous sommes partis du postulat que pour chaque site test, il y a un site témoin qui est défini.

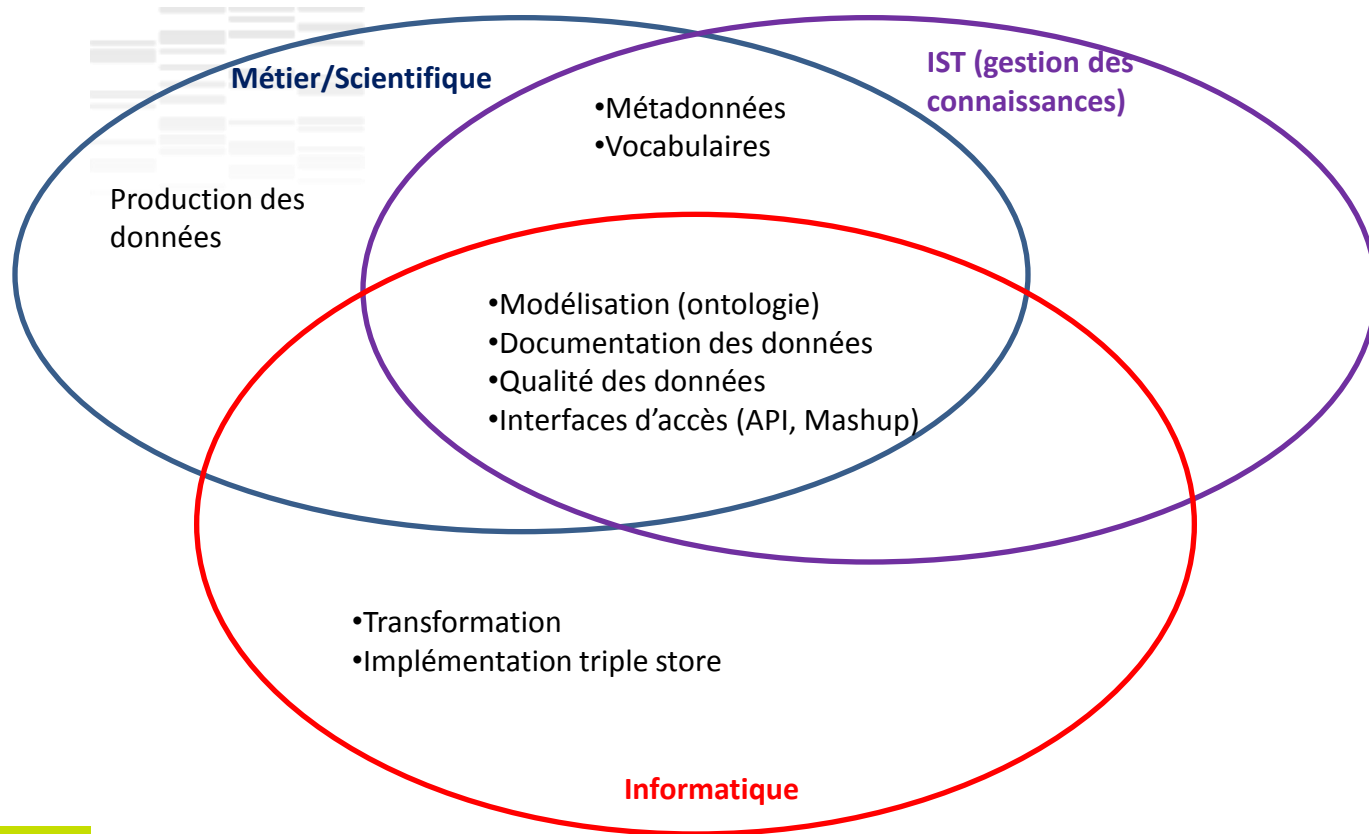
**Observation\_ThaumetopoeaPityocampa** : ce concept englobe l'ensemble des relevés d'observation (nombre de nids, nombre de papillons, etc.) sur

## L'ontologie

### La base de connaissances

La page d'accueil décrit le concept de CampagneObservation, ses propriétés, ses relations et donne la liste des objets de type CampagneObservation. Chaque propriété, relation, objet est cliquable. Pour explorer la base de connaissances, il suffit de naviguer en cliquant sur les différents liens hypertextes.

# Compétences mobilisées





# Résultats

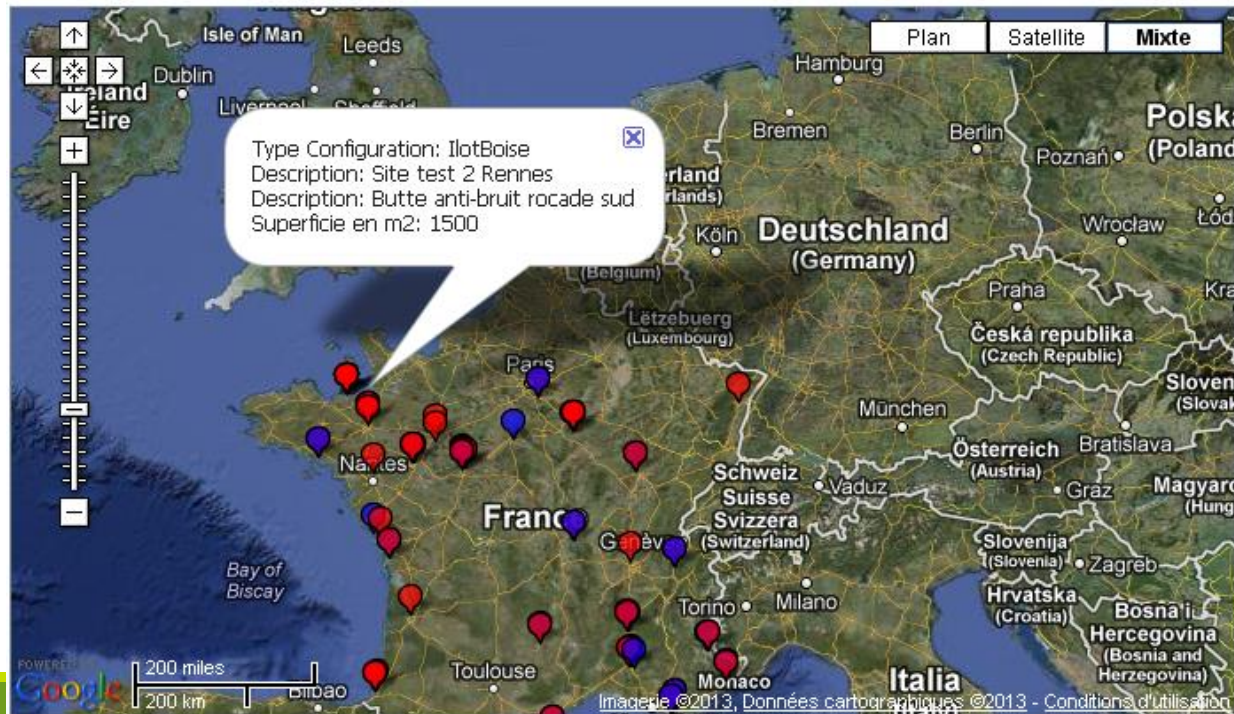
- ❖ Chaque donnée d'intérêt est représentée par un URI dérérérençable
- ❖ Des données décrites à l'aide de métadonnées documentées
- ❖ Des données faciles à comprendre, et à réutiliser : vocabulaires connus et/ou documentés, formats et protocoles standards
- ❖ Des données enrichies et mises en contexte grâce à divers liens : geospecies, qudt, Insee
- ❖ Des vocabulaires réutilisables : pièges, phéromones , configurations paysagères

# Sites d'observation

87 SiteObservation

CARTE • DETAILS

5 résultats sur 87 ne peuvent pas être tracés.



● Temoin ● Test

Search

## Localisation

- 2 Aix-les-Bains
- 8 Angers
- 2 Arzon
- 4 Avignon
- 3 Bayonne
- 2 Bordeaux

## Configuration

- 2 (missing this field)
- 16 Alignement
- 3 ArbreSeul
- 66 IlotBoise

## Type Site

- 28 Temoin
- 59 Test

# Les bénéfices internes

- ❖ Une manière efficace de répondre à un impératif : l'ouverture des données de la recherche
- ❖ Une occasion d'augmenter la qualité des données produites : clarification, harmonisation, enrichissement, explicitation et documentation des données de manière standardisée
- ❖ Une voie de sortie de la logique des silos
- ❖ L'opportunité de s'insérer dans un écosystème
- ❖ Un moyen supplémentaire de valoriser son expertise

# Les bénéfices pour la communauté

- ❖ Un gain de temps grâce à l'expertise disponible réutilisable : modèle, données
- ❖ Une réutilisation des données facilitée : données structurées, documentées, accessibles par programme
- ❖ Une possibilité d'enrichir ses propres données : liens



**Merci de votre attention!**