

# SERVIST : PLATEFORME D'EXPLORATION, DE RECHERCHE ET DE VISUALISATION DE DONNEES MULTIMEDIAS

Valérie BONVALLOT <sup>(\*)</sup>, Philippe HOUDRY <sup>(\*)</sup>, François PARMENTIER <sup>(\*)</sup>, Alain TRAMONTI <sup>(\*)</sup>, Solveig VIDAL <sup>(\*)</sup>  
[valerie.bonvallot@inist.fr](mailto:valerie.bonvallot@inist.fr), [philippe.houdry@inist.fr](mailto:philippe.houdry@inist.fr), [francois.parmentier@inist.fr](mailto:francois.parmentier@inist.fr), [alain.tramonti@inist.fr](mailto:alain.tramonti@inist.fr), [solveig.vidal@inist.fr](mailto:solveig.vidal@inist.fr)

<sup>(\*)</sup> INIST-CNRS

2, Allée du Parc de Brabois

CS 10310

F-54519 Vandœuvre-lès-Nancy.

## Mots clefs :

Veille scientifique et technologique, Gestion des connaissances, Ingénierie documentaire, Innovation, Accès à l'information, Collecte d'informations, Circulation de l'information, Cartographie information

## Keywords :

Scientific and technical observation, Knowledge management, Knowledge engineering, Innovation, Information access, Information gathering, Information circulation, Information mapping

## Palabras clave :

Escudriñar científico y tecnológico, Administración del conocimiento, Innovación, Acceso información, Reunir de información, Circulación información, Cartografía información

## Résumé

L'analyse de quantité importante d'informations de nature hétérogène a conduit le service Veille de l'INIST à développer un Système d'Exploration, de Recherche et de Visualisation de l'Information Scientifique et Technique : Servist. Servist est une plateforme de veille traitant de gros volumes de données multisources, multiformats et multimédias. Répondant à un besoin interne de capitalisation des développements, d'amélioration et de diversification des traitements, Servist est à présent accessible à des utilisateurs extérieurs établissant ainsi une nouvelle relation entre le demandeur et le veilleur.

La première partie de cet article décrit les principales fonctionnalités de Servist, la seconde détaille plusieurs scénarios mettant en évidence les besoins auxquels l'outil peut répondre (exploration de données, analyses statistiques et construction d'indicateurs de production scientifique) et la troisième partie présente les perspectives envisagées (importation de données, accroissement du nombre de sources et de ressources, amélioration de l'interface web et des possibilités d'échanges sécurisés).

# 1 Introduction

Progrès et évolutions des TIC facilitent la production et l'accès à l'information. À cet aspect quantitatif s'ajoutent des caractéristiques qualitatives et techniques : les données factuelles, textuelles et structurées ne sont plus les seules à être traitées ; viennent s'associer les images, les sons, etc. Volume et hétérogénéité des données ainsi que variété des sources à surveiller sont désormais à prendre en considération par les outils d'analyse. En effet, l'abondance d'information disponible devient pertinente quand elle est exploitée, capitalisée afin d'en extraire une réelle plus-value informationnelle qui débouchera sur des orientations et des prises de décisions stratégiques d'ordre scientifique, économique, social et/ou politique.

Pour répondre à ce besoin de traitement de grande masse d'information hétérogène, en un temps réduit, de nombreux logiciels et plateformes ont été développés et mis sur le marché, économique ou académique. En tant qu'utilisateur et producteur/développeur d'outils, au sein de l'INIST-CNRS<sup>1</sup>, le service Veille<sup>2</sup> effectue une veille régulière afin de répondre aux besoins évolutifs de ses clients. Il fait bénéficier, à l'ensemble de la communauté des spécialistes de l'information et des éditeurs, de ses activités en mettant en ligne des fiches descriptives d'outils ainsi que les résultats des tests d'évaluation réalisés dans le cadre du projet de benchmarking<sup>3</sup> [1, 2, 3, 4]. Son domaine de compétence étant l'IST (Information Scientifique et Technique) issue de bases de données bibliographiques, avec notamment la production de PASCAL<sup>4</sup> et FRANCIS<sup>5</sup>, ses travaux se concentrent essentiellement sur les techniques de fouilles de données structurées à des fins de génération d'indicateurs bibliométriques. À ces fonctionnalités de base, appliquées à du document multimédia, se rajoutent les techniques de visualisation auxquelles seront ultérieurement intégrées des ressources terminologiques.

Dès 1999, en collaboration avec le LORIA<sup>6</sup>, l'INIST mettait en place grâce à une ingénierie SGML-XML les serveurs d'investigation [5, 6, 7] remplacés en 2006 par Servist. Ces Systèmes d'Exploration, de Recherche et de Visualisation de l'IST continuent de s'appuyer sur la bibliothèque DILIB<sup>7</sup> [8] associée aux langages modernes de programmation (PHP, XSL, ...).

Le projet Servist repose sur la volonté de prendre en compte des besoins spécifiques, aujourd'hui en partie pris en charge par les différentes solutions à disposition des veilleurs. Si l'on reste dans un budget économiquement comparable, trois possibilités d'utilisation se dégagent :

- outils indépendants d'interrogation, de traitements ou de visualisation de l'information<sup>8</sup> ;
- les outils de veille intégrés mais centrés sur une source d'information<sup>9</sup> ;
- plateformes intégrées et multiformats dont l'offre est importante et certaines étudiées dans le cadre du projet benchmarking (voir ci-dessus).

---

<sup>1</sup> <http://www.inist.fr/>

<sup>2</sup> <http://veille.inist.fr/>

<sup>3</sup> <http://outils.veille.inist.fr/>

<sup>4</sup> Base de données bibliographiques, multidisciplinaire et multilingue qui couvre l'essentiel de la littérature mondiale en Sciences, Technologie et Médecine : <http://ingenierie.inist.fr/rubrique27.html>

<sup>5</sup> Base de données bibliographiques multidisciplinaire qui couvre l'essentiel de la littérature en Sciences Humaines et Sociales : <http://ingenierie.inist.fr/rubrique1.html>

<sup>6</sup> <http://www.loria.fr>

<sup>7</sup> <http://dilib.inist.fr/>

<sup>8</sup> <http://www.visokio.com/>

<sup>9</sup> <http://www.hubmed.org/>

Chacune de ces possibilités présente ses propres limites :

- l'approche indépendante, qui offre parfois des réalisations très abouties, engendre l'inconvénient majeur de ralentir le cycle de veille en obligeant à de multiples reformatages et/ou interfaçage et impliquant la maîtrise de différents logiciels ;
- les solutions d'exploration, de recherche et de visualisation de l'information contenue dans une base de données bibliographique ne permettent l'exploitation que d'une ressource ou, pour certaines d'entre elles, nécessitent des développements complémentaires importants pour pouvoir exploiter les données issues d'une nouvelle base ;
- les plateformes intégrées proposent un niveau élevé de fonctionnalité mais ne traitent pas encore les données multimédias.

Un dernier élément important face à cet ensemble de solutions est la capacité à contrôler les évolutions de la plateforme et la liberté d'exploitation.

Cet article évoquera, dans une première partie, Servist en tant que plateforme d'intégration de compétences et en tant que plateforme collaborative et interactive de fouille de données multimédias. Alors que la première partie décrira les principales fonctionnalités, la seconde détaillera plusieurs scénarios mettant en évidence les besoins auxquels l'outil peut répondre. La troisième partie, quant à elle, se focalisera sur les perspectives envisagées.

## 2 Servist

### 2.1 Pour le service Veille

Le service Veille de l'INIST est composé de huit ingénieurs et rassemble des compétences documentaires, scientifiques et informatiques. Ce service réalise des prestations pour le CNRS, pour des clients, pour des partenaires (autres EPST<sup>10</sup>) ou, en interne, pour la Direction de l'Institut. Il propose ainsi une offre diversifiée allant de la recherche bibliographique au dossier de synthèse en passant par la veille sur internet. Il s'appuie pour cela sur une collecte multisource et multibase. Pour le traitement de l'information, il dispose d'outils commerciaux (plateforme KeyWatch – iScope<sup>11</sup> de veille sur internet, LexiQuest Mine – SPSS<sup>12</sup> pour la cartographie) et, pour des raisons de coûts ou de fonctionnalités, développe ses propres programmes.

Servist s'intègre dans ce schéma de fonctionnement avec trois objectifs :

- capitaliser les développements jusqu'alors réalisés de manière ponctuelle et répartis en fonction du sujet traité ;
- améliorer la qualité des traitements ;
- diversifier les traitements de l'information plus rapidement au travers d'une interface conviviale.

Servist reprend ainsi l'ensemble des développements déjà réalisés mais sous la forme d'une plateforme unique. Les versions successives sont testées avant mise en ligne. Les compétences du service sont donc progressivement intégrées et capitalisées sous forme de développements informatiques aboutissant à la mise à disposition de modules de fonctionnalités.

Le gain de temps obtenu sur les traitements permet soit de se focaliser davantage sur l'analyse soit de traiter plus de dossiers pendant la même période.

Devant le double constat d'une demande client évolutive en fonction des résultats obtenus et d'une interface web à la prise en main aisée, nous avons convenu d'un usage possible par des utilisateurs externes préalablement formés.

---

<sup>10</sup> Établissement Public Scientifique et Technique.

<sup>11</sup> <http://www.iscope.fr/>

<sup>12</sup> [http://www.spss.com/fr/produits\\_solutions/Data\\_Text\\_Mining/Lexiquest\\_Mine.htm](http://www.spss.com/fr/produits_solutions/Data_Text_Mining/Lexiquest_Mine.htm)

## 2.2 Pour nos utilisateurs

Dans un premier temps, l'ouverture de Servist à des utilisateurs « extérieurs » a été faite afin d'accéder en ligne à des corpus bibliographiques via une interface sécurisée<sup>13</sup>. Les premiers modules développés répondent parfaitement à ce besoin en autorisant des requêtes et des recherches.

Les fonctionnalités proposées autorisent à présent un travail plus orienté veille sur les données. Ce travail nécessite une collaboration importante entre le veilleur et le demandeur. En effet, le cycle de veille entraîne une remise en cause permanente de la recherche et donc de l'analyse de l'information. Dans ce cadre, Servist se positionne comme une solution où le demandeur devient lui-même le veilleur pouvant ainsi adapter le travail de veille et de recherche d'information en fonction de ses propres résultats.

Lorsqu'un chercheur dispose d'une grande quantité de données bibliographiques, le plus difficile est d'y trouver les quelques articles qui apporteront le plus d'information pertinente. Servist facilite la découverte de ces informations grâce à une manipulation aisée de ces données (ex : liste des index avec le nombre de termes dans chacun d'eux et le nombre de documents concernés, bibliométrie avancée, etc.).

L'accès à la plateforme est contrôlé, ce qui permet de n'en réserver l'usage qu'à une communauté (par adresse IP, ou par login/mot de passe).

Suivant les compétences documentaires des utilisateurs, il est possible de faire varier la complexité (et aussi la puissance) de son interface. Il suffit pour cela de ne retenir que les modules pertinents pour l'utilisateur. Le système étant développé module par module, ses versions les plus avancées se distinguent souvent par l'ajout de nouvelles possibilités fonctionnelles (modules). Ainsi, lorsqu'un nouveau module se révèle intéressant pour des utilisateurs avancés, il est facile d'ajouter ce module afin qu'ils en profitent. Voici les modules d'ores et déjà développés, dont certains sont encore en phase de test :

- **Exploration**

- **Recherche rapide** : lance une recherche quel que soit l'index (présent sur toutes les pages).
- **Liste des termes d'un index** : visualise le contenu d'un champ en liste de termes, avec le nombre de notices contenant ce terme. On peut ainsi accéder à toutes ces notices.
- **Nuage des termes d'un index** : visualise le contenu d'un champ comme un nuage de termes, le nombre de notices contenant un terme étant affiché au survol de celui-ci par le curseur de la souris. Les termes du nuage sont également combinables en une requête booléenne.
- **Notices** : affiche une liste de notices correspondant à une requête, permet aussi de visualiser chacune de ces notices.
- **Recherche** : lance une recherche dans tout index en laissant le choix du mode (strict, par troncature à droite, par expression régulière – sous-chaîne).
- **Historique** : liste l'historique des requêtes et le nombre de notices trouvées, permet la combinaison deux à deux de chacune de ces requêtes avec les opérateurs ET, OU et SAUF. Tout clic sur un terme est historisé.

- **Calcul**

- **Bibliométrie avancée** : affiche la répartition de termes, avec tout ou partie du corpus, d'un croisement entre deux ou trois index.
- **Cooccurrences** : affiche les cooccurrences entre des termes d'un même index pour tout ou partie du corpus.

- **Visualisation**

- **Croisement** : visualise, à partir du choix de deux requêtes dans le module *Historique*, les volumes correspondant à chaque requête et à leur conjonction (opérateurs ET / SAUF).

---

<sup>13</sup> Voir 4.4.

- **Cartographie** : projette une répartition calculée dans le module *Bibliométrie avancée* sur une carte (par exemple, une mappemonde). En général, il s'agit de données géographiques (les pays de publication ou d'affiliation).
- **Histogramme** : projette une répartition calculée dans le module *Bibliométrie avancée* sur un diagramme en barres.
- **Visualisation des cooccurrences** : projette les cooccurrences calculées dans le module *Cooccurrences* sur un plan, et permet de filtrer les liens ainsi que les nœuds en fonction de leurs poids.

Quelques uns de ces modules seront illustrés dans les scénarios qui vont suivre.

## 2.3 Partenariat veille / utilisateurs

La plateforme Servist donne la possibilité à l'utilisateur de piloter sa recherche avec une plus grande autonomie. La nécessité de mettre en place un partenariat étroit avec le veilleur demeure essentielle, chacun préservant ses compétences.

Étant développeur de Servist, le service Veille peut intégrer des demandes complémentaires dans son plan de développement. Il suffit pour cela d'exprimer les besoins et de s'entendre sur les moyens et les délais de livraison. En général, cela passe par l'amélioration ou la création d'un module.

### 2.3.1 Collecte des données

Les utilisateurs peuvent confier directement leurs fichiers au service Veille ou fournir leurs stratégies de recherche. Les ingénieurs documentalistes de l'INIST-CNRS se chargent alors de la collecte qui peut être multibase. La plateforme Servist peut aujourd'hui charger et traiter des notices bibliographiques en provenance de plusieurs bases de données, ensemble ou séparément au sein d'un même serveur, par exemple PASCAL, FRANCIS, SCI/SSCI<sup>14</sup>, MEDLINE<sup>15</sup> et HAL<sup>16</sup>. En complément, des tables d'enrichissement peuvent accompagner les corpus à traiter. Ces dernières peuvent concerner des données internes aux partenaires ou utilisateurs.

Dans le cas de données multimédias, comme des images ou des vidéos, les métadonnées doivent accompagner le corpus non bibliographique.

### 2.3.2 Prétraitements

Une fois les données collectées, des prétraitements sont nécessaires pour leur permettre d'être complètement exploitables :

- reformatage en XML bien formé ;
- modification de la description documentaire (champs) ;
- traitement sur les données (contenus) ;
- transcodage (jeux de caractères).

<sup>14</sup> SCI/SSCI - Science Citation Index / Social Science Citation Index : <http://scientific.thomson.com/products/sci/> et <http://scientific.thomson.com/products/ssci/>

<sup>15</sup> MEDLINE – Une de plus importantes base de données bibliographique produite par l'U.S. National Library of Medicine (NLM). Elle couvre les sciences de la vie et le domaine biomedical. Elle propose plus de onze millions de notices provenant de plus de 7 000 périodiques de 1965 à aujourd'hui.  
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>

<sup>16</sup> HAL – Hyper Article en Ligne : <http://hal.archives-ouvertes.fr/>

Le format informatique des données doit d'abord être converti en XML bien formé. Cela nécessite quelquefois l'écriture de programmes à façon (en partant de formats type serveur, de formats tabulés ou autres).

La fourniture de la description documentaire des données à traiter est importante, pour connaître le découpage en champs, la sémantique et les contraintes associées. Cela peut passer par une DTD ou un schéma (pour le SGML ou le XML), ou une liste des champs avec leurs intitulés.

L'utilisateur précise les champs qui l'intéressent et sur lesquels il souhaite faire des traitements. Les contenus de ces champs ne sont pas toujours manipulables tels quels et demandent :

- le choix de la granularité de la structure (on peut simplifier une structure complexe, ou au contraire détailler certains champs comme l'extraction du pays de l'affiliation dans un nouveau champ) ;
- l'homogénéisation à l'intérieur d'un champ (ex : écriture des auteurs différente suivant le type de publication ou la base de données, variantes des affiliations, etc.) ;
- l'enrichissement à l'aide de tables extérieures (ex : verbalisation de codes de classement ou de pays d'affiliation, traduction de termes, etc.).

Enfin, il ne faut pas oublier la conversion des jeux de caractères. Certaines bases ont leur propre encodage. Les partenaires doivent obtenir les informations concernant ces jeux de caractères pour permettre leur transcodage en Unicode.

### **3 Exemples de scénarios et du Servist multimédia ENERGIGÉO**

Paramétré selon les besoins exprimés par les utilisateurs finals puis chargé avec les données sélectionnées dans les gisements d'informations adéquats, données structurées éventuellement converties en XML voire enrichies d'informations externes, Servist offre une large gamme de fonctionnalités pour la fouille. Celles-ci, au premier niveau, peuvent s'appuyer sur un moteur de recherche classique associé à des liens hypertextes alors que les explorations approfondies profitent d'approches bibliométriques couplées à des modes de visualisation avancée.

Les quatre scénarios qui suivent vont illustrer comment Servist peut répondre à des besoins classiques de veille :

- exploration de premier niveau (comptages) ;
- analyse statistique croisant deux champs, ici dans un contexte plutôt bibliothéconomique (gestion de périodiques) ;
- construction d'un indicateur de production scientifique avec visualisation géographique des résultats ;
- analyse statistique croisant trois champs (une particularité de Servist).

Ces scénarios ont été construits à partir du Servist de démonstration, accessible à partir de l'url suivante : <http://vsst2007.veille.inist.fr> (codes d'accès *vsst*, en minuscules, tant pour le login que pour le mot de passe). Un Servist avec données multimédias est consultable à : <http://energigeo.veille.inist.fr> (accès public). Ce dernier a été généré avec la version 2 de Servist qui, à ce stade des développements, n'intègre pas encore toutes les fonctionnalités de la version précédente mais qui respecte les standards du web 2.0. Ces serveurs sont optimisés pour Mozilla Firefox.

Le premier serveur est consacré aux organismes génétiquement modifiés (OGM). Ses notices bibliographiques, extraites de la base de données PASCAL de l'INIST, sont au nombre de 6 346 et couvrent la période 2004 à 2007 (années de production de la base). Ces notices ont été arbitrairement enrichies, pour l'élément « Périodique », d'un indicateur de fourniture de copies de document « Classe FDP ».

Le second serveur, ENERGIGÉO, a été constitué pour le 18<sup>ème</sup> Festival International de Géographie de Saint-Dié-des-Vosges (04/10-07/10/2007) sur le thème « la planète en mal d'énergies ». Il contient des notices bibliographiques extraites de la base de données FRANCIS de l'INIST, au nombre de 577 sur la période 1992 à 2007 (années de production de la base) ainsi qu'une sélection de 120 photographies accompagnées de leurs métadonnées.

### 3.1 Scénario n°1 : Exploration de premier niveau

Un auteur cherche une revue représentative de son domaine pour publier un article. Servist lui permet de consulter directement, à l'aide du tableau des index dans sa partie gauche de l'écran, l'index « Périodique ».

Par défaut, le classement lié au nombre de notices bibliographiques dans la liste est par ordre décroissant. Les revues, dans lesquelles le nombre de publications du domaine des OGM est le plus élevé, sont donc directement visibles en un clic sur « Périodique », comme le montre la *figure 1*.

The screenshot shows the VSST 2007 search interface. At the top, it says 'VSST 2007 : Thème OGM'. There is a search bar with 'Auteur' selected in the dropdown menu. Below the search bar are buttons for 'Liste de Notices' (selected) and 'Liste de descripteurs'. On the left side, there are two tables: 'Bibliométrie' and 'Descripteurs'. The 'Bibliométrie' table shows counts for various categories like 'Auteur', 'Affiliation', etc. The 'Descripteurs' table shows counts for 'Mot-clé français', 'Mot-clé anglais', etc. In the center, there is a navigation bar with buttons for 'Index', 'Notices', 'Recherche', 'Historique', 'Bibliométrie', 'Croisement', 'Carte', and 'Histo.'. Below this, there are options for 'Sans filtrage' (selected) and 'Avec filtrage'. On the right, there are buttons for 'Export Excel' and 'Imprimer'. At the bottom, there is a list of journals sorted by 'Nb docs' in descending order, with 'Périodique' selected in the dropdown menu. The list includes 'Plant physiology : (Bethesda)', 'Plant journal', 'Transgenic research', etc.

Bibliométrie		
Nom	Nb d'items	Nb docs
Auteur	28306	6322
Affiliation	12984	6201
Pays d'affiliation	100	6205
Année	10	6346
Périodique	812	6273
Classe FDP	4	5996

Descripteurs		
Nom	Nb d'items	Nb docs
Mot-clé français	9664	6337
Mot-clé anglais	8440	6336
Mot du titre français	710	196
Mot du titre anglais	11412	6241

Périodique	Nb docs
Plant physiology : (Bethesda)	220
Plant journal	193
Transgenic research	174
Plant science : (Limerick)	171
Plant cell reports : (Print)	145
Planta	120
Applied and environmental microbiology : (Print)	118
Cancer research : (Baltimore)	104
Journal of agricultural and food chemistry : (Print)	101
Blood	97

Figure 1

### 3.2 Scénario n°2 : Analyse statistique croisant deux champs

Dans le cadre d'une campagne de réabonnement, le responsable d'une bibliothèque ou le gestionnaire de la couverture scientifique d'une base de données cherche à savoir quels sont les périodiques les plus demandés en fourniture de document (FDP) qui doivent absolument être pris en compte. Il lui faut donc créer un état statistique croisant les noms des périodiques et le nombre de fourniture de documents associé. Pour ce scénario, on considérera un classement des périodiques du corpus dans 4 classes pour décrire la fourniture de documents selon une approche de type « Boxplot »<sup>17</sup> [9] : Extrême, Forte, Moyenne et Faible<sup>18</sup>.

Outre faire une collecte d'informations appropriée, pour chaque notice bibliographique, le corpus doit contenir la valeur de la classe de fourniture de documents associée au périodique. Avec des notices bibliographiques PASCAL qui ne contiennent pas cette information, il a fallu procéder à un enrichissement par emploi de données externes. Avant chargement dans Servist, une table d'enrichissement a été constituée en prenant comme clé de rapprochement le nom des périodiques.

L'état statistique complet s'obtient simplement grâce à l'onglet « Bibliométrie ». Les deux index à croiser sont « Classe FDP » et « Périodique ». Le résultat obtenu présente un tri des périodiques triés par classe de fourniture de document, comme l'illustre la *figure 2*.

Le même type d'analyse pourrait être mené à partir d'un autre type d'indicateur évaluant la qualité scientifique des périodiques, par exemple le facteur d'impact<sup>19</sup>.

### 3.3 Scénario n°3 : Production d'indicateurs et visualisation géographique

On cherche à savoir quels sont les pays qui coopèrent ces dernières années avec la Chine sur le thème des OGM, et à évaluer ces collaborations par le nombre d'articles publiés en commun (entre équipes de recherche). Servist permet ce croisement d'informations, comme on a pu le voir plus haut, mais il propose, en plus une visualisation cartographique à l'aide d'une mappemonde (avec échelle de couleurs).

Avec le Servist de démonstration, il suffit d'interroger sur Chine dans l'index « Pays d'affiliation » puis de demander dans l'onglet « Bibliométrie » d'abord « Nouvelle Bibliométrie » puis le tri par « Pays d'affiliation » sur la dernière requête. Une fois le tableau des résultats obtenu, il faut alors cliquer sur « Visualisation cartographique » afin d'obtenir la carte de la *figure 3*.

On constate aisément que, ces dernières années et autour des OGM, ce sont principalement les pays dits occidentaux (notamment les États-Unis) et émergents, comme l'Inde, qui coopèrent avec la Chine.

---

<sup>17</sup> Ou boîte à moustaches : <http://www.sfds.asso.fr/groupe/statvotre/Boite-a-moustaches.pdf>

<sup>18</sup> Données fictives dans le cadre de cet exemple

<sup>19</sup> Facteur d'impact : <http://scientific.thomson.com/free/essays/journalcitationreports/impactfactor/>

VSST 2007 : Thème OGM

Recherche

Auteur  Liste de Notices  Liste de descripteurs

[Index](#)
[Notices](#)
[Recherche](#)
[Historique](#)
[Bibliométrie](#)
[Croisement](#)
[Carte](#)
[Histo.](#)


**Nouvelle bibliométrie**

**Voir l'histogramme en barres (Classe FDP)**

Répartition par Classe FDP (4) ▲ - +

**Faible : 2207**

Répartition par Périodique (189) ▲

- Plant physiology : (Bethesda) : 220
- Plant journal : 193
- Transgenic research : 174
- Plant science : (Limerick) : 171
- Plant cell reports : (Print) : 145
- Planta : 120
- Plant cell, tissue and organ culture : 71
- The American journal of pathology : 55
- Trends in biotechnology : (Regular ed.) : 50
- Neuroscience : 46
- Vaccine : 46
- Phytochemistry : 40
- Plant physiology and biochemistry : (Paris) : 38
- Science : (Washington, D.C.) : 37
- Neuroscience letters : 36
- Plant, cell and environment : (Print) : 33
- Pest management science : 31
- Plant and soil : 31
- Trends in plant science : 31
- Theoretical and Applied Genetics : 28
- Physiological and molecular plant pathology : 23
- The Journal of nutrition : 21
- Neurobiology of aging : 19
- Scientia horticultrae : 18
- Weed technology : 18
- New phytologist : 16
- Transplantation : 16
- Neuroreport : (Oxford) : 14
- Neuropeptides : (Edinburgh) : 13
- Plant disease : 12

Figure 2



6346 notices

**Bibliométrie**

Nom	Nb d'items	Nb docs
<b>Auteur</b>	28306	6322
<b>Affiliation</b>	12984	6201
<b>Pays d'affiliation</b>	100	6205
<b>Année</b>	10	6346
<b>Périodique</b>	812	6273
<b>Classe FDP</b>	4	5996

**Descripteurs**

Nom	Nb d'items	Nb docs
<b>Mot-clé français</b>	9664	6337
<b>Mot-clé anglais</b>	8440	6336
<b>Mot du titre français</b>	710	196
<b>Mot du titre anglais</b>	11412	6241



Guide de navigation

Recherche

 Auteur

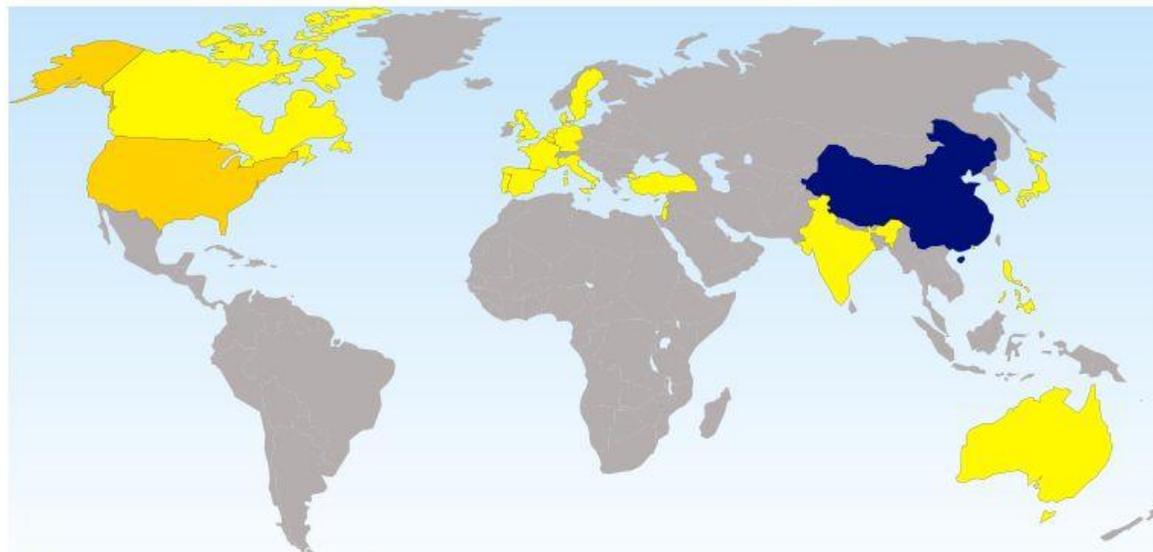
Liste de Notices

Liste de descripteurs

OK

- Index
- Notices
- Recherche
- Historique
- Bibliométrie
- Croisement
- Carte
- Histo.
- 

**Visualisation Géographique / COUNTRY: "Chine" (373 notices)**



Chine : 373

Figure 3

### 3.4 Scénario n°4 : Analyse statistique croisant trois champs

Toujours dans le contexte des OGM, on cherche à savoir quelles sont les habitudes en matière de publication scientifique des chercheurs selon leur pays avec, en plus, l'évolution de celles-ci dans le temps.

Servist, à l'aide de l'onglet « Bibliométrie », permet d'obtenir la réponse à cette question déjà assez complexe. Dans cet onglet, il faut d'abord cliquer sur « Nouvelle bibliométrie ». Puis les trois index à croiser sont, et dans cet ordre, « Pays d'affiliation », « Périodique » puis « Année » (il s'agit de l'année de publication), par exemple pour « Tout le corpus ».

On peut ainsi voir, comme le montre la *figure 4*, que les chercheurs affiliés aux États-Unis semblent moins publier dans la revue « Plant Physiology (Bethesda) » en 2003, 2006 et 2007. Ces résultats, comme ils devraient toujours l'être, doivent être pondérés par une bonne connaissance de la couverture des gisements d'informations utilisés.

De la même manière, Servist permet de voir quels sont les auteurs ou les organismes (index « Auteur » ou « Affiliation ») qui travaillent dans tel ou tel domaine (index « Mot-clé français » ou « Mot-clé anglais »), avec l'évolution de ces travaux thématiques dans le temps (index « Année »).

### 3.5 Servist multimédia ENERGIGÉO

A l'occasion du 18<sup>e</sup> Festival International de Géographie de Saint-Dié-des-Vosges, un Servist a été créé sur le thème « la planète en mal d'énergies ». Il est librement consultable à partir de l'url suivante : <http://energigeo.veille.inist.fr/>

La particularité de ce serveur est de traiter des données multimédias. En effet, ce Servist mélange des notices bibliographiques (extraites de la base de Géographie dans FRANCIS) à une collection d'images numériques provenant de la photothèque du laboratoire PRODIG-CNRS (détenteur des droits). Ces dernières ont été chargées dans le serveur accompagnées de leurs métadonnées descriptives en XML avec un vocabulaire d'indexation de type FRANCIS. Il a ainsi été possible de créer des index communs et homogènes avec les descripteurs du corpus bibliographique, tant en français qu'en anglais.

La *figure 5* illustre l'intégration des images au sein de ce Servist, à l'aide de leurs métadonnées descriptives (ici titres et auteurs/photographes).

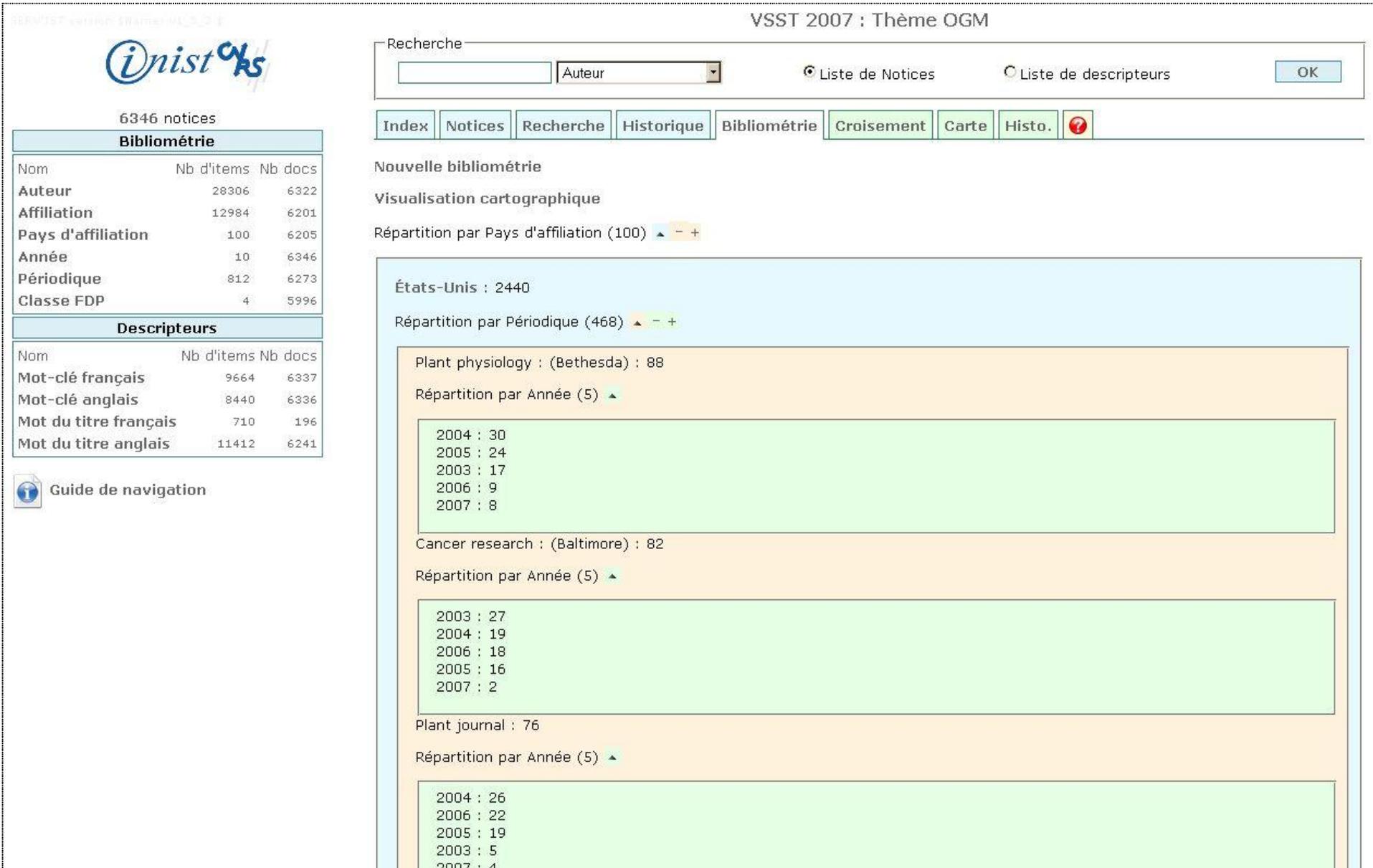


Figure 4

Liste des index

697 notices

catalog

Nom	Occurrences	Notices
auteur	906	685
type	2	697
stype	8	697
annee	21	697
journal	204	547
affiliation	40	41

desc

Nom	Occurrences	Notices
mcfr	1612	697
mcan	1475	697

Notices

Page N°3 / 8

Page n°  / 8 [Atteindre](#)

Requête : annee: "2005" (38 Notices)

Score	Auteurs	Titre
★★★★	 EUZEN (A.) ;	Brésil. Vue du barrage d'Itaipú, construit sur le Paraná
★★★★	 EUZEN (A.) ;	Brésil. Vue panoramique du barrage d'Itaipú, construit sur le Paraná
★★★★	 EUZEN (A.) ;	Brésil. Paraná, transport de l'électricité fournie par le barrage d'Itaipú
★★★★	BUCKLEY (G.L.) ; BAIN (N.R.) ; SWAN (D.L.) ;	When the lights go out in Cheshire
★★★★	SCHAAL (P.) ; KOLSHORN (J.) ;	Windenergie quo vadis ? Entwicklungspotenziale der Windenergie auf Binnenlandstandorten in Niedersachsen und Sachsen-Anhalt

Figure 5

## 4 Perspectives

Le projet Servist, présente de nombreuses perspectives. Sa modularité permet d'envisager une multiplication des développements spécifiques tout en garantissant à l'utilisateur la mise à disposition des fonctionnalités dont il aura réellement besoin.

Les perspectives de ce projet peuvent être segmentées en quatre axes :

- importer des données ;
- augmenter le nombre de sources et ressources disponibles ;
- améliorer l'interface web ;
- accroître les possibilités d'échange sécurisé entre utilisateurs.

Ces perspectives, décrites ci-après, s'inscrivent dans la continuité de la première phase des développements. Elles ne pourront être implémentées sur la base des seules ressources affectées par l'INIST et devront s'appuyer sur l'intérêt de partenaires ou sur l'obtention de financements spécifiques.

### 4.1 Importer des données

Les Servist sont générés après les opérations de collecte effectuées par le service Veille. Le temps de mise en place de l'interface est alors limité à quelques heures. Cependant, dans les projets où les données sont susceptibles de varier ou d'être mises à jour régulièrement, ce délai peut être un frein. Il convient aussi de signaler que, succès aidant, le nombre de serveurs à maintenir est devenu important. Il faut donc que les utilisateurs aient la possibilité d'intégrer eux-mêmes leurs données et que les prétraitements soient aisément utilisables à partir de l'interface. La mise en place d'un module d'importation est donc une priorité.

### 4.2 Augmenter le nombre de sources et ressources disponibles

Des Servist peuvent aujourd'hui être générés à partir des principales bases de données bibliographiques (SCI/SSCI, MEDLINE, PASCAL/FRANCIS, etc.), à partir de fichiers Excel ou texte structuré, la limite étant fixée par les possibilités de conversion au format XML. Servist peut aussi être utilisé avec des données multimédias (images, sons, vidéos). Les données hétérogènes peuvent ensuite être combinées au sein d'un même serveur.

D'autres ressources peuvent venir compléter le dispositif de collecte :

- ressources terminologiques (par exemple via l'API de TermSciences<sup>20</sup> pour une aide à la navigation ou à la saisie des requêtes) ;
- moteurs web (par exemple Google Scholar) ;
- lien vers le texte intégral d'un article ou vers sa commande, ...

---

<sup>20</sup> TermSciences, portail terminologique initié par la Direction de l'Information Scientifique du CNRS et développé par l'INIST en association avec le LORIA et l'ATILF a pour but de valoriser et de mutualiser les ressources terminologiques (lexiques, dictionnaires, thesaurus) des organismes publics de recherche et d'enseignement supérieur pour aboutir à la constitution d'un référentiel terminologique commun. <http://www.termsciences.fr/>

### **4.3 Améliorer l'interface web**

L'objectif principal est d'améliorer les fonctionnalités actuelles en fonction des retours de nos utilisateurs qui, pour certains, financent des développements spécifiques à leurs besoins en acceptant qu'ils soient ensuite mis à disposition des autres utilisateurs.

Un point important concerne l'amélioration du moteur de recherche. Le basculement de la gestion des index à l'aide de la plateforme DILIB vers une solution MySQL est en cours pour la version 2 de Servist. Cela va apporter de nouvelles fonctionnalités, plus proches du Web 2.0.

Les différentes étapes de travail sur les données (recherche – statistiques – visualisation) doivent être intégrées de manière plus dynamique, permettant au veilleur de rebondir en permanence sur de nouvelles requêtes (notion de fouille).

Le nombre de fonctionnalités, et donc de modules, augmentant, il convient de revenir sur des problématiques d'ergonomie. Cette nouvelle étape intégrera les notions d'accessibilité et de personnalisation.

### **4.4 Accroître les possibilités d'échange sécurisé entre utilisateurs**

La plateforme Servist est orientée vers un travail de veille pour des groupes d'utilisateurs-veilleurs. Ceux-ci doivent donc être en mesure d'apporter de l'information par un taggage des données, par le partage de scénarios de traitements des données ou d'analyse. Le lien avec un Content Management System de type SPIP pourrait être envisagé.

Les Servist sont hébergés sur les serveurs de l'INIST. Actuellement, l'identification des ayants droit se fait par login/mot de passe ou par adresse IP. Le cryptage des échanges pourrait accroître le niveau de sécurisation. Des améliorations doivent être apportées pour faciliter la portabilité du système afin de permettre l'hébergement des données sur les serveurs de nos utilisateurs s'ils le souhaitent.

## **5 Conclusion**

Les veilleurs doivent aujourd'hui analyser toujours plus de documents de nature hétérogène provenant de sources diversifiées. Les outils à disposition ne répondent qu'en partie à leurs besoins. C'est pourquoi le service Veille de l'INIST a entrepris le développement d'une plateforme de veille traitant de gros volumes de données multisources, multiformats et multimédias. D'abord réservée à un usage interne, cette plateforme s'ouvre à présent à des utilisateurs externes.

L'approche modulaire de l'interface permet de leur proposer les seules fonctionnalités dont ils ont besoin dans le cadre de leur projet. Le service Veille de l'INIST a pour ambition de continuer les développements, via de nouveaux partenariats et la recherche de financements spécifiques, afin de proposer de nouvelles fonctionnalités avec une meilleure ergonomie et de plus grandes facilités de partage de l'information.

## 6 Bibliographie

- [1] BRUGIDOU M., ESCOFFIER C., FOLCH H., LAHLOU S., LE ROUX D., MORIN-ANDRÉANI P., PIAT G., Les facteurs de choix et d'utilisation de logiciels d'analyse de données textuelles, Journées d'Analyse des Données Textuelles (JADT 2000), Lausanne, mars 2000. [http://solcidsp.upmf-grenoble.fr/cidsp/publications/articles/brugidou\\_choix\\_logiciels.htm](http://solcidsp.upmf-grenoble.fr/cidsp/publications/articles/brugidou_choix_logiciels.htm)
- [2] PERRIN C., Logiciels : le bon choix pour veiller, Archimag, p 42-45, 2000.
- [3] QUATRRAIN Y., PERADOTTO A., NUGIER S., Évaluation d'outils de Text Mining dans un contexte industriel, Colloque International sur la Fouille de Texte (CIFT'04), p 103-115, La Rochelle, 2004. [http://archivesic.ccsd.cnrs.fr/sic\\_00001256.en.html](http://archivesic.ccsd.cnrs.fr/sic_00001256.en.html)
- [4] GARNIER A., Les fonctions émergentes des outils de veille, Bases, p 1-5, 2005.
- [5] DUCLOY J., ARNAUD I., CZYSZ C., HOUDRY P., MAZUR F. et MOREL-PAIR C., Serveurs d'investigation pour la veille scientifique et technologique à l'INIST, Veille stratégique scientifique et technologique, VSST 2001, p 359-366.
- [6] VIDAL S., DUCLOY J., HOUDRY P., Mining medical data using multiple corpora interaction: the Transcriptomics investigation server experiment, Systemics Cybernetics and Informatics, Orlando (USA), 2003. [http://dilib.inist.fr/documents/2003/sci/SCI\\_2003\\_S113GI.pdf](http://dilib.inist.fr/documents/2003/sci/SCI_2003_S113GI.pdf)
- [7] VIDAL S., L'approche diachronique dans la détection de signaux faibles : Une illustration dans le bioterrorisme. Une comparaison entre deux outils de veille, Mémoire DEA SIC, Universités Nancy 2 / Metz, 2004. [http://dilib.inist.fr/documents/2004/rapports/Rapport\\_DEA\\_Solveig\\_VIDAL.pdf](http://dilib.inist.fr/documents/2004/rapports/Rapport_DEA_Solveig_VIDAL.pdf)
- [8] DUCLOY J., DILIB, une plate-forme XML pour la génération de serveurs WWW et la veille scientifique et technique, Le Microbulletin du CNRS, mars 1999. <http://dilib.inist.fr/documents/1999/lmb/article.1.pdf>
- [9] MAGRI M.H., SOLARI A., RERAT K., Les périodiques scientifiques d'audience internationale au travers du Journal Citation Reports : analyse du système d'évaluation de l'ISI. Application à l'étude de la production de l'INRA, Colloque INRA : L'information scientifique et technique : nouveaux enjeux documentaires et éditoriaux, Tours, Octobre 1996. [http://www.inra.fr/internet/Unites/UCD\\_Jouy/scientom/magri\\_fr.htm](http://www.inra.fr/internet/Unites/UCD_Jouy/scientom/magri_fr.htm)