



Les études institutionnelles : périmètres, nomenclatures « à la carte »

Françoise ROJOUAN

Responsable du projet IPERU à l'Observatoire des Sciences et des Techniques (OST)

Cette présentation vise à montrer comment l'OST peut répondre à un besoin qu'expriment les établissements, celui d'études plus individualisées que celles qu'ils reçoivent dans le cadre du projet IPERU et tenant mieux compte de leurs besoins spécifiques. Je vais donc essayer de vous montrer comment grâce à sa base, l'OST peut répondre par des études que l'on peut dire « à la carte » à tout niveau institutionnel dans des contrats spécifiques.

La base de données de l'OST est construite à partir d'une vingtaine de corpus sur différents aspects des activités de R&D. Les données sont acquises auprès de fournisseurs extérieurs de référence tels que l'OCDE, Eurostat, le ministère de l'Enseignement supérieur et de la Recherche, Thomson Reuters ou les offices de brevets. Le travail de l'OST consiste ensuite à mettre en cohérence ces corpus, au travers par exemple de nomenclatures géographiques et thématiques qui s'appuient sur des standards internationaux. Cette base de données permet des études à l'échelle macroscopique (l'espace mondial, les continents ou l'agrégation de différents espaces nationaux...), mais aussi à l'échelle microscopique comme celle des établissements.

Si l'on prend l'exemple du corpus bibliographique de la base de l'OST, il est structuré comme les autres corpus selon trois dimensions (P3 du diaporama associé) :

- Une dimension géographique et géopolitique dont la totalité représente le monde et qui peut ensuite être divisée en pays, régions, départements, etc. C'est aussi sur cette dimension que sont repérés, grâce à leurs adresses, les établissements. Ils ne s'inscrivent pas forcément dans une réalité géographique standard car un établissement peut par exemple, avoir des implantations dans plusieurs régions ;
- Une dimension temporelle car la base de l'OST a des données « biblio » qui débutent en 1999. Cela permet de s'intéresser lors d'une étude à des évolutions annuelles mais aussi à des périodes de temps ;
- Une dimension thématique qui est construite à partir des spécialités des journaux dans lesquels se trouvent les publications de la base Thomson Reuters.

Dans cette base, la production d'un établissement du projet IPERU est représentée sur le graphique par le quadrilatère vert. Il est repéré sur les adresses des publications d'un certain nombre d'années auxquelles on applique des nomenclatures thématiques standard construites par agrégation de spécialités des publications. Ce qui fait la force de cette base c'est que l'établissement identifié ici en vert peut être positionné dans un espace plus large comme par exemple sa région, son pays ou le monde.



Ce qui est aussi très important c'est que pour des études « à façon » les trois dimensions de la base sont modulables à partir de l'élément de base, la publication. L'approche « micro » peut donc être flexible.

Si l'on s'intéresse à la dimension thématique, comment construit-on une nomenclature « à la carte » ? Les nomenclatures standard sont construites à partir de l'agrégation de spécialités dans lesquelles sont classés les journaux et leurs publications. Rien n'empêche de revenir à la publication pour s'intéresser à une thématique et construire son périmètre propre de journaux. Ce travail se fait avec des experts scientifiques de la thématique. Ainsi, la construction d'une thématique peut se faire à partir :

- d'un « panier de journaux » spécifiques de la thématique. On regroupe des journaux qui tous ensemble vont représenter le périmètre de référence de la thématique ;
- d'une construction de spécialités. On regroupe des spécialités qui ne sont pas assemblées de cette façon dans les nomenclatures standard ;
- d'une analyse textuelle des résumés des publications. On repère les journaux du périmètre à partir du contenu de leurs publications.

Sur la dimension thématique « à la carte », l'OST a par exemple mené les études suivantes :

- *Recherche et analyse bibliographique des publications scientifiques dans le domaine E-santé* (2010) ;
- *Constitution d'un corpus sciento-technométrique pour l'étude des pratiques scientifiques et techniques des chercheurs en génomique* (2009) ;
- *Positionnement international de la recherche dans le domaine de la neuroendocrinologie* (2009) ;
- *Etude pour l'aide à la décision en matière de coopération scientifique internationale : le cas des STIC* (2008) ;
- *Positionnement international de la recherche française dans le domaine du handicap* (2008).

Si l'on s'intéresse à la dimension géographique et institutionnelle : les regroupements géographiques ou géopolitiques standard sont construits par l'attribution de codes de la nomenclature internationale des unités territoriales statistiques (NUTS) à chaque adresse. Dans les différents projets récurrents de l'OST, les établissements, voire les unités de recherche, qui effectuent leur propre repérage dans la base de l'OST se reconnaissent à partir du nom du laboratoire et/ou de l'établissement dans l'adresse de la publication. La construction d'un périmètre géo-institutionnel spécifique « à la carte » est donc possible par l'agrégation soit d'unités géographiques de base comme le code postal (les établissements d'un territoire), soit encore par agrégation de noms de laboratoires et/ou d'établissements (le périmètre géo/institutionnel d'un PRES, un RTRA, etc.). Récemment, l'OST a finalisé les études suivantes :

- *Identification de domaines scientifiques d'excellence du « site de Strasbourg » mesurés par les publications en sciences de la matière et de la vie* (2010) ;



- *Positionnement scientifique du PRES de l'Université de Lorraine en sciences de la matière et de la vie* (2010) ;
- *Indicateurs pour un tableau de bord des activités scientifiques des régions Pays-de-la-Loire et Bretagne* (2010) ;
- *Indicateurs pour un tableau de bord scientifique du Réseau thématique de recherche avancée « Recherche agroalimentaire et développement durable »* (2009).

Si l'on s'intéresse à la dimension temporelle, il est possible de faire des études en évolution sur les années individuelles et cela depuis 1999. Il est aussi possible de faire des études sur ce qu'on appelle des années lissées (moyenne triennale glissante) qui peuvent être intéressantes au niveau institutionnel car elles estompent les variations annuelles non significatives de la base, assurent une meilleure robustesse des résultats et permettent de produire des estimations.

Ainsi, toutes les combinaisons sont possibles : pour les établissements d'un territoire, pour un regroupement d'établissements ou une partie d'établissement (département scientifique, par exemple), sur une thématique donnée par des paniers de publications, de journaux, de spécialités ou par nomenclature standard, sur une période ou par groupe de période, etc. Il est également envisageable de réaliser une étude comparative entre deux voire plusieurs établissements sur la base d'une volonté commune.

En conclusion, la base de données est flexible dans toutes ses dimensions dès l'élément de base et les périmètres aux quels on s'intéresse sont toujours positionnables dans des dimensions plus larges, nationales ou mondiales par exemple. Au-delà du rapport IPERU dont les choix de périmètres en particulier thématiques peuvent sembler rigides, il est possible de travailler au cas par cas, de manière très ciblée.

Ghislaine FILLIATREAU

Directrice de l'Observatoire des Sciences et des Techniques (OST)

La base de données peut en effet être utilisée de manière ciblée, car l'on peut se focaliser sur un point de vue et à partir de là construire des périmètres « à façon ». Il est également possible d'aller un cran plus loin, en travaillant sur le contenu scientifique des articles, d'après les mots que l'on trouve dans les abstracts, les titres par exemple. Une base de données aussi riche permet une exploitation libre et puissante des informations. Si le travail courant et le plus visible de l'OST en lien avec ses missions peut paraître très cadré, la base n'en demeure pas moins un espace d'information riche et ouvert sur les contenus scientifiques. Par ailleurs, un autre point sur lequel je souhaite revenir, c'est que la base de données ne prend qu'une fraction de la production des établissements et des chercheurs. Nous ne nous inscrivons donc pas dans une logique de rapport d'activité et d'évaluation scientifique de la fonction de production, mais dans une logique de compte rendu des mouvements de la recherche, focalisé sur une partie seulement des publications scientifiques mondiales. On a un point de vue particulier lorsque l'on travaille avec ce type de bases qui servent à positionner les différents acteurs dans un espace international et disciplinaire. La



bibliométrie a une entrée par la recherche et le contenu scientifique des articles et pas d'entrée par les chercheurs.

Q : Dans certains cas, je ne comprends pas la valeur ajoutée qu'apporte l'OST. Par exemple si je m'intéresse au positionnement de la Bourgogne en France en optique, une recherche sur web of science en ligne me permet d'obtenir les informations recherchées sur la bourgogne et sur la France et je fait le rapport. D'ou ma question, les études de l'OST exploitent-elles les repérages des établissements ?

R : En ce qui concerne la recherche thématique, lorsque vous êtes sur le WoS en ligne, votre démarche est différente de la notre : l'interrogation que vous menez s'effectue par mot-clé que vous pensez opportuns, vous pouvez vous débarrasser du « bruit » que cela occasionne, mais vous ne savez pas ce que vous manquez, c'est donc une approximation. Tandis que l'OST fait d'abord une sorte de cartographie en analysant l'ensemble des contenus et en choisissant seulement ensuite des mots-clés. Notre logique de couverture du thème de recherche est plus systématique et inclusive et rendue possible par le fait que nous avons la base. En ce qui concerne les repérages des établissements, oui nous les exploitons dans ce genre d'études, ce qui permet aussi d'être moins approximatif qu'avec un repérage automatique.

Q : Nous complétons les repérages que nous effectuons par le marquage des adresses avec les noms des laboratoires ou des départements, Est il possible de les utiliser pour des indicateurs ?

R : Dans le cadre d'IPERU, nous ne les utilisons pas, par contre ces marquages des adresses sont stockés comme un enrichissement de la base et pourraient être exploités lors d'un contrat individuel avec votre établissement selon ses besoins.

Q : En ce qui concerne le repérage, l'indication sur les noms des publiants nous a manqué cette année et d'autre part qu'en est-il des mathématiques et de l'informatique qui comme les SHS sont des disciplines mal représentées dans le WoS.

R : En effet, il semble que l'indication des auteurs puisse être importante pour l'aide au repérage des établissements, En même temps, il n'y a pas jusqu'à aujourd'hui dans la base de lien direct entre une adresse et un auteur. Lorsque vous avez un doute sur une adresse, le fait d'avoir le nom de tous les auteurs de la publication ne vous permet de savoir comment a signé celui de votre établissement et du coup l'aide nous semble assez subjective. Par ailleurs, le périmètre du projet IPERU est défini par les laboratoires du contrat quadriennal de l'établissement (repérables dans les adresses) et nous maintenons l'objectif de respecter ce périmètre sans dériver vers un périmètre de « publiants ».

En ce qui concerne les mathématiques et l'informatique, le WoS s'est beaucoup enrichi dans les dernières années, en particulier avec l'apport d'un nouveau produit Proceedings et on fait une étude comparative avec ACM qui montre qu'aujourd'hui le WoS est tout à fait à la hauteur dans ces disciplines.

Q : Avez-vous des informations sur le groupe NormAdresses, car la problématique de la signature est souvent un moment difficile lors des accords de collaboration et dans les conventions, notamment du point de vue de la propriété intellectuelle. Nous avons besoin d'arguments.



R : Le groupe NormAdresses existe toujours potentiellement. Il y a aujourd'hui des freins institutionnels à son opérationnalisation, mais il peut être réactivé à tout moment. Je rappelle qu'à ce titre l'OST est un opérateur et non un décideur : il possède une mission d'incitation mais ne fait pas autorité. Nous étudions l'impact de la signature multi-lignes dans les repérages des établissements qui, en 2007 n'était pas flagrant. Nous en sommes à l'analyse pour 2008 mais il faudrait aller plus loin pour voir l'impact sur les indicateurs et c'est effectivement une question importante car les indicateurs sont structurants. Il semble qu'il y ait une contradiction entre le fait de demander à chacun de développer des collaborations, et celui de privilégier le niveau établissement dans les signatures. Il ne faut pas oublier que la signature multi-lignes n'est pas conforme aux standards internationaux et que s'éloigner des standards internationaux entraîne inévitablement une perte de visibilité au niveau mondial. La tension entre adaptation locale spécifique et respect des standards internationaux est permanente et conduit à poser des arbitrages entre ces deux voies. Les unités mixtes françaises ne sont pas comprises à l'étranger.

Q : Ma question porte sur la dimension temporelle. Vous avez dit pouvoir remonter jusqu'à l'année 1999 alors que les repérages dans le projet IPERU ont démarré en 2004.

R : Mon propos portait sur la base en général, les données sont disponibles de puis 1999. Pour un établissement IPERU, une telle étude nécessiterait d'effectuer un repérage complémentaires sur les années antérieures à 2002 ou 2003.

Q : J'aimerais revenir sur les présentations de ce matin et la difficulté d'interprétation des résultats envoyés par l'OST. Nous repérons la production de l'établissement par laboratoire. Or l'existence des partenariats peut constituer un biais et modifier l'interprétation des résultats. Par exemple, l'Université Lyon II essentiellement SHS est associée à des laboratoires d'informatique et de mécanique dans lesquels elle est peu impliquée mais qui lui donne sur les documents « radars » un développement important dans ces domaines et du coup nous ne nous retrouvons pas vraiment dans cet affichage.

R : Votre propos, illustre ce que nous venons d'évoquer sur les spécificités composites et mixtes des laboratoires. Un établissement est-il la somme de ses enseignants-chercheurs ou des personnels de ces laboratoires ? Ce problème est récurrent, mais il est lié aux institutions elles-mêmes. Les unités mixtes ont été conçues comme telles institutionnellement. Le fait qu'elles regroupent des personnels ayant plusieurs employeurs reflète simplement la réalité de l'établissement. On pourrait faire le repérage de façon différente si un établissement le souhaitait pour une étude spécifique, mais ce serait un point de vue individuel dans lequel l'établissement déciderait de s'identifier par son personnel. Quant à la lecture des radars, elle est toujours délicate et c'est pourquoi nous réfléchissons à des propositions de graphiques plus lisibles et par groupe d'établissements.

Q : Est-il possible d'indiquer la part de l'implication des enseignants-chercheurs des établissements ? L'établissement, attaché au critère de performance, tente d'évaluer la performance de ses équipes dans les unités mixtes. Il serait souhaitable de pouvoir connaître la part de leur implication dans la recherche.

R : Ce type d'indicateurs existe mais dans une logique individuelle. Il requiert que l'établissement collecte à son niveau les *curriculum vitae* des chercheurs et enseignants-chercheurs et relève leur production. Etablir des indicateurs par employeur qui ignorent les unités mixtes créerait des tensions nouvelles. La définition de plusieurs périmètres, contradictoires, mérite donc réflexion. En



outre, les classements internationaux privilégient la logique des laboratoires et des établissements et non les chercheurs. Ce n'est pas propre à l'OST. Un établissement est supposé être une institution qui s'identifie à un laboratoire.