

## RAPPORT SCIENTIFIQUE DU PROGRAMME CONVEX

### Table des matières

Présentation .....	2
Rappel des opérations annoncées .....	2
Résumé du projet .....	2
Résultats – résumé des actions menées .....	2
Traitement documentaire et indexation : .....	2
Traitement technique des documents : .....	5
Livrables.....	9
Documents numérisés et fichiers pour publication .....	9
Instruments de recherche encodés en XML-EAD.....	9
Publication des contenus sur le portail PANDOR de la MSH de Dijon .....	9
Évènements scientifiques.....	9
Perspectives scientifiques .....	10
Projet CONVEX 2.....	10
Autres perspectives.....	11
Annexes .....	11
Annexe 01 : Description et indexation du Bulletin international de la répression des fraudes (1910-1916) et des Annales des falsifications et des fraudes (1917-1938) .....	11
Annexe 02 : Instrument de recherche décrivant le Bulletin international de la répression des fraudes.....	13
Annexe 03 : Instrument de recherche décrivant les Annales des falsifications et des fraudes ....	13
Annexe 04 : Programme de la journée de restitution de CONVEX .....	13

CONVEX\_MSHDijon\_Rapport scientifique

## Présentation

**Programme** : Collection numérique vitivinicole d'excellence

**Porteur du projet** : Université de Bourgogne -

**Porteur scientifique** : Jean-Pierre Garcia.

**Durée du projet** : 18 mois, octobre 2018-mars 2020

**Budget total du projet** : 61 850 €

**Financement demandé et obtenu** : 14 750 €

## Rappel des opérations annoncées

Traitement documentaire, indexation, numérisation et mise en ligne des ressources suivantes :

- ❖ Bulletin international de la répression des fraudes
- ❖ Annales des falsifications et des fraudes

## Résumé du projet

Il s'agit tout d'abord, de poursuivre et développer, à l'aide de corpus anciens, inédits sous format numérique, la politique documentaire 3.0 amorcée depuis quelques années à la MSH de Dijon. Mais ici, le caractère diversifié et rare des nouveaux corpus à traiter permettent de surcroît, d'envisager des approches holistiques et multiscalaires d'un grand intérêt scientifique. Ces corpus apportent en outre une contribution à une meilleure connaissance de l'histoire des sciences, techniques et réglementations vitivinicoles par le traitement et la diffusion des contenus publiés dans le Bulletin International de la répression des fraudes. Enfin, ce projet a permis de mieux appréhender, au niveau national comme international, la construction et l'ancrage des normes de production et de commercialisation des vins, offrant des outils d'analyse rares aux juristes, historiens, géographes ou économistes, ainsi qu'aux professionnels du vin.

## Résultats – résumé des actions menées

*Préambule : au regard de la situation sanitaire générale, la fin du projet CONVEX – en accord avec CollEx-Persée – a été reportée au 31/03/2021.*

Le Bulletin international de la répression des fraudes et les Annales des falsifications et des fraudes ont fait l'objet du même processus de traitement documentaire et technique.

## Traitement documentaire et indexation :

- ❖ Le plan de classement est chronologique (un niveau par année, puis un niveau pour chacun des numéros), le sommaire de chaque numéro est détaillé sur chaque notice, un index thématique annuel est également proposé.

## CONVEX\_MSHDijon\_Rapport scientifique

- ❖ L'élaboration des instruments de recherche archivistiques électroniques a été effectuée par saisie dans un tableau par les archivistes recrutés dans le cadre du projet, la transformation au format XML-EAD a été réalisée par le pôle informatique de la MSH, à l'aide d'un logiciel développé en interne.
- ❖ Le lien vers les fichiers numérisés de chaque numéro a été créé au moment de la transformation en EAD.
- ❖ Plusieurs tests de publication ont été réalisés pour vérifier la qualité et l'exhaustivité des données ainsi que leur affichage sur un environnement de travail interne.
- ❖ Le recettage (vérification de conformité) de l'instrument de recherche a été effectué sur une interface de pré-production de l'outil PANDOR.
- ❖ La mise en ligne effective de l'instrument de recherche archivistique et des documents numériques associés sur le portail PANDOR (Portail Archives Numériques et DONNÉES de la Recherche) de la MSH de Dijon sera réalisée à la livraison de la nouvelle version du portail (juin 2021).
- ❖ La notice-type a été conçue à la fois pour s'intégrer à la logique du portail PANDOR, notamment aux autres fonds de la thématique vigne, vin et gastronomie labellisés par CollEx-Persée et pour répondre aux mieux aux attentes des scientifiques quant aux possibilités d'interrogation de ce corpus spécifique.

Plusieurs réunions associant ingénieurs et scientifiques ont eu lieu (10 juillet 2019, 14 octobre 2019, 22 novembre 2019, 16 décembre 2019, 31 janvier 2020). Voir la synthèse des échanges en annexe 1 *Description et indexation du Bulletin international de la répression des fraudes (1910-1916) et des Annales des falsifications et des fraudes (1917-1938)*

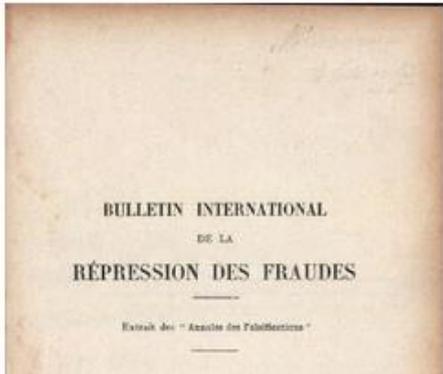
Voir **Annexe01 Description et indexation du Bulletin international de la répression des fraudes (1910-1916) et des Annales des falsifications et des fraudes (1917-1938)**, p.11-13.

Les index suivants ont été retenus :

- ❖ **Auteurs des articles** lorsqu'ils sont signés
- ❖ **Dénomination du produit** utilisé spécifiquement pour les produits du domaine de la vigne et du vin. Le même index est déjà utilisé pour le Bulletin de l'OIV présent sur le portail PANDOR
- ❖ **Noms d'organismes** auxquels sont rattachés les auteurs
- ❖ **Juridiction** renvoyant aux cours d'appel et à leurs arrêts
- ❖ **Pays** pour un article qui concerne un pays particulier
- ❖ **Entrée(s) de la table des matières**, il s'agit d'entrées thématiques

CONVEX\_MSHDijon\_Rapport scientifique

Copies d'écran depuis l'outil de pré-production, visuellement proche du futur portail PANDOR

Cote	BIRF_1915_01_n075
<b>Intitulé de l'unité documentaire</b>	<b>N°75</b>
Date de l'unité documentaire	Janvier 1915
Description matérielle	44 p.
Documents numériques	

Présentation du contenu	<p>La répression des fraudes et la guerre</p> <p><b>Toubeau, Maxime</b> Service de la répression des fraudes (France) p.1-3</p> <p>La répression des fraudes aux colonies. Décret du 14 septembre 1914</p> <p>Anonyme 14 Septembre 1914 p.3-9</p> <p>La répression des fraudes au Maroc</p> <p>Anonyme 14 Octobre - 19 Octobre 1914</p>
-------------------------	---

*Bulletin international de la répression des fraudes, 1915, n°75*

## CONVEX\_MSHDijon\_Rapport scientifique

### Table des matières

#### A

Acétone / Acide benzoïque / Acide carbonique / Acide mucique / Acide nitreux / Acide phosphorique / Acide salicylique / Acide tartrique / Airelle myrtille / Alcaloïdes / Alcool / Aldéhyde / Aldéhyde formique / Aliments du bétail / Aliments vitaminés / Allemagne / Aluminium / Amandes / Amidon / Angleterre / Antimoine / Appellation d'origine / Argent / Arrêté / Arsenic

#### B

Baking powder / Buile de bancoulier / Belgique / Benzène / Beurre de cacao / Beurre de murumuru / Bicarbonate de soude / Bismuth / Blé

#### C

Cacao / Caféine / Calcium / Calcul des probabilités / Carbonates alcalins / Carthame / Cassis / Cerneaux / Charbon / Charcuterie / Chimie physique / Chocolat / Choucroute / Chula / Circulaire / Cire / Citron / Cobalt / Cobolt / cognac / Colloïdes / Confiserie / Conserves / Conventions internationales / Crème / Cuivre / Cumin / Cyanamide

#### D

Décret / Dicyandiamide / Distinctions honorifiques

## *Index thématique annuel de l'année 1925 des Annales des falsifications et des fraudes*

### Traitement technique des documents :

#### ✓ Numérisation de conservation

Le plateforme Archives Documentation Numérisation (ADN) de la MSH de Dijon s'appuie sur les recommandations du Service Interministériel des Archives de France (SIAF), du Centre Informatique National de l'Enseignement Supérieur (CINES) et du TGIR Huma-Num en termes de numérisation, choix des formats numériques en fonction des différentes typologies et résolution (« *Guide méthodologique pour le choix de formats numériques pérenne* » et « *Guide des bonnes pratiques numériques* »). Une numérisation dite de conservation est effectuée, le format de numérisation pour les archives de type textuel ou iconographique est le format Tagged Image File Format (TIFF) sans compression et la résolution est de 400 DPI optique.

#### **Matériel utilisé pour la numérisation**

Scanner de livres I2s COPIBOOK 300 RGB et scanner de livres I2s COPIBOOK ONYX 600

## CONVEX\_MSHDijon\_Rapport scientifique



### *Scanner de livres I2S COPIBOOK ONYX 600*

- ✓ Création d'unités documentaires à la numérisation et en post traitement si nécessaire
- ✓ Contrôle qualité

Un premier contrôle qualité est réalisé sur la totalité des fichiers produits lors de la numérisation de conservation (TIFF). Ce contrôle s'appuie sur une partie de la norme NF Z42-013 (Archivage électronique). Les vérifications suivantes sont effectuées : bonne ouverture des fichiers TIFF, vérification de la qualité de la numérisation en fonction des originaux, vérification du nommage des fichiers, le nombre de pages, les incidents survenus lors de la numérisation (cas très rares).

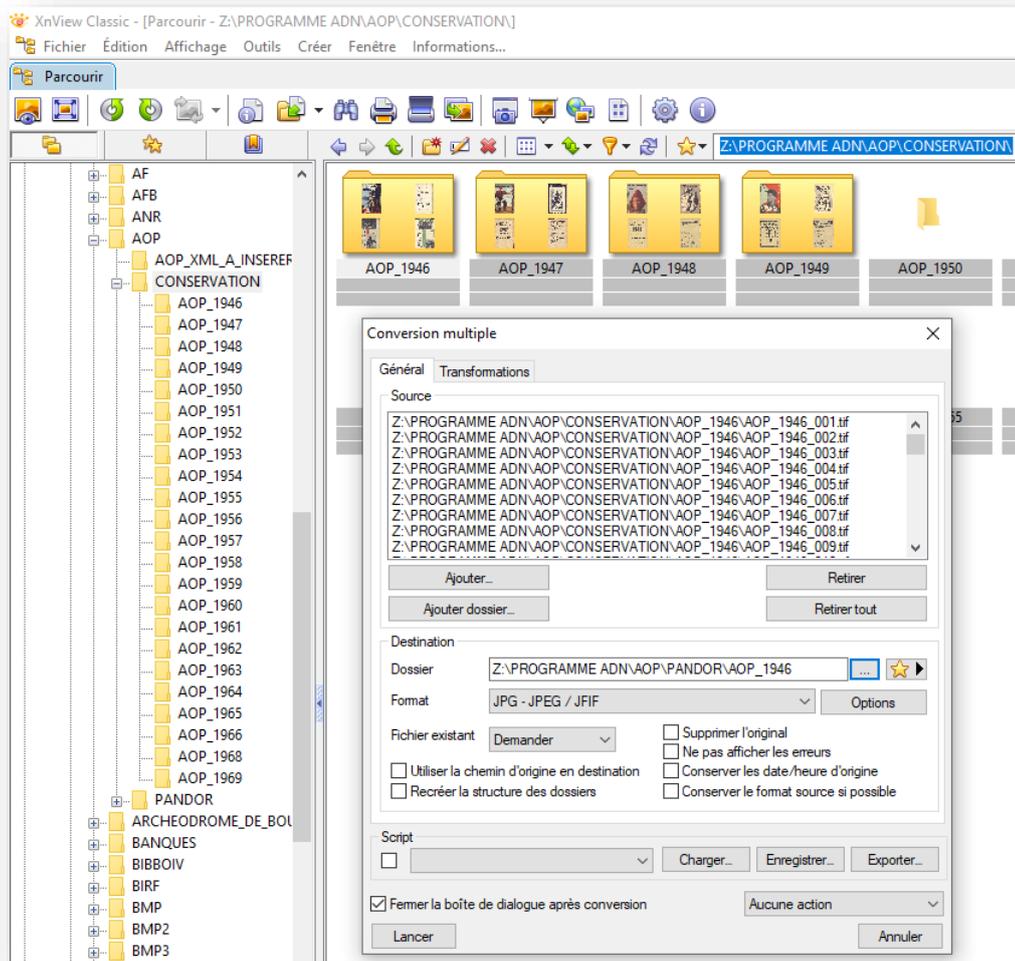
- ✓ Recadrage, redressement

Lors du contrôle des documents, il arrive que nous constatons que des documents ont été numérisés de manière non satisfaisante. Au moyen de la suite Adobe, (Adobe Bridge, camera raw ainsi que Photoshop), un travail de redressement et de recadrage est alors réalisé.

## CONVEX\_MSHDijon\_Rapport scientifique

- ✓ Production de fichiers pour le web

Après validation des fichiers numérisés, la plateforme ADN produit des fichiers dits pour le web. Un traitement par lot est effectué à partir du logiciel gratuit XNVIEW. La solution de publication des fichiers numérisés nous impose le format de fichier JPEG. La résolution est de 150 DPI à 300 DPI.



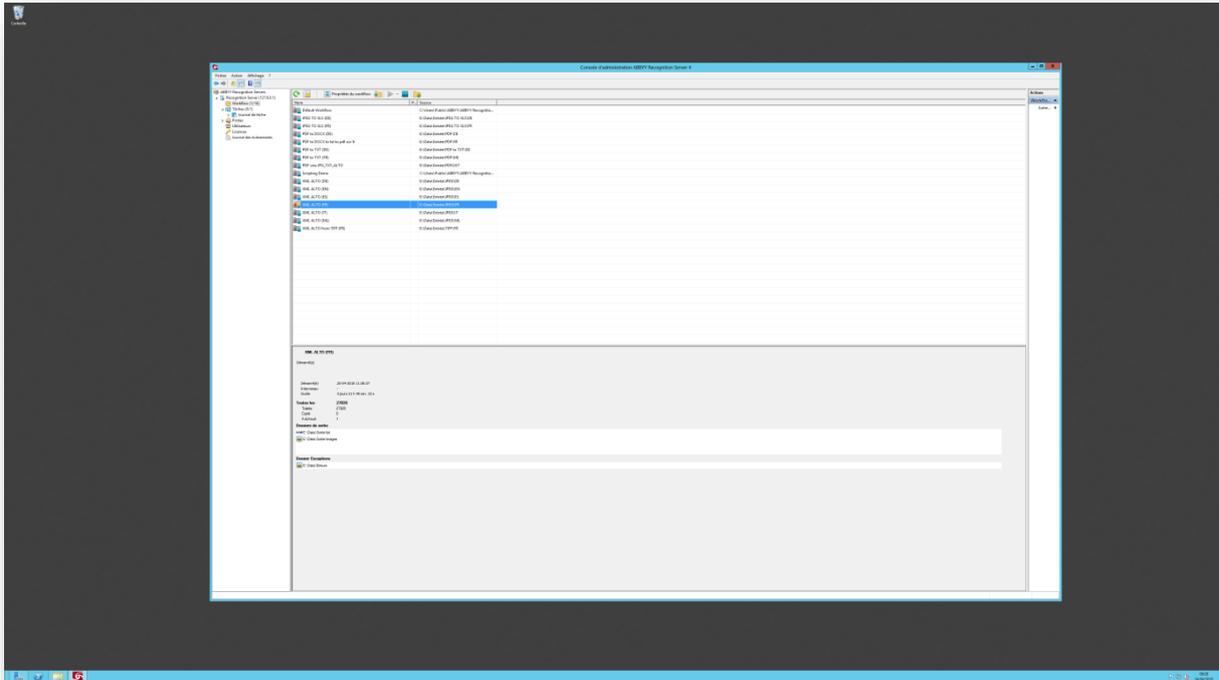
Logiciel XNVIEW V2.49.5

- ✓ Reconnaissance optique de caractères

Une reconnaissance optique de caractères a été réalisée à partir de l'application Abbyy recognition server 4.0. Le logiciel permet de convertir par automatisation des documents au format TIFF, JPEG ou PDF en fichiers numériques interrogeables en plein texte. Il est utilisé dans la chaîne de traitement pour produire des fichiers TXT, des fichiers DOCX, des fichiers PDF/A ou des fichiers XML-ALTO ; ces

## CONVEX\_MSHDijon\_Rapport scientifique

derniers alimentent le portail PANDOR et permettent la recherche en plein texte dans les documents océrisés.



### Logiciel Abbyy recognition server 4.0.

- ✓ Production des fichiers XML-METS

Des fichiers XML-METS sont générés au moyen d'un programme informatique développé en interne en JAVA. Sur le portail PANDOR, ces fichiers servent à l'affichage des résultats après une recherche à l'intérieur des documents océrisés, ils donnent également accès au contenu recherché dans la visionneuse de documents.



### Affichage de résultats depuis l'interrogation METS

- ✓ Stockage des fichiers

## CONVEX\_MSHDijon\_Rapport scientifique

Un stockage quotidien des fichiers de conservation dont la qualité a été jugée satisfaisante est effectuée sur le serveur de la MSH. La même opération est réalisée sur les fichiers compressés pour le web.

- ✓ Sauvegarde des fichiers

Une sauvegarde sur bande magnétique de l'intégralité de notre serveur de stockage est effectuée chaque nuit par le Direction du Numérique (DNUM) de l'université de Bourgogne. A la fin des programmes cofinancés par le GIS CollEx-Persée, une sauvegarde des programmes sera mise en place sur les serveurs de la TGIR HUMA-NUM dans le cadre de la HUMA-NUM BOX.

## Livrables

### Documents numérisés et fichiers pour publication

**Bulletin international de la répression des fraudes** : 93 bulletins numérisés soit 3225 fichiers, production de 3225 fichiers de publication au format JPEG, 3225 fichiers XML-ALTO (OCR pour le portail), 93 fichiers PDF/A (avec OCR).

**Annales des falsifications et des fraudes** : 280 numéros numérisés soit 18264 fichiers, production de 18264 fichiers de publication au format JPEG, 18264 fichiers XML-ALTO (OCR pour le portail), 280 fichiers PDF/A (avec OCR).

### Instruments de recherche encodés au format XML-EAD

Fichiers joints en annexe

Bulletin international de la répression des fraudes :

**CONVEX\_Annexe02\_FRMSH021\_00054.xml**

Annales des falsifications et des fraudes :

**CONVEX\_Annexe03\_FRMSH021\_00056.xml**

### Publication des contenus sur le portail PANDOR de la MSH de Dijon

Les instruments de recherche en XML-EAD et les documents numériques associés ont vocation à être publiés sur le portail de publication PANDOR de la MSH. Ce dernier est actuellement en cours de mise à jour et la nouvelle version de l'outil n'est pas encore mise en production (la mise en œuvre est attendue pour juin 2021).

En attendant, les instruments de recherche ont été publiés sur une application de test et sont fonctionnels.

## Évènements scientifiques

Journée de restitution du programme CONVEX et de lancement du programme CONVEX2 le 24 février 2021 : Enjeux et usages des collections numériques sur la vigne et le vin : sciences, normes et transmissions



CONVEX\_MSHDijon\_Rapport scientifique

Programme joint en annexe : **CONVEX\_Annexe04\_ProgrammeRestitution.pdf**

## Synthèse

25 personnes issues des sciences humaines et sociales comme des sciences expérimentales étaient présentes lors de ce web-séminaire dont le but était en premier lieu de présenter les perspectives de recherches et déjà quelques résultats relatifs à la mise en ligne de la collection des Annales des Falsifications et des Fraudes.

Après une présentation de l'outil de recherche en ligne, les interventions se sont portées sur l'intérêt de cet instrument de recherche dans le cadre de travaux transdisciplinaires sur l'histoire des sciences et des normes du vin. Les résultats phares des premières recherches entreprises montrent, en particulier, l'importance de la revue pour mieux cerner les évolutions en matière de mesure des composés du vin et, au-delà, pour appréhender les constructions et les mutations des goûts des vins et des consommateurs au long du XXe siècle.

Dans un second temps, les chercheurs et ingénieurs présents se sont penchés sur le programme en cours de numérisation et diffusion de la collection du Progrès Agricole et Viticole. Après une présentation de l'outil en cours d'élaboration, les scientifiques de Montpellier SupAgro ont pu éclairer l'assemblée sur l'intérêt bibliographique qu'ils avaient à accéder aux contenus.

Enfin, par des remises en contexte des cadres pédagogiques, scientifiques et institutionnels de genèse et de développement de la revue entre 1884 et 1950, les historiens ont caractérisé les différents sujets qu'il sera désormais possible d'aborder : histoire de l'encépagement dans le monde, évolution des techniques viticoles, histoire des pratiques au vignoble et du rapport à l'environnement, etc.

## Perspectives scientifiques

Le traitement et la diffusion de corpus relevant des thématiques de la vigne et du vin, entamé de longue date à la MSH et soutenu dès 2014 dans le cadre du dispositif BSN5, se poursuit actuellement par le projet CONVEX2, lauréat de CollEx-Persée en 2021.

### Projet CONVEX 2

L'objectif recherché est dans un premier temps de poursuivre la politique documentaire 3.0 initiée par la MSH de Dijon depuis plusieurs années, et ce grâce à la diffusion en accès ouvert et gratuit d'une revue professionnelle ancienne, inédite au format numérique : *Le Progrès agricole et viticole (1884-1945)*. La seconde partie du projet consiste en l'élaboration d'un corpus numérique signalant des publications anciennes, du XVIIIe à 1950, déjà présentes en ligne ou restant à numériser. Le traitement de ces deux corpus, distinct en matière de méthodologie, favorise la mise en place d'approches holistiques et multiscalaires d'un intérêt scientifique indéniable. En outre, l'histoire des sciences et techniques vitivicoles n'est pas la seule à profiter de ce projet puisque ce dernier englobe dans le même souci d'exhaustivité certains aspects juridiques, économiques et socio-culturels inhérents au monde de la vigne et du vin. L'objectif affiché permettra non seulement de mieux appréhender la place de la France dans le monde en tant que pays producteur, mais aussi d'envisager ses interactions avec d'autres nations viticoles au cours des trois derniers siècles. Enfin en se plaçant sous le signe de l'histoire globale, CONVEX 2 ne s'adresse pas uniquement aux chercheurs en sciences humaines et



CONVEX\_MSHDijon\_Rapport scientifique

sociales, mais aussi aux professionnels de la filière vitivinicole ainsi qu'à un public curieux de découvrir un monde en constante mutation.

Ces ressources viendront alimenter les collections relatives à la vigne et au vin diffusées par la MSH de Dijon et labellisées par CollEx en 2017.

### Autres perspectives

Dans le cadre de la Bibliothèque Numérique des fonds conséquents et pour certains inédits (compte rendus RONEO de congrès ou autres publications issues d'organisations professionnelles nationales et internationales) sont en cours de fléchage au sein des bibliothèques de l'**INAO** et de l'**Organisation Internationale de la Vigne et du Vin**.

Afin de numériser et mettre en ligne la collection complète du Progrès Agricole et Viticole (1951-2018) des échanges réguliers sont en cours avec l'éditeur de la revue, le **Groupe international d'Experts en Systèmes vitivinicoles pour la CoOpération** (GiESCO), qui possède les numéros antérieurs à ceux déjà numérisés. Des recherches prospectives communes sont lancées pour parvenir à un financement complet de ce projet qui possède un immense intérêt pour les chercheurs en sciences de la vigne comme, autant que pour les Sciences Humaines et Sociales et la filière de production.

## Annexes

### Annexe 01 : Description et indexation du Bulletin international de la répression des fraudes (1910-1916) et des Annales des falsifications et des fraudes (1917-1938)

Synthèse des échanges ayant eu lieu entre le 10 juillet 2019 et le 31 janvier 2020

### **Points discutés avec les scientifiques**

Titre de l'article (*head*) : chaque titre noté en « gras » sera indexé.

- Auteur (*persname*) : Personne qui est à l'origine de l'écriture de l'article, sauf les ministres et le président de la République. Si pas d'auteur, on note « anonyme » ou « non signé » plus neutre
  - o On conserve la mention Anonyme pour tous les articles non signés, dans un souci de cohérence.
    - Saisie : la mention anonyme en tant qu'item simple, elle n'est pas considérée comme un auteur. Prévoir un traitement du fichier par lot plutôt qu'une saisie de chaque occurrence.
- Institution (*corpname*) : rattachement institutionnel de l'auteur.  
On n'indexe ni les services de l'état ni les collectivités territoriales (Ministères, préfectures)

## CONVEX\_MSHDijon\_Rapport scientifique

On ajoute les noms du pays de toutes les institutions indexées, qu'elles soient situées en France ou à l'étranger

- Pays : On indexe tous les pays
- Date : À la fois des dates précises, mais aussi plus larges (année/mois).  
Il s'agit des dates des textes officiels (arrêtés, circulaires, lois) ou de jugements
- Quand il y a plusieurs dates :
  - o Indiquer toutes les dates dans le titre et indexer un intervalle de dates
- Pagination : page de début – page de fin.
- Juridiction : on indique le lieu du tribunal où est rendu le jugement, pour les tribunaux parisiens si le lieu n'est pas indiqué, indexer Paris, ex : Tribunal de la Seine
- On indexe le lieu de juridiction à la suite de la date
  
- Pour les décrets : la date et le nom du décret sont importants. Pas besoin de préciser le nom du Président de la République et des ministres à l'origine du texte
- Concernant la « juridiction » : On isole le lieu de jugement afin qu'il soit cliquable.
- Pour les laboratoires situés à l'étranger : on mentionne le pays entre parenthèses. On l'indique pour tous

Les scientifiques souhaiteraient, si c'est possible pour ADN, que les noms de vins soient cliquables (en créant des mots-clés). Et notamment pour les appellations viticoles. Voir à partir de quelle liste on travaille et si c'est pertinent de n'offrir ce traitement fin que sur les termes qui concernent le vin pour une revue plus large.

Voir le traitement proposé pour les tables des matières annuelles qui listent les thématiques abordées, voir table des matières de l'année 1910.

Traitement des tables des matières > validé, à poursuivre sur le modèle de l'année 1910

Ce travail d'indexation permet de ne pas traiter la revue comme une ressource exclusivement vigne & vin

Index des dénominations de vins

Lire les articles traitant du vin pour essayer de repérer leurs noms. On n'utilise pas de liste spécifique car la période considérée de la revue concerne surtout ce qui précède les AOC. Si des questions surviennent, interroger les scientifiques. *A noter* : les noms de vins ne prennent pas de majuscules.

Ex : champagne, bourgogne

- Quand il y a plusieurs dates dans le titre, il faut les indiquer toutes dans le titre, et indexer seulement un intervalle.
- Quand il n'y a pas d'auteur, indiquer « anonyme »
- Validé : Indiquer anonyme pour l'ensemble des articles non signés alourdit la saisie et l'affichage à l'écran. Est-ce qu'on conserve ou est-ce qu'on ne renseigne un auteur que pour les articles signés hors auteurs institutionnels de type ministre, préfet, élu ?

Cas d'un article signé de deux auteurs rattachés à deux laboratoires de recherche distincts



CONVEX\_MSHDijon\_Rapport scientifique

Voir :AF\_1927\_578 p.574

On renseigne les éléments dans l'ordre dans lequel ils sont écrits

On considère que le premier auteur cité est rattaché au premier laboratoire cité, etc

Dans Excel, on saisit dans l'ordre suivant qui sera reproduit à l'affichage dans PANDOR

Auteur 1

Labo 1

Auteur 2

Labo 2

### **Travail interne à la MSH relatif au traitement informatique**

- Enlever toutes les balises <item/> : empêchent le logiciel de transformation de fonctionner.
- Pour les auteurs : on garde seulement le nom de celui qui est sous le titre. Pareil avec les corpname .
- Pour les articles nommés « Jurisprudence française relative à la répression des fraudes » et aux autres articles construits de la même manière : on conserve un head pour chaque titre, avec à chaque fois des items. Si le reste du titre n'est pas en gras, il sera tout de même ajouté dans le tableau.
- Pour certains services comme la « Répression des fraudes », et il est important de préciser dans le corpname « service de la répression des fraudes ».
- Concernant les noms des services, un travail sera de nouveau fait à la fin afin de vérifier l'écriture exacte de chacun (le BIRF ou l'AF ont tendance à mettre des majuscules un peu partout).
- S'il est écrit « Service de surveillance, à Bruxelles », il faut indiquer dans corpname « Service de surveillance de Bruxelles (Belgique) ».
- Quand plusieurs pays sont listés en gras, il faut tous les indiquer en geogname.
- Concernant les tables des matières de BIRF : Voir comment adapter le modèle
- Pour les tables des matières de BIRF et AF : on indique cela en head, et on ajoute seulement la pagination (pour AF seulement la première page). Voir si on a le temps de les reproduire intégralement en EAD, ce qui permettrait de créer une indexation matière.

### **Annexe 02 : Instrument de recherche décrivant le Bulletin international de la répression des fraudes**

Instrument de recherche joint : CONVEX\_Annexe02\_FRMSH021\_00054.xml

### **Annexe 03 : Instrument de recherche décrivant les Annales des falsifications et des fraudes**

Instrument de recherche joint : CONVEX\_Annexe03\_FRMSH021\_00056.xml

### **Annexe 04 : Programme de la journée de restitution de CONVEX**

Annexe04\_ProgrammeRestitutionCONVEX.pdf