



**HAL**  
open science

## Etude de faisabilité du développement et de la valorisation d'une base de données sur l'évolution des pressions biotiques dans les parcelles agricoles

Vincent Cellier, Julia Agerberg, Michel Cariolle, David Gouache, Jean-Noël Aubertot

### ► To cite this version:

Vincent Cellier, Julia Agerberg, Michel Cariolle, David Gouache, Jean-Noël Aubertot. Etude de faisabilité du développement et de la valorisation d'une base de données sur l'évolution des pressions biotiques dans les parcelles agricoles. *Innovations Agronomiques*, 2013, 28, pp.141-153. 10.17180/r30x-3r77. hal-02650516

**HAL Id: hal-02650516**

**<https://hal.inrae.fr/hal-02650516v1>**

Submitted on 29 May 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

## Etude de faisabilité du développement et de la valorisation d'une base de données sur l'évolution des pressions biotiques dans les parcelles agricoles

Cellier V.<sup>1</sup>, Agerberg J.<sup>1</sup>, Cariolle M.<sup>2</sup>, Gouache D.<sup>3</sup>, Aubertot J.N.<sup>4</sup>

<sup>1</sup> Réseau PIC, INRA - Unité Expérimentale du domaine d'Epoisses, 21110 Bretenière

<sup>2</sup> ITB, 45 rue de Naples, 75008 Paris

<sup>3</sup> Arvalis – Institut du végétal, Service Génétique Physiologie et Protection des Plantes, Rue de Noetzelin - Bât. 630, 91405 - ORSAY CEDEX ; Coordonnées actuelles : Nestlé R& D, Tours

<sup>4</sup> INRA UMR 1248 AGIR, chemin de Borde-Rouge, BP52627, 31326 Castanet-Tolosan CEDEX

Correspondance : [vincent.cellier@epoisses.inra.fr](mailto:vincent.cellier@epoisses.inra.fr)

### Résumé

De 1943 à 2008, les Services de la protection des Végétaux (SRPV) ont élaboré des documents destinés à tenir informés les agriculteurs des pressions biotiques s'exerçant sur les cultures. Cette étude a permis de montrer l'intérêt de collecter les Avertissements Agricoles, les bilans de campagne nationaux des SRPV et certains documents des instituts techniques agricoles, qui permettront de constituer des séries documentaires longues. Ces corpus documentaires seront mis à disposition des utilisateurs avec des outils permettant la fouille et l'extraction d'informations. Des exemples d'études qu'il serait possible de réaliser avec ces informations ont été identifiés. L'étude a également montré l'intérêt de relier les observations sur les pressions biotiques à des données connexes permettant de mieux exploiter ces informations et propose de les collecter dans le cadre de l'actuel dispositif d'épidémiologie. Cette étude montre que le développement d'une base de données documentaire sur l'évolution des pressions biotiques dans les parcelles agricoles à partir de documents historiques était faisable et que cette base de données permettrait de générer des connaissances nouvelles, utiles à la gestion durable de la santé des cultures.

**Mots-clés** : pressions biotiques, Services de la Protection des Végétaux, Avertissements Agricoles, bilans de campagne nationaux, base de données documentaire, parcelles agricoles, recherche sémantique.

**Abstract: Feasibility of developing and potential uses of a database on the evolution of biotic constraints in agricultural plots.**

From 1943 to 2008, the French plant health protection agency (SRPV) produced documents intended to inform farmers of the biotic constraints in their crops. This study showed it was worthwhile to collect these pest warnings documents, national annual assessments of the SRPV and certain documents from agricultural technical institutes, because they will allow to constitute long documentary series. The documentary corpuses constituted would be made accessible to the users with tools allowing to search and to extract contained information. Examples of studies using this information were identified. The study also showed the interest of connecting these observations on biotic constraints with related data (climate, soil, agricultural practices...) to better exploit the information and propose their collection in the current epidemiological surveillance schemes. This study showed that the development of a documentary database on pest dynamics in commercial fields using historical documents was feasible, and that this database would provide new knowledge useful for sustainable management of crop health.

**Keywords**: biotic constraints, plant health protection agency, pest warnings system, annual national reports on pests, documentary database, commercial fields, semantic mining.

## Introduction

Pendant plusieurs décennies, les Services Régionaux de la Protection des Végétaux (SRPV) ont produit des documents destinés à tenir informés les agriculteurs et leurs conseillers des pressions biotiques subies par les différentes cultures : avertissements agricoles, bilans de campagne régionaux et nationaux. L'ensemble de ces données historiques de la Protection des Végétaux représente une source d'informations extrêmement précieuse, d'autant plus qu'elles recouvrent une très large gamme de situations de production, sur des séries temporelles longues et sur l'ensemble du territoire national.

Un projet intitulé « Etude de faisabilité du développement et de la valorisation d'une base de données sur l'évolution des pressions biotiques dans les parcelles agricoles » (mai 2010 – décembre 2011) a été soutenu par le Groupement d'Intérêt Scientifique Grande Culture à Hautes Performances Économiques et Environnementales (GIS GC-HP2E), afin d'étudier la faisabilité de la récupération et de la structuration de ces données historiques sur les pressions biotiques et de faciliter leur mobilisation par les acteurs de la recherche-développement. D'autres données, notamment celles des Instituts et Centres Techniques Agricoles (ICTA) et des chambres d'agriculture, pourraient s'y ajouter.

Au-delà de l'usage immédiat de préconisation pour lequel ces données ont été collectées, elles pourront servir à réaliser des études sur un plus long terme en analysant par exemple l'évolution spatio-temporelle de la répartition des bioagresseurs suivis, et en étudiant l'influence de différents facteurs sur cette répartition. Les connaissances ainsi générées pourront avoir des retombées sur les stratégies de protection intégrée des cultures en permettant d'anticiper l'évolution des pressions biotiques. Elles pourront également apporter des pistes de réflexion sur le recueil des données dans le cadre de la surveillance biologique du territoire réalisée actuellement.

### Les partenaires de l'étude de faisabilité

Le projet « Etude de faisabilité du développement et de la valorisation d'une base de données sur l'évolution des pressions biotiques dans les parcelles agricoles », coordonné par Vincent Cellier du réseau Protection Intégrée des Cultures de l'INRA (Réseau PIC), a regroupé 7 partenaires et un expert, issus de différentes structures de la recherche-développement :

- INRA : Jean-Noël Aubertot (UMR AGIR, Réseau PIC), Marie-Hélène Jeuffroy (UMR Agronomie), Julia Agerberg, (ingénieur CDD) et Vincent Cellier (UE Domaine expérimental d'Epoisses, Réseau PIC).
- Arvalis-Institut du végétal : David Gouache
- ITB : Michel Cariolle
- ITL : Florence Gorrias
- CETIOM : Xavier Pinochet
- UNIP : Véronique Biarnès
- APCA : Claire Vingut
- SRAI Midi-Pyrénées : Marc Delos (expert)

### L'étude de faisabilité s'est articulée autour de 3 axes principaux :

- Inventaire des données sur les pressions biotiques (Avertissements Agricoles, BSV, données des instituts, ...) et choix des documents qu'il est raisonnable d'espérer pouvoir collecter et structurer dans une base documentaire,
- Analyse des valorisations potentielles futures des documents collectés,
- Analyse de la mise en relation avec des données sur les pratiques agricoles et les états du peuplement végétal et réflexion sur les données connexes à acquérir : météo, sol, ...

## 1. La méthodologie employée

Le travail d'inventaire des documents intéressants à prendre en compte a été réalisé à partir d'enquêtes dans plusieurs structures : Services Régionaux de l'Alimentation (SRAI) qui sont dépositaires des documents des SRPV, Bibliothèque nationale de France (BnF), archives nationales, Instituts et Centres Techniques Agricoles (ICTA), Chambres d'Agriculture. Un travail d'inventaire plus poussé a été réalisé dans 3 SRAI (Bourgogne, Champagne-Ardenne et Midi-Pyrénées), ainsi qu'à l'ITB et au Cetiom, afin de recueillir des informations sur la volumétrie des documents à collecter et les problèmes techniques pouvant être rencontrés lors de la numérisation des documents papier ou la récupération des documents électroniques.

Le choix des documents à prendre en compte dans un projet de numérisation et de mise à disposition d'un fonds documentaire a été fait en prenant en compte les critères suivants :

- Disponibilité des documents
- Possibilité et facilité d'inventaire des documents,
- Centralisation des documents en cas de numérisation,
- Homogénéité dans le temps afin de pouvoir constituer des séries temporelles.

Le travail sur les valorisations des données a été fait, dans un premier temps, à l'aide d'un questionnaire en ligne abordant les questions des domaines de valorisation des données et des types de données souhaitées. Il a été diffusé à un public identifié par les partenaires du projet et potentiellement intéressé par une telle base documentaire. Des entretiens téléphoniques plus détaillés ont permis de compléter les réponses au questionnaire. Dans certains cas, si un projet précis pouvait être décrit, des fiches-projet ont été élaborées et pourront servir de base pour une première valorisation de la base documentaire.

Ces entretiens ont également permis de mettre en avant des fonctionnalités de la future base de données jugées intéressantes par les utilisateurs potentiels. La réflexion sur la mise à disposition des données ainsi engagée a été complétée par un travail sur la fouille et l'extraction de données avec deux équipes de l'INRA.

Le travail sur les données connexes a été réalisé en identifiant et en hiérarchisant les principaux facteurs physiques et biotiques influençant la dynamique de développement des bioagresseurs, grâce à des entretiens téléphoniques avec des experts. Cet inventaire a permis de sélectionner les éléments principaux à mettre en relation avec les informations sur les pressions biotiques de la future base documentaire, et, dans une approche plus prospective, à identifier des observations pertinentes qui pourraient être faites à l'avenir sur le terrain par les acteurs de l'épidémiosurveillance du territoire.

## 2. Résultats

### 2.1 Choix des documents à collecter

#### 2.1.1 Les documents des Services Régionaux de la protection des Végétaux (SRPV)

La collecte de quatre types de document des SRPV a été envisagée dans le projet :

- Les avertissements agricoles,
- Les fiches de notation,
- Les bilans de campagne régionaux,
- Les bilans de campagne nationaux.

Malgré l'intérêt des informations contenues dans les fiches de notation des SRPV, nous avons jugé que leur exploitation n'était pas prioritaire car impliquant beaucoup de problèmes techniques. En effet, ces fiches sont très nombreuses et hétérogènes, et, de ce fait, difficiles à rassembler et à exploiter. De plus,

elles sont manuscrites (tableaux imprimés remplis à la main sur le terrain), ce qui imposerait une saisie manuelle des informations et donc des coûts très élevés.

Nous préconisons également de ne pas inclure les bilans régionaux des SRPV dans une première campagne de numérisation, en raison de la dispersion de ces documents, disponibles uniquement dans le SRAI de la région concernée et d'une certaine redondance avec les bilans de fin de campagne présents dans les avertissements agricole, ainsi qu'avec les bilans de campagne nationaux comportant un tableau détaillant la pression de l'année par bioagresseur et par région. Nous proposons au contraire une organisation permettant de s'affranchir au maximum des recensements individuels dans chaque région, en se concentrant sur les collections de la BnF à Paris et sur un petit nombre de services régionaux sélectionnés.

Parmi les documents des SRPV, on privilégiera donc les **Avertissements Agricoles et les bilans de campagne nationaux**.

### **2.1.2 Les Avertissements Agricoles**

Les Avertissements Agricoles étaient des bulletins d'alerte de 2 à 3 pages destinés aux agriculteurs, rédigés par les Services Régionaux de la Protection des Végétaux (SRPV) et envoyés aux abonnés à une fréquence variable selon les périodes. Ils étaient basés sur les observations d'un réseau d'agents et d'experts qui notaient l'évolution des maladies (bactériennes, fongiques, parasitaires, etc.) ainsi que des insectes, dans les principales cultures de la région concernée. Des prévisions, parfois basées sur des modèles mathématiques, étaient également envoyées aux bénéficiaires de la veille phytosanitaire. En fonction des observations et des prévisions, des conseils sur les stratégies et les moyens de lutte étaient donnés par des spécialistes (préconisation de traitement ou de non-traitement). Des informations ou rappels sur l'évolution de la réglementation, les formations existantes, les prescriptions en matière d'usage, ... pouvaient accompagner les conseils.

Les Avertissements Agricoles ont existé sur la période 1943-2008, essentiellement sur les cultures pérennes au départ. A partir de 1960, les Avertissements Agricoles deviennent une publication payante sur abonnement et le système est étendu aux cultures légumières, au maïs et au colza. L'année 1975 marque le début des grands réseaux de surveillance des céréales à pailles en lien avec les agriculteurs et les instituts. A partir de 1980, les FREDON s'associent aux SRPV pour rédiger les avertissements, en équipe mixte jusqu'en 2000 puis avec des relations contractuelles de 2000 à 2009. En 2009, les Avertissements Agricoles sont remplacés par les Bulletins de Santé du Végétal (BSV).

### **2.1.3 Les bilans de campagne nationaux**

Des bilans nationaux étaient réalisés par thématique (généralement par pathosystème). Ces documents contenaient une partie « Bilan phytosanitaire » et une partie « Expérimentation ». Pour les rédiger, un rapporteur chargé d'un type de culture et d'un groupe de bioagresseurs donné envoyait un questionnaire aux spécialistes des SRPV et il faisait ensuite la synthèse des réponses, en s'appuyant sur son expérience personnelle. La partie « Bilan phytosanitaire » du document contient généralement un tableau donnant pour chaque région les informations suivantes : évolution de la maladie, perte de rendement, niveau d'attaque (note de 0 à 3 ou de 1 à 3), gravité par rapport à l'année précédente (inférieure, égale ou supérieure), traitements et seuils conseillés dans les avertissements de l'année, les traitements qu'il aurait fallu conseiller et des observations et remarques. Les bilans de campagne nationaux sont de volumétrie très variable : de quelques dizaines à quelques centaines de pages.

### **2.1.4 Documents des Instituts techniques agricoles**

Les fonds de tous les instituts techniques n'ont pas pu être inventoriés avec précision. Parmi ceux que nous avons pu consulter et étudier de manière détaillée, certains fonds de l'ITB et du Cetiom semblent en bonne adéquation avec les objectifs de la base documentaire et faciles à exploiter dans la mesure où ils sont centralisés et bien organisés.

Il s'agit, pour le Cetiom :

- De rapports d'activité diffusés depuis 1951 et contenant des résultats d'essais et d'études regroupés par catégorie (agronomie, lutte antiparasitaire, technologie ...). Certaines années, des bilans de campagne apparaissent.
- De bulletins, publiés mensuellement entre 1969 et 1990. Ils contenaient des articles sur différentes thématiques, avec une volonté de vulgarisation de l'information. Quelques bilans phytosanitaires apparaissent à partir de la 2<sup>ième</sup> moitié de la décennie 1970-1980, moins détaillés que ceux des SRPV.

En ce qui concerne l'ITB, les documents suivants ont été recensés :

- Des comptes-rendus de travaux : bilans annuels réalisés à l'échelle nationale, ils comportent une partie consacrée aux ennemis des cultures et, à partir des années 1980, un bilan synthétique de la campagne. La partie bioagresseurs contient des résultats d'expérimentation par bioagresseur ou type de bioagresseur : tests d'efficacité de produits, tests de doses de produits, tests variétaux...
- Des comptes-rendus d'essais : bilans annuels réalisés à l'échelle des régions betteravières, ils comportent un bref bilan de l'année, qui reprend les événements marquants de la campagne pour chaque type de bioagresseur. Les essais réalisés dans la région sont ensuite décrits. Quelques lignes en début de description de l'essai retracent la situation du bioagresseur concerné dans la région et sur l'année.

### **2.1.5 Documents des Chambres d'agriculture**

Malgré l'intérêt évident de ces documents, leur collecte n'a pas été jugée prioritaire en raison de leur hétérogénéité, de leur dispersion géographique et des difficultés d'inventaire.

## *2.2 Organisation de la collecte des documents*

Nous préconisons une démarche chronologique partant des documents les plus récents et remontant progressivement dans le temps, afin de ne pas avoir de coupure avec la période actuelle. Cela permet également de commencer par des documents informatisés (à partir de 1995), en théorie plus faciles à collecter, avant de s'atteler à la numérisation de documents papier, suivie obligatoirement par une opération de reconnaissance de caractères afin de pouvoir exploiter le contenu des documents. De plus, cette approche permettra de ne pas être confronté d'entrée de jeu aux documents les plus anciens qui risquent de poser des problèmes de numérisation et de reconnaissance de caractères en raison de leur âge, de la qualité d'impression et de leur fragilité.

Préalablement à toute opération de numérisation, un inventaire précis devra être réalisé afin de décrire chaque document à numériser, d'indiquer son état, de préciser s'il est relié de façon permanente ou non, et s'il est éventuellement possible de le détruire lors de la numérisation. Ces paramètres influenceront fortement sur le coût de l'opération de numérisation, qui pourra également être réduit en regroupant les différents fonds documentaires à numériser.

### **2.2.1 Les Avertissements Agricoles**

#### 2.2.1.1 Documents sous format électronique

La collecte des Avertissements Agricoles débutera par les documents sous format électronique même si une enquête réalisée par la Direction Générale de l'Alimentation du Ministère en charge de l'agriculture (DGAL) montre de nombreuses lacunes dans ce fonds. Des problèmes de lecture des fichiers collectés sont à prévoir, soit en raison du support utilisé qui peut être endommagé soit qu'ils

nécessitent un équipement ou un logiciel particulier pour relire les documents. Cette opération de collecte devra être lancée avec l'accord de la DGAL.

### 2.2.1.2 Documents papier

Nous préconisons l'exploitation de 2 sources centralisées où sont rassemblés les Avertissements Agricoles sous format papier :

- La Bibliothèque nationale de France (BnF) qui possède les Avertissements Agricoles toutes cultures sur la période 1960-2009, au titre du dépôt légal ;
- Les archives nationales qui possèdent des collections plus anciennes (1940-1980), versées par le Ministère en charge de l'agriculture.

Dans un premier temps, la collection de la BnF est la plus intéressante car, plus récente, elle permet de faire la jonction avec les Bulletins de Santé du Végétal (BSV) qui ont remplacé les avertissements agricoles à partir de 2009. La numérisation à la BnF s'effectue obligatoirement selon les cahiers des charges établis par celle-ci et la reconnaissance de caractère ne sera vraisemblablement pas suffisante pour exploiter valablement tous les documents numérisés. La négociation avec la BnF devra donc prévoir la mise à disposition des fichiers sources (Alto OCR et fichiers image), afin de pouvoir réaliser une reconnaissance de caractères de meilleure qualité. Le fonds de la BnF comportant des lacunes, il faudra les identifier avant l'opération de numérisation et rechercher les documents manquants dans les Services Régionaux de l'Alimentation (SRAI) afin qu'ils puissent être empruntés par la BnF pour être numérisés en même temps que ses propres collections. Les documents numérisés par la BnF seront mis à disposition du public dans la bibliothèque numérique Gallica.

S'il s'avère intéressant de remonter plus loin dans le temps, une opération de numérisation des archives nationales pourra être lancée en recourant aux services d'un prestataire. Il est à noter qu'un inventaire précis préalable à la numérisation sera nécessaire et que, pour certains documents, il faudra lever une interdiction de consultation.

## **2.2.2 Les bilans de campagne nationaux**

### 2.2.2.1 Documents sous format électronique

Les bilans de campagne nationaux disponibles sous format électronique sont rassemblés sur un serveur du Centre d'Etude et de Réalisation Informatiques de Toulouse (CERIT), mais cette collection comporte des lacunes. Le téléchargement de ces documents peut se faire à partir des sites des SRAI qui sont habilités à se connecter au serveur.

### 2.2.2.2 Documents papier

Etant donné le caractère national de ces documents, ils sont susceptibles de se trouver dans chaque SRAI. Nous proposons de démarrer la collecte dans quelques SRAI connus pour une bonne complétude de leurs collections, notamment les SRAI de Champagne-Ardenne et de Midi-Pyrénées dont nous avons inventorié les documents. Cette opération visera à collecter les documents disponibles uniquement sous forme papier et à compléter les lacunes observées dans les documents électroniques.

## **2.2.3 Les documents des Instituts et Centres Techniques Agricoles**

Les documents du Cetiom et de l'ITB que nous avons sélectionnés, sont centralisés et leur numérisation en sera facilitée. Après récupération des documents disponibles sous format électronique, il faudra réaliser un inventaire précis des documents papier disponibles afin de pouvoir préparer un appel d'offres de numérisation.

### 2.3. La valorisation des documents collectés

Une partie de l'étude de faisabilité a été consacrée à l'analyse des valorisations possibles d'une base documentaire sur les pressions biotiques. Il s'agissait de recueillir des idées précises sur les besoins en termes de données sur les pressions biotiques et sur d'éventuels projets d'exploitation de ces données, ainsi que de déterminer avec les utilisateurs potentiels les fonctionnalités importantes de la future base de données documentaire.

#### 2.3.1 Les pistes de valorisation envisagées

Nous nous sommes d'abord intéressés aux valorisations imaginées par les futurs utilisateurs potentiels, s'ils avaient accès aux données des SRPV ou des instituts techniques. Les réponses au questionnaire en ligne concernant cette question ont été regroupées en grandes catégories, les principales étant présentées ci-dessous :

Domaine de valorisation	Nombre de réponses (% de personnes enquêtées)
Le développement de modèles et/ou d'outils d'aide à la décision	9 (43%)
Une meilleure description des évolutions spatio-temporelles des populations de bioagresseurs et/ou des dégâts à différentes échelles	7 (33%)
L'analyse des relations entre pratiques agricoles et pressions biotiques (comparaison d'impacts de pratiques...)	7 (33%)
La prévision et le conseil	3 (14%)
L'analyse des relations entre variables climatiques et pressions biotiques	3 (14%)
L'analyse de relations entre bioagresseurs	2 (10%)
L'optimisation des pratiques agricoles (évaluation de risque...)	3 (10%)
La gestion des résistances	2 (10%)
La description de l'évolution spatio-temporelle des pratiques agricoles	2 (10%)

**Tableau 1** : Domaines de valorisation cités par les personnes ayant répondu au questionnaire en ligne

On constate un très fort intérêt pour les jeux de données sur les pressions biotiques dans le domaine de la conception et la validation de modèles et d'outils d'aide à la décision (Wallach *et al.*, 2006). Les personnes enquêtées citent aussi leur volonté de mieux décrire et expliquer les pressions biotiques. De nombreuses personnes enquêtées souhaitent explorer les liens entre les pressions biotiques et d'autres facteurs, les pratiques agricoles notamment. Citons également, en termes de valorisations envisagées, l'étude de la phénologie des cultures et la réflexion sur le système de suivi épidémiologique actuel par les Bulletins de Santé du Végétal (BSV), sur la base des lacunes constatées dans les observations réalisées pour l'élaboration des Avertissements Agricoles (données connexes manquantes, manque d'homogénéité dans les notations...).

Lors des entretiens, plusieurs interlocuteurs ont déclaré avoir déjà eu recours aux documents historiques des SRPV dans le cadre de projets, en particulier pour les usages suivants :

- Le calibrage de modèles et d'OAD,
- L'estimation des tendances générales d'évolution des bioagresseurs,
- La sélection des principaux bioagresseurs à étudier (à partir de leurs fréquences de citation dans les documents).

Dans la grande majorité des cas, ce sont les Avertissements Agricoles qui ont été exploités, probablement en raison de leur disponibilité et de la meilleure connaissance de leur existence et de leur contenu, par rapport aux bilans de campagne. Ces exemples montrent que les possibilités de valorisation des documents de la PV sont nombreuses et prouvent la faisabilité pratique de l'exploitation de ces données, surtout si l'accès en est facilité.

### **2.3.2 Données souhaitées**

Au sein d'un document, les personnes enquêtées ont signalé qu'elles cherchaient d'abord les informations suivantes :

- La date et la localisation géographique de l'observation,
- Le pathosystème étudié,
- L'organe observé,
- L'interprétation du risque (faible, moyen, fort), par situation agronomique,
- La phénologie des cultures et des bioagresseurs,
- Les liens entre bioagresseurs,
- La présence d'un bioagresseur peu répandu (différent des principaux bioagresseurs abordés systématiquement dans chaque document).

Les échelles spatio-temporelles recherchées dépendent évidemment de l'utilisation envisagée. On note des préférences pour l'échelle de la région, de la parcelle et du pays. La plupart des personnes enquêtées s'accordent pour dire que les jeux de données fournis doivent être les plus larges possibles : des durées de 40 ou 50 ans sont évoquées. Quant à la fréquence d'observation, les échelles mensuelle et hebdomadaire sont souvent citées.

Les données textuelles, comme les données quantitatives, intéressent les futurs utilisateurs. Beaucoup estiment que plus les données sont précises plus les exploitations pourront être poussées mais, dans beaucoup de cas, des données qualitatives et assez globales (notes annuelles de pressions ...) pourraient suffire.

Il est mis en avant que les données anciennes présentent un avantage certain puisqu'elles sont peu disponibles généralement et que des bases de données existent déjà pour les données plus récentes.

L'origine des données (parcelles d'agriculteurs et non parcelles expérimentales) est aussi considérée comme un atout par rapport à d'autres données disponibles.

### **2.3.3 Difficultés d'exploitation identifiées**

Nous avons recueilli l'avis des personnes ayant déjà travaillé sur des données issues des SRPV ou des données similaires, sur les difficultés éventuelles qu'elles ont pu rencontrer. Nous avons fait la liste de ces difficultés ci-dessous :

- L'hétérogénéité des protocoles : ceux-ci sont théoriquement harmonisés, mais varient suivant les régions. De plus les observations ont été menées par des organismes différents (PV, FREDON, FREDEC...),
- Les erreurs d'observation,
- L'imprécision des données (beaucoup de caractérisations qualitatives des pressions),
- Le manque de références : dans certains cas d'attribution d'une note annuelle de pression par exemple, aucune échelle de notation rigoureuse expliquant les scores attribués n'est disponible,
- La pression décrite est parfois liée au stade végétatif de la culture plutôt qu'aux observations réellement réalisées,
- Un échantillonnage des parcelles suivies mauvais ou insuffisant,
- L'hétérogénéité du vocabulaire utilisé (par exemple mauvaise orthographe de mots, changements de dénominations au cours du temps...),
- Des séries de données pas toujours complètes si elles n'ont pas été reprises certaines années dans les documents,
- Le temps important demandé pour récolter les documents, repérer les informations pertinentes et les traduire en variables utilisables.

Les personnes enquêtées ont cependant trouvé des solutions pour contourner certaines de ces difficultés.

Pour pallier en partie la variabilité des parcelles suivies d'une année à une autre, certains interlocuteurs ont travaillé sur 80% de points fixes et 20% de points variables dans chaque région.

Par ailleurs, ils n'ont comparé les données qu'entre elles (pas de croisement avec des données extérieures), en faisant l'hypothèse que les biais étaient les mêmes pour toutes les données. En adoptant ce type de méthodes, les problèmes d'homogénéité ont été maîtrisés.

Dans certains cas, les données présentes dans les Avertissements Agricoles peuvent être corrigées par les bilans de campagne nationaux.

Dans beaucoup de cas, les données de la PV sont jugées de bonne qualité, et suffisantes pour des études « grosse maille ». A ce sujet, la pertinence d'une échelle large (grandes tendances régionales ou nationales), *a priori* la plus adaptée pour les données disponibles, a été soulignée par d'autres interlocuteurs. En effet, les liens entre pratiques agricoles et bioagresseurs à l'échelle de la parcelle sont déjà connus et étudiés par les différents organismes à travers des expérimentations ciblées : dans tous les cas, la base de données à construire ne serait pas le meilleur outil pour cette échelle.

### 2.3.4 Projets pilotes

Lors des entretiens réalisés avec de futurs utilisateurs potentiels, nous avons identifié un certain nombre de projets de valorisation des données, textuelles ou chiffrées. Ces projets pilotes montrent la diversité des questions techniques et scientifiques à laquelle la base de données documentaire pourra répondre.

Thématique
Construction d'un OAD sur l'antracnose du pois.
Analyse et hiérarchisation des dommages causés par les bioagresseurs du blé à l'échelle nationale.
Analyse des relations entre occupation des terres et intensité de bioagresseurs à l'échelle de petites régions.
Etude de la co-évolution de bioagresseurs par analyse textuelle.
Etude des savoirs et dispositifs des transitions vers une production agricole durable (passage Avertissements Agricoles - BSV).
Reconstitution du climat passé en utilisant des vieilles séries d'observation phénologiques et modélisation des effets du changement climatique sur l'évolution des bioagresseurs des principales cultures à l'échelle nationale.
Etude de l'évolution des épidémies de mildiou sur pomme de terre au cours du temps.

**Tableau 2** : Liste des projets pilotes identifiés lors de l'étude.

Il est apparu important d'impliquer fortement les porteurs de ces projets à la conception de la base de données documentaire et aux outils qui y seront associés afin qu'elle corresponde à leurs besoins. En retour, ces porteurs de projet bénéficieront d'un accompagnement dans l'exploration des documents et la constitution de séries de données.

## 2.4. La mise à disposition des documents collectés

Comme l'indiquent les entretiens avec des utilisateurs potentiels, les documents que nous proposons de collecter contiennent une quantité importante de données valorisables et il apparaît difficile de chercher à extraire et structurer l'ensemble de ces informations sans savoir si elles vont effectivement être utilisées ni quand elles le seront. La proposition est de mettre à disposition le fonds documentaire avec des outils permettant la fouille des documents, afin que chaque utilisateur puisse rechercher les informations dont il a précisément besoin et générer lui-même la base de données idoine.

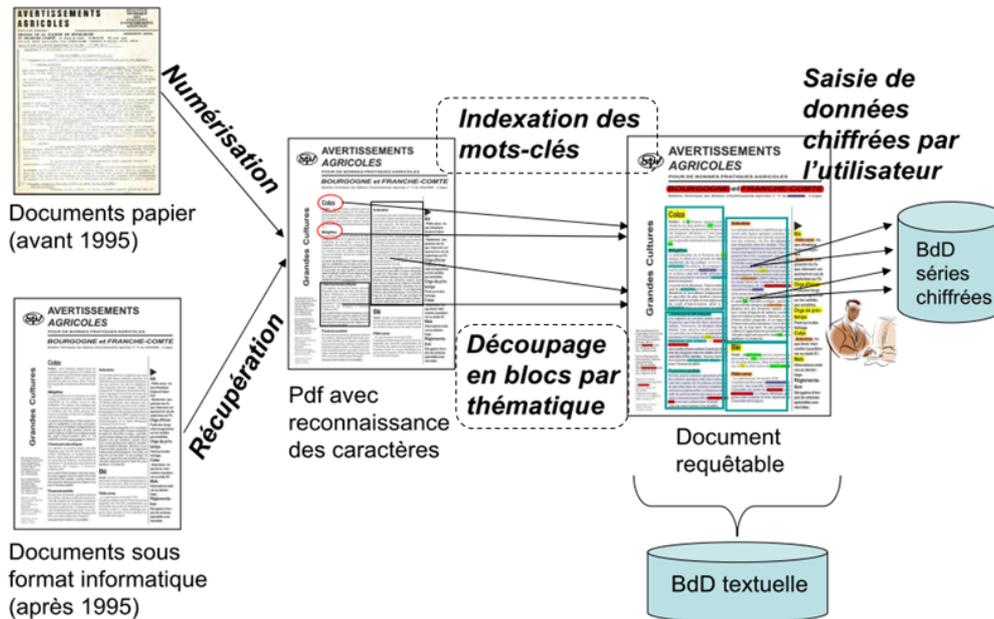


Figure 1 : Proposition détaillée d'exploitation et de mise à disposition des documents

L'utilisateur pourra ainsi saisir les données particulières qui l'intéressent (notamment les séries de données chiffrées), afin qu'elles puissent être mises à disposition des autres utilisateurs. La démarche est donc participative.

La mise à disposition des ressources documentaires nécessitera donc de créer ou d'adapter un outil existant aux besoins des utilisateurs. En particulier, un travail sur les terminologies<sup>1</sup> et les ontologies<sup>2</sup> du domaine devra être entrepris. L'outil devra pouvoir répondre aussi bien à des besoins simples de classement et de recherche de documents qu'à des besoins plus complexes de recherche sémantique et d'extraction d'informations. L'utilisation des documents par ces outils nécessite une adaptation de leur format, en particulier s'ils comportent plusieurs colonnes de texte.

Nous avons pu tester 2 outils :

- La plateforme CorText, soutenue par l'unité de recherche SenS 1326 de l'Institut Francilien « Recherche, Innovation et Société » (IFRIS), propose une extraction de l'information des documents pour générer une base structurée, et la production d'outils d'analyse d'exploration (analyse spatiale, analyse temporelle, combinaisons des associations). L'application de ces méthodes aux bulletins d'épidémiologie peut permettre de répondre à des questions agronomiques sur l'évolution d'un bioagresseur ou un couple culture-bioagresseur au cours du temps, voire sur l'interaction entre des bioagresseurs. Elle permet aussi d'aborder des

<sup>1</sup> Terminologie : ensemble des termes, rigoureusement définis, qui sont spécifiques d'une science, d'une technique, d'un domaine particulier de l'activité humaine.

<sup>2</sup> Ontologie : ensemble structuré des termes et concepts représentant le sens d'un champ d'information.

questions sociologiques, par exemple l'évolution de la construction des documents d'épidémiosurveillance (acteurs impliqués ...).

- La technologie Alvis, utilisée par l'équipe Bibliome de l'INRA (Claire Nédellec), qui réalise une analyse sémantique fine des textes, basée sur une termino-ontologie du domaine pour le développement de services en ligne à base de connaissance :
  - Moteur de recherche sémantique
  - Aide à la lecture rapide de document
  - Extraction d'information et formalisation dans une base de données

Sur les avertissements agricoles et les BSV, il serait possible d'étudier les relations entre :

- la plante, la variété, le stade de végétation, le bioagresseur, la maladie,
- les facteurs environnementaux (sol, rotation des cultures, pluviométrie, température),
- la localisation géographique,
- la date de l'observation.

### 2.5. Les données connexes à recenser

Les réponses aux entretiens sur les facteurs importants influençant le développement des bioagresseurs sont résumées ci-dessous. Pour établir une hiérarchie entre les différents facteurs cités lors des entretiens, nous nous sommes basés sur leur fréquence de citation, c'est-à-dire le nombre de fois où un facteur a été cité sur les 11 entretiens.

	Les plus cités
<b>Facteurs climatiques</b>	Température Pluviométrie Type de climat
<b>Facteurs agronomiques</b>	Rotation Travail du sol Date de semis Gestion des résidus Variété Fertilisation Densité de semis
<b>Facteurs environnementaux</b>	Espèces végétales présentes dans les bordures Assolement de la petite région agricole Présence de bois et de bosquets Présence de haies
<b>Facteurs pédologiques</b>	Texture du sol
<b>Facteurs topographiques</b>	Exposition de la parcelle Pente

**Tableau 3** : Principaux facteurs influençant le développement des bioagresseurs cités.

Nous avons vérifié si des données relatives à ces facteurs jugés importants à prendre en compte étaient présentes dans les documents historiques de la PV. Il en ressort que des données climatiques apparaissent fréquemment dans les documents, mais souvent à des échelles assez globales. Quelques données agronomiques telles que la date de semis ou la variété peuvent y figurer, mais pas systématiquement. Les facteurs environnementaux, pédologiques et topographiques ne sont quasiment jamais décrits.

Au vu du nombre important de facteurs non recensés dans les documents sur les pressions biotiques, on peut envisager d'utiliser des bases de données existantes susceptibles d'apporter les informations

manquantes mais, dans beaucoup de cas, la mise en relation entre les observations présentes dans les documents et les bases de données connexes nécessite une bonne localisation, indications rarement précises dans les documents historiques des SRPV. Dans le meilleur des cas le nom de la commune est précisé. Seules des données connexes assez générales (par commune, petite région agricole, département...) pourront donc être obtenues. Ce degré d'information est néanmoins suffisant dans de nombreux cas d'exploitation des documents évoqués par des futurs utilisateurs potentiels.

## En conclusion

L'étude dont les résultats sont présentés ici a montré la pertinence du développement d'une base de données documentaire sur les pressions biotiques pour répondre à une large gamme de questions scientifiques et techniques exprimées par les acteurs de la recherche agronomique et du développement agricole. La collecte et la mise à disposition des documents et des outils nécessaires à leur exploitation nécessiteront l'accord et la collaboration de la DGAL pour les documents issus des SRPV, et du Cetiom et de l'ITB pour leurs documents respectifs.

L'approche étant innovante dans le domaine de la gestion de la santé des cultures, il est important de prévoir un accompagnement des premiers utilisateurs des outils qui seront mis en place, en particulier les outils d'extraction d'informations sémantiques.

Il serait également intéressant que les propositions concernant les données connexes qui permettraient de mieux exploiter ces sources d'information soient étudiées par les structures en charge de la surveillance biologique du territoire et de l'élaboration des BSV. Quelques propositions d'amélioration du système actuel, les Bulletins de Santé du Végétal (BSV), peuvent être formulées :

- Rendre obligatoires certaines données déjà demandées de manière facultative dans les outils utilisés actuellement (principalement VGObs et Vigicultures pour les grandes cultures) : travail du sol, précédents culturaux, exposition de la parcelle, géoréférencement de la parcelle (pour faciliter la mise en relation avec des bases de données connexes : Météo France, Registre Parcellaire Graphique, Géoportail, référentiel régional pédologique pour n'en citer que quelques-unes) et les faire remonter dans la base de données centrale de la DGAL.
- Ajouter d'autres données importantes citées par les personnes enquêtées : des précisions sur les pratiques (gestion des résidus, fertilisation) et des informations sur l'environnement de la parcelle (assolement de la petite région agricole, espèces présentes dans les bordures de parcelles, présence d'éléments paysagers remarquables du type bois, bosquet, haie), les plus légères possibles, pour ne pas alourdir le travail d'observation, déjà conséquent.

Des propositions d'amélioration plus conséquentes ont également été formulées :

- Un suivi plus approfondi des auxiliaires des cultures et des facteurs qui influencent leur développement,
- Le suivi systématique d'un réseau de parcelles non traitées afin de disposer de références sur les dynamiques de bioagresseurs non biaisées par la lutte chimique.

Enfin, cette étude a été réalisée dans le cadre du GIS GC-HP2E et ses résultats ne concernent donc que la filière des grandes cultures, même si certains documents sont communs à d'autres filières. Il semble évident que la démarche devrait être étendue à ces autres filières afin de maximiser l'efficacité de la démarche mise en œuvre.

Il ne sera sans doute pas possible de collecter l'ensemble des fonds documentaires mentionnés dans cette étude au sein d'un unique projet et il semble judicieux de commencer par les documents des SRPV qui représentent un volume important, assez homogène et bien connu des utilisateurs potentiels. Une collecte, même partielle, de ce fonds permettrait de réaliser une première version de la base documentaire et de ses outils d'extraction d'informations et de montrer l'intérêt de poursuivre le travail de collecte documentaire.

## Remerciements

Les auteurs tiennent à remercier le GIS GC-HP2E qui a financé cette étude, les représentants des partenaires, qui ont participé au pilotage du projet, le personnel des SRAI dans lesquels les inventaires précis ont été réalisés (Jean-blaise Davaine, Thierry Durand et Dominique Jacquin (Bourgogne), Olivier Pilon (Champagne-Ardenne) et Marc Delos (Midi-Pyrénées)) ainsi que toutes les personnes qui ont eu la gentillesse de répondre à leurs sollicitations et en particulier M. Frédéric Vey, chef du Bureau des biotechnologies, de la biovigilance et de la qualité des végétaux du Ministère en charge de l'agriculture qui nous permis d'accéder aux informations des Services Régionaux de l'Alimentation concernant la protection des végétaux et Mme Muriel Amar qui nous a fait découvrir certains aspects du fonctionnement de la BnF.

## Références bibliographiques

- Aubertot J.-N., Barbier J.-M., Carpentier A., Gril J.-J., Guichard L., Lucas P., Savary S., Savini I., Voltz M., 2005. Pesticides, agriculture et environnement. Réduire l'utilisation des pesticides et limiter leurs impacts environnementaux. Rapport d'expertise scientifique collective, INRA et CEMAGREF (France).
- Aubin S., Bessieres P., Bossy R., Gillard L., Jourde J., Papazian F., Veber P., Nedellec C., 2009. BioAlvis II, NLP-based semantic mining of literature on molecular biology of bacteria. BioCreative II.5 workshop: Special session on digital annotations, Madrid (ESP).
- Claerr T., Westeel I. *et al.*, 2010. Numériser et mettre en ligne. Presses de l'ENSSIB. 200 p.
- Fitt B.D.L., Hu B.C., Li Z.Q., Liu S.Y., Lange R.M., Kharbanda P.D., Butterworth M.H., White R.P., 2008. Strategies to prevent spread of *Leptosphaeria maculans* (phoma stem canker) onto oilseed rape crops in China; costs and benefits. *Plant Pathology* 57, 652–664.
- Fussler L., Kobes N., Bertrand F., Maumy M., Grosman J., Savary S., 2008. A Characterization of Grapevine Trunk Diseases in France from Data Generated by the National Grapevine Wood Diseases Survey. *Phytopathology* 98, 571-579.
- Gladders P., Paveley N.D., Barrie I.A., Hardwick N.V., Hims M.J., Langton S., Taylor M.C., 2001. Agronomic and meteorological factors affecting the severity of leaf blotch caused by *Mycosphaerella graminicola* in commercial wheat crops in England. *Annals of Applied Biology* 138, 301-311.
- Jeger M.J. et Pautasso M., 2008. Plant disease and global change – the importance of long-term data sets. *New Phytologist* 177, 8–11.
- Shaw M.W., Bearchell S.J., Fitt B.D.L., Fraaije B.A., 2008. Long-term relationships between environment and abundance in wheat of *Phaeosphaeria nodorum* and *Mycosphaerella graminicola*. *New Phytologist* 177, 229–238.
- te Beest D.E., Shaw M.W., Pietravalle S., van den Bosch F., 2009. A predictive model for early-warning of Septoria leaf blotch on winter wheat. *European Journal of Plant Pathology* 124, 413-425.
- Turenne N., Barbier M., BELUGA : un outil pour l'analyse dynamique des connaissances de la littérature scientifique d'un domaine - Première application au cas des maladies à prions.
- Hébrail G., Lebart L. eds., 2004. Extraction et gestion des connaissances, dans les actes de la conférence : Extraction et Gestion de Connaissances (EGC'04), Clermont-Ferrand, France, pp 423-428.
- Wallach D., Makowski D., Jones J.-W., 2006. Evaluating crop models, Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications. Elsevier, Amsterdam, pp. 11-53
- Welham S.J., Turner J.A., Gladders P., Fitt B.D.L., Evans N., Baierl A., 2004. Predicting light leaf spot (*Pyrenopeziza brassicae*) risk on winter oilseed rape (*Brassica napus*) in England and Wales, using survey, weather and crop information. *Plant Pathology* 53, 713-724.
- Zwankhuizen M.J., Zadoks J.C., 2002. *Phytophthora infestans*'s 10-year truce with Holland: a long-term analysis of potato late-blight epidemics in the Netherlands. *Plant Pathology* 51, 413–423.