Séminaire

LiSSa

Littérature Scientifique en Santé

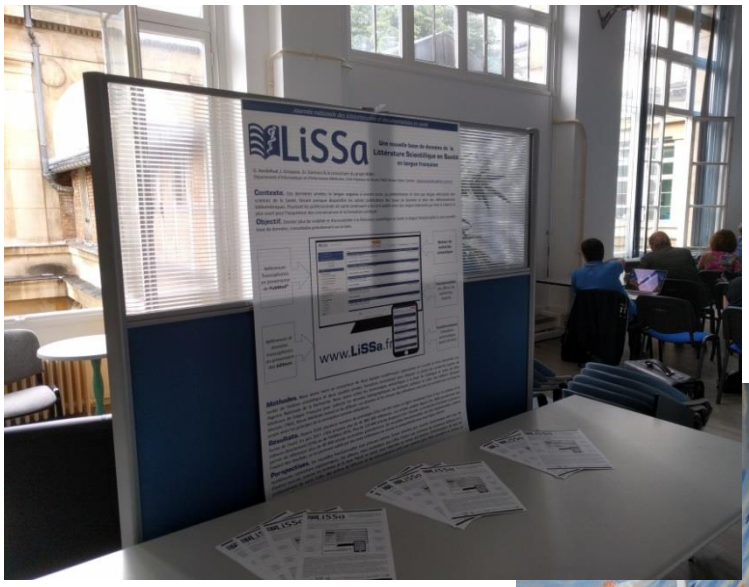
Faire connaître les publications médicales en langue française grâce à cette nouvelle base de données bibliographiques

27 juin 2017

Partenaires

Financement

Le 27 juin 2017 un séminaire consacré à la base de données LiSSa s'est déroulé dans les locaux du laboratoire LIMICS, 15 rue de l'école de médecine à Paris. Vous retrouverez dans ce document les diaporamas présentés ainsi qu'une retranscription des propos tenus et des débats effectuée à partir d'un enregistrement audio partiel. Merci aux participants pour leur attention et leurs questions.



Contenu

Introduction du Pr Stéfan Darmoni, CHU de Rouen 3
Interventions Gaétan Kerdelhué et Julien Grosjean, CHU de Rouen 4
Intervention de Julien Grosjean, CHU de Rouen 18
Gaétan Kerdelhué (suite) 25
Intervention de Sandrine Lebastard, documentaliste à la Cité de la santé 39
Gaétan Kerdelhué (suite) 44
Intervention du Dr Matthieu Schuers 48
Débat 74

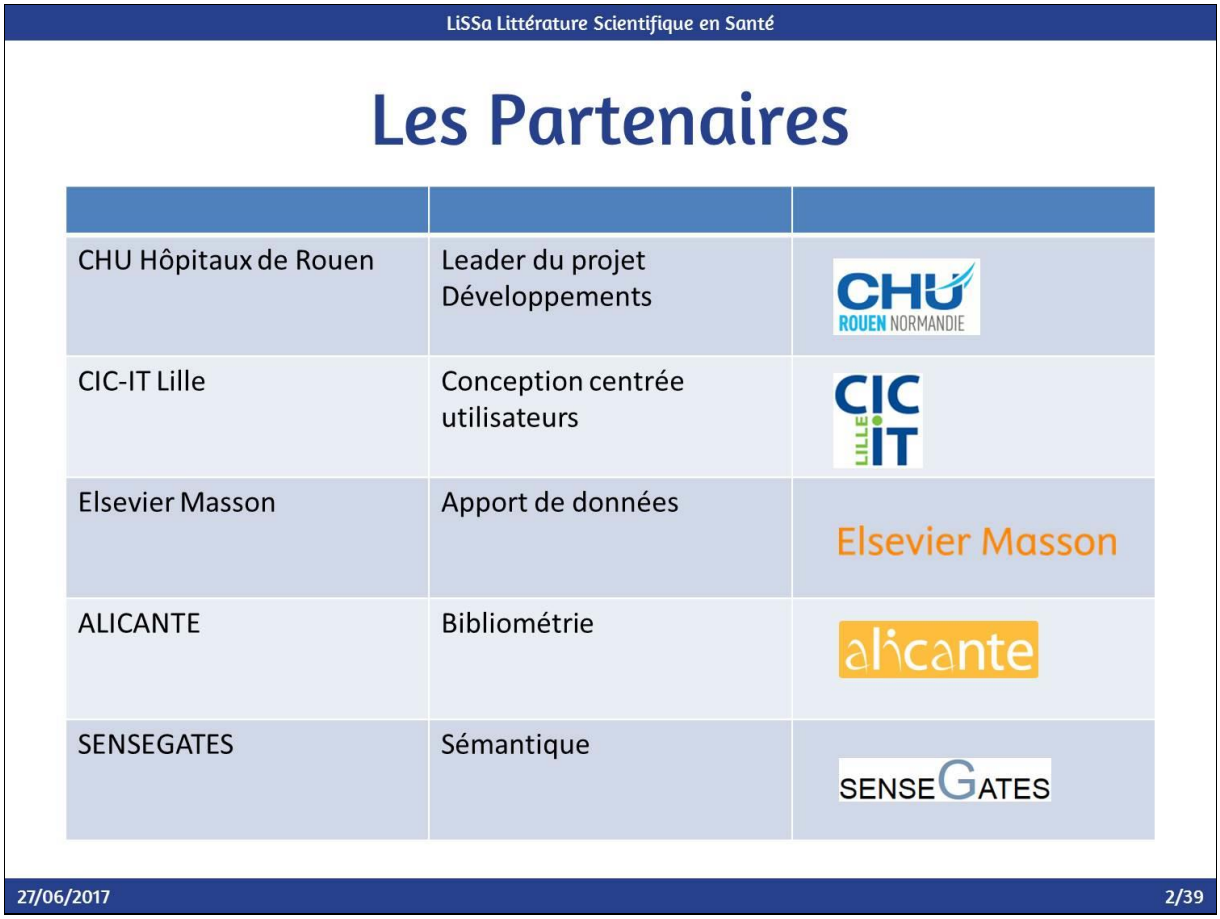
Introduction du Pr Stéfan Darmoni, CHU de Rouen

LiSSa est un acronyme qu'on a choisi et qui veut dire littérature scientifique en Santé. Il faut savoir que c'est un projet qui est toujours financé par l'ANR sous l'intitulé de BDBfr pour base de données bibliographiques en français. Ça c'est le nom du projet de Recherche, il se terminera en mars 2018. Les membres du consortium sont écrits en bas donc il y a, à part notre équipe à Rouen, il y a le CIC-IT de Lille, représentée par Laura Douze. Ce sont des spécialistes de l'ergonomie de l'interface homme machine. Il y a Elsevier-Masson qui est représenté par Phil Woodall. Il y a Alicante qui représenté par Julien Taillard. Et vous avez, in fine, SenseGates qui est une TPE représentée par Dominique Dutoit. Je ne vais pas en dire beaucoup plus, je vais aller m'asseoir et je vais écouter ce qu'ils vont vous dire.

Interventions Gaétan Kerdelhué et Julien Grosjean, CHU de Rouen

Merci beaucoup Stéfan pour cette introduction. Je vais te présenter donc : le Pr Stéfan Darmoni du CHU de Rouen et du laboratoire LIMICS. Je peux revenir sur ce qui a été dit. Le consortium, l'équipe

qui vous propose le site LiSSa aujourd'hui. Ce sont plusieurs partenaires qui sont réunis par un projet financé par l'Agence National de la Recherche.



Tout d'abord notre équipe au CHU Hôpitaux de Rouen qui est en charge du projet et qui s'occupe des développements informatiques. Notre équipe vous la connaissez surement grâce à nos réalisations précédentes notamment CISMeF, le Catalogue et Index des Sites Médicaux de langue Française qui existe depuis 1995 et aussi le Portail des Terminologies HeTOP pour Health Terminology/Ontology Portal. On a acquis une certaine expérience à la fois en documentation et en terminologies de la Santé.

Avec nous notre partenaire de Lille, le CIC-IT de Lille et l'équipe qu'on appelle Evalab. Une équipe d'ergonomes. Dans cette diapositive, j'ai essayé de résumer très brièvement le rôle de chaque partenaire. Donc ce que nous a apporté Evalab, c'est la conception centrée utilisateurs, on va en reparler, c'est-à-dire non seulement le regard des ergonomes mais les ergonomes comme porte-paroles de nos utilisateurs et qui vont intervenir tout au long du projet pour aider au développement et à l'orientation de notre site pour qu'il réponde à des besoins.

Elsevier-Masson, c'est bien sûr le leader de l'édition scientifique qui propose plus de 100 revues médicales francophones si je ne dis pas de bêtises. Nous, bien sûr, comptons sur ce partenaire pour nous apporter des données bibliographiques à côté de ce que l'on pouvait récupérer par ailleurs.

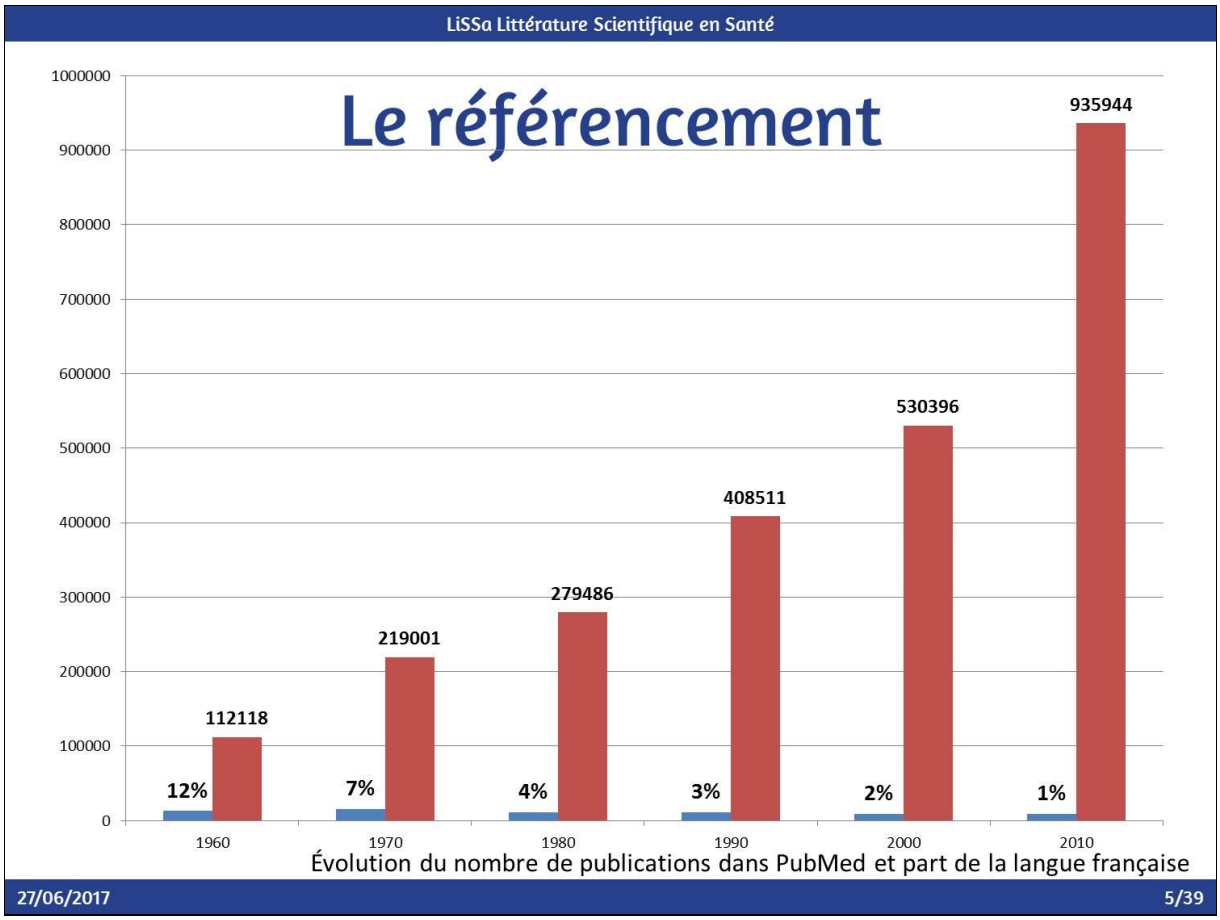
Et puis enfin, d'autres sociétés, plus petites mais pas sans talent bien sûr. Alicante, spécialiste en bibliométrie et SenseGates qui a une expertise en tout ce qui est moteur de recherche et technologie sémantique.

Tous les partenaires sont représentés aujourd'hui. On va avoir un moment de pause toute à l'heure et un moment de débat donc n'hésitez pas à leur poser toutes les questions nécessaires qui vous viendraient.

Le projet BDBfr c'est le nom du consortium, le nom du projet que nous avons soumis à l'ANR. C'est la création d'une base de données bibliographiques avec une dimension bibliométriques qui vise à rassembler les principaux contenus scientifiques en santé en français. Cet appel d'offres, on y a répondu en 2014 et on a obtenu un financement d'octobre 2014 à mars 2018. Ce qui est intéressant c'est qu'on propose un outil qui encore en cours de développement et on a voulu rendre cet outil public assez tôt dans le projet pour avoir des retours et pouvoir proposer, non pas un projet clé en mains mais quelque chose qui soit en évolution et qu'on puisse réorienter en fonctions des attentes des utilisateurs et des besoins qui se sont exprimés.

Qu'est-ce qui a motivé notre projet ? Ce sont tout d'abord des interrogations autour du rôle de la langue française pour les publications en santé. Alors ici je vais parler un peu de cette question mais uniquement pour introduire le sujet et on attend surtout du débat que nous aurons à 17h de pouvoir développer ce thème-là. Mais je vous apporte déjà quelques éléments.

Tout d'abord, bien longtemps, quand on a voulu chercher des publications en langue française et bien la meilleure façon c'était de passer par un outil américain, c'était la base PubMed. Il y a plein de données francophones dans PubMed mais pour y accéder il faut savoir interroger en anglais, passer par une interface en anglais, et même les références francophones qu'on peut découvrir dans PubMed auront été traduites vers l'anglais. Mais paradoxalement, pendant longtemps c'était quand même un des meilleurs endroits où trouver ces publications francophones.



Et puis PubMed donne une vision un petit peu triste des publications en langue française puisqu'on voit que la part relative des publications en langue française n'a pas arrêté de décliner depuis les années 50. Ici on commence en 1960, on avait 12% des publications en langue française. Au fur à mesure des années, ici j'ai fait de 10 ans en 10 ans, on a pu voir cette proportion décliner jusqu'à 1% en 2010. Le chiffre actuel est certainement encore plus faible. Non seulement le nombre relatif de publications a diminué mais, dans PubMed, le nombre absolu de publications en langue française diminue. Cela donne une vision un peu pessimiste d'un domaine francophone qui serait en déclin. On verra tout à l'heure que ce n'est pas une vision très juste de la réalité.

Le passage à l'anglais

« il a été décidé les réformes suivantes :

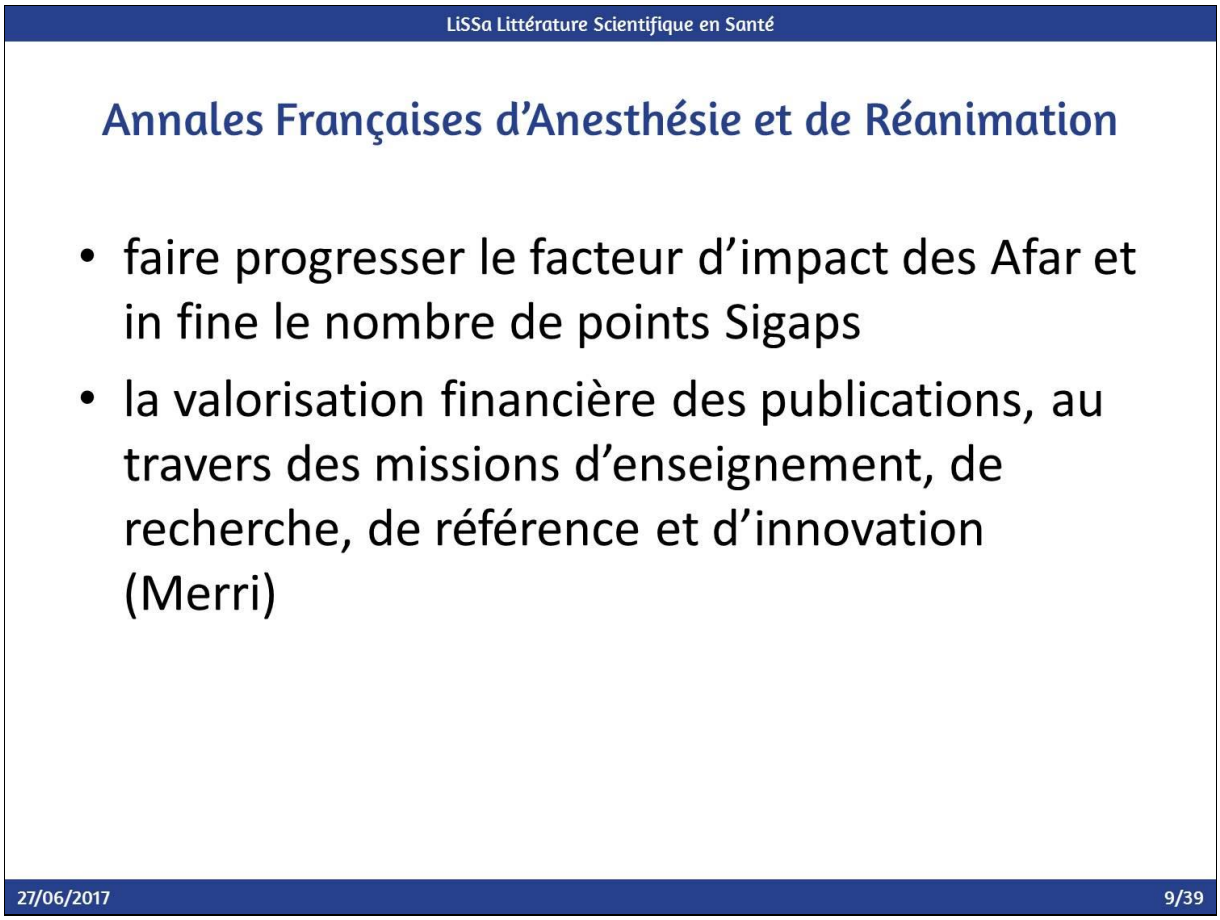
- Transformation de la revue Archives de Pédiatrie en une revue totalement anglophone, indexée, avec un contenu centré sur des revues générales, des articles originaux, des éditoriaux, ou des publications de recommandations.
- Création d'une revue pédagogique non indexée, restant totalement francophone, dont le titre reste à définir. L'objectif de cette revue sera la diffusion d'articles didactiques pour la pratique pédiatrique ou de recommandations et d'informations professionnelles. »

Chabrol B, Delacourt C, Sarles J. **Archives de Pédiatrie : quelle revue pour demain ?** Archives de Pédiatrie. mars 2017;24(3):209-10.

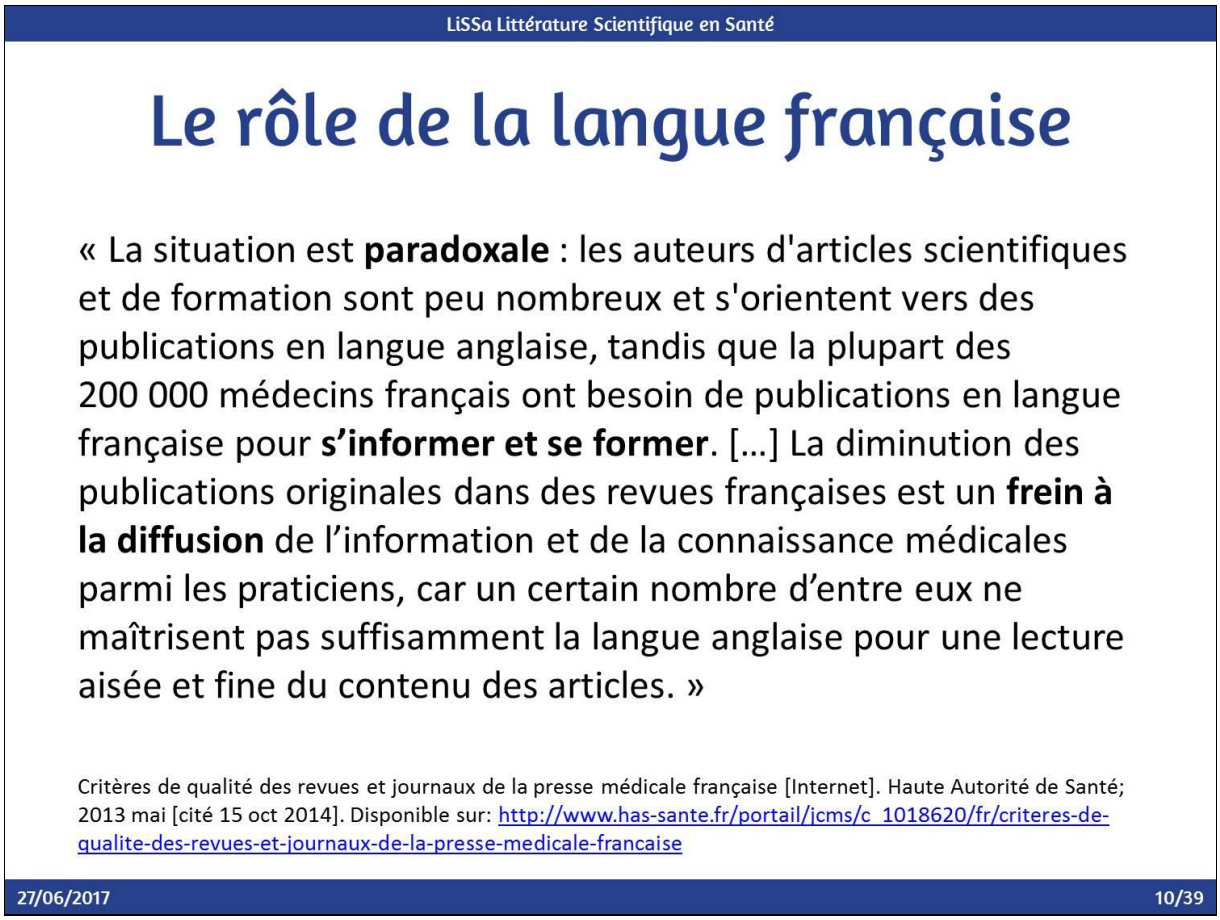
Archives de Pédiatrie

- la visibilité internationale reste faible
- pas assez attractive pour les auteurs non francophones

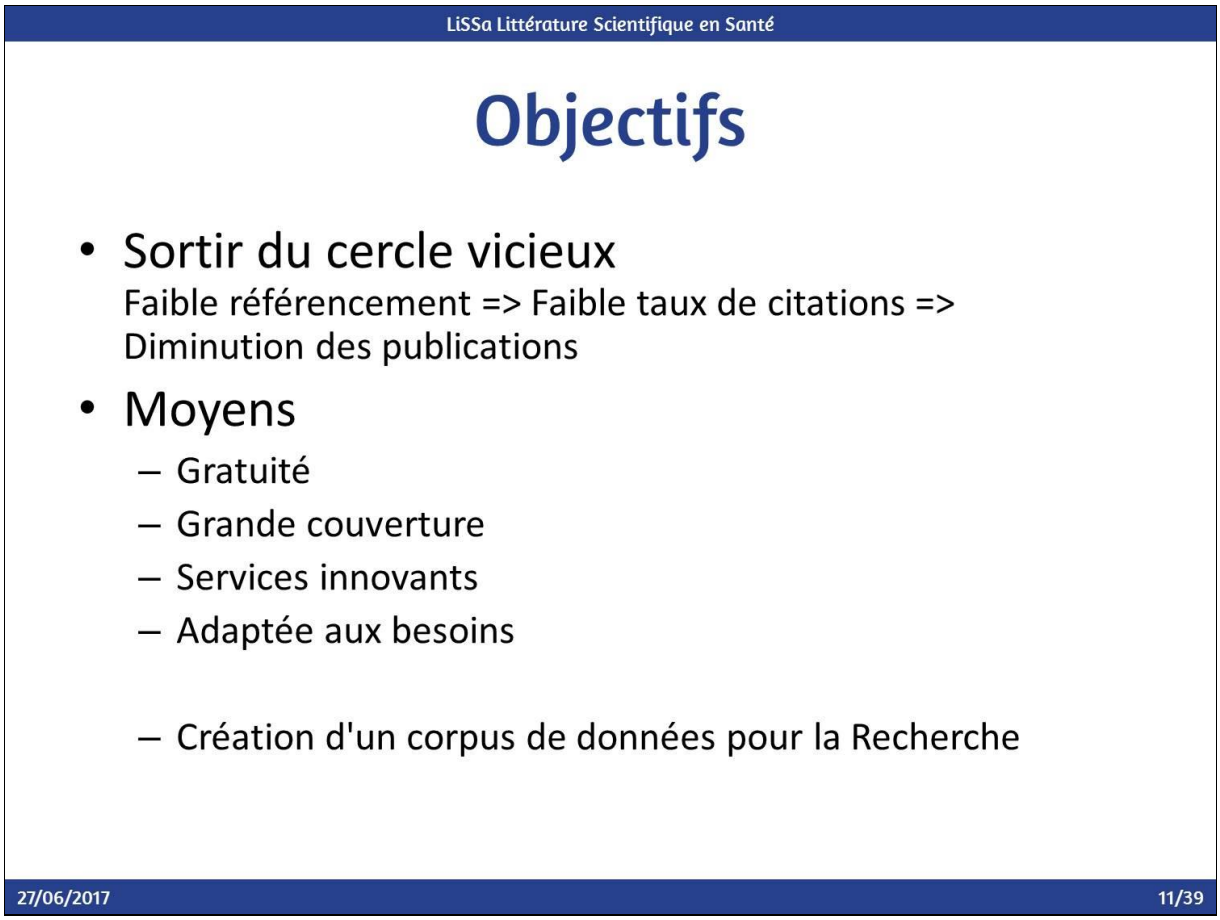
Qu'est-ce qui se passe ? On peut aller voir les revues elles-mêmes. Que nous disent-elles ? Et notamment on a l'exemple de plusieurs revues qui ont fait le choix de passer à la langue anglaise. Un exemple très récent est « les Archives de pédiatrie ». Cette décision a été rendu publique en mars 2017 dans un édito. Et donc les Archives de pédiatrie ont décidé de scinder leur revue en deux. Tous d'abord une revue totalement anglophone, qui resterait indexé dans PubMed et qui contiendrait les articles de fond, revues générales, articles originaux, éditoriaux et même les recommandations. Et d'un autre côté une revue qui serait non indexée -ça reste à voir, on espère que non- et qui serait à but pédagogique, didactique. Si on regarde leur argumentation, qu'est-ce qu'ils nous disent ? Ils nous disent : la visibilité internationale reste faible et « notre revue n'est pas assez attractive pour des auteurs non francophones ».



Plus tôt, une autre revue avait fait ce choix, ce sont « les annales françaises d'anesthésie et de réanimation ». L'option choisie a été à peu près la même c'est-à-dire scinder la revue en deux. Le français continue pour tout ce qui est éducation médicale et une revue continue pour tout ce qui est articles originaux en langue anglaise. Une différence notable est tout ce qui est recommandations qui resterait francophone. Ce qui n'est pas le cas pour les Archives de pédiatrie, on verra ce qui se passera ensuite. Si on regarde cet éditorial et leur argumentation, il rentre plus dans la question cruciale, le nerf de la guerre, la question des financements et donc, plutôt que de parler de visibilité internationale, on parle des données chiffrées et notamment du rôle du facteur d'Impact et au-delà du facteur d'Impact, des points SIGAPS. J'imagine que cela est familier à tout le monde dans la salle. Le facteur d'Impact détermine le rang d'une revue et à partir de ce rang, les auteurs toucheront un certain nombre de points. Ce système s'appelle SIGAPS. Ces points détermineront la part de l'enveloppe MERRI (pour missions d'enseignement, de recherche, de référence et d'innovation) qui sera attribuée à tel ou tel établissement. On voit qu'il n'y a pas seulement une question de visibilité, il y a tout simplement une question financière, cruciale pour les établissements.



Toute cette situation a été bien décrite dans un rapport de la Haute de Santé qui s'appelle « Critères de qualité des revues et des journaux de la presse médicale française » qui a été publié en 2013 et qui met bien le doigt sur le paradoxe actuel : d'un côté des médecins qui ont besoin du français, notamment pour s'informer et se former (la langue française est tout de même le vecteur le plus efficace pour la formation, pour l'apport de nouvelles connaissances) et d'un autre côté la diminution des publications francophones, un frein à la diffusion des connaissances.



C'est dans ce contexte que se situe notre projet et on s'est donné des objectifs ambitieux. Dans un premier temps on cherche à sortir, ou peut-être à ralentir ou contrer, une sorte de cercle vicieux qui serait : faible référencement des revues, faible taux de citation, diminution des publications et on continue ainsi. Moins il y en a, moins elles sont citées etc. Le taux de citation qu'il faut bien distinguer du taux de lecture car il y a beaucoup de revues qui sont lues, où il y a des abonnés, où on se félicite du nombre de lecteurs. Ce sont vraiment deux questions différentes.

C'est avec cet objectif que nous avons créé la base de données LiSSa. Les moyens que nous nous sommes donné. Nous avons visé la gratuité. Il faut que la base de données soit accessible pour tout le monde. On cherche à avoir la plus grande couverture possible. On veut apporter des services innovants qui soient au moins aussi bien que ce qui peut se faire à l'international et on veut que ce soit bien adapté aux besoins d'information de nos utilisateurs. Un objectif secondaire que nous nous sommes fixés, car c'est dans l'ADN de notre équipe, c'est de pouvoir se servir de LiSSa comme un support pour la Recherche, cela peut être un corpus intéressant pour la recherche en Sciences de l'information en Santé.

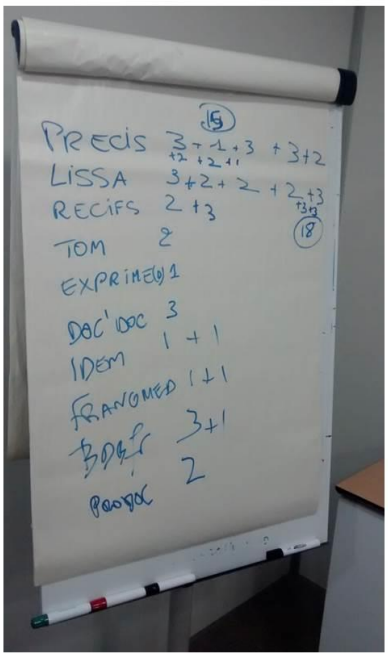
Historique

MLPubMed(fr)
Multilingual PubMed (French version)

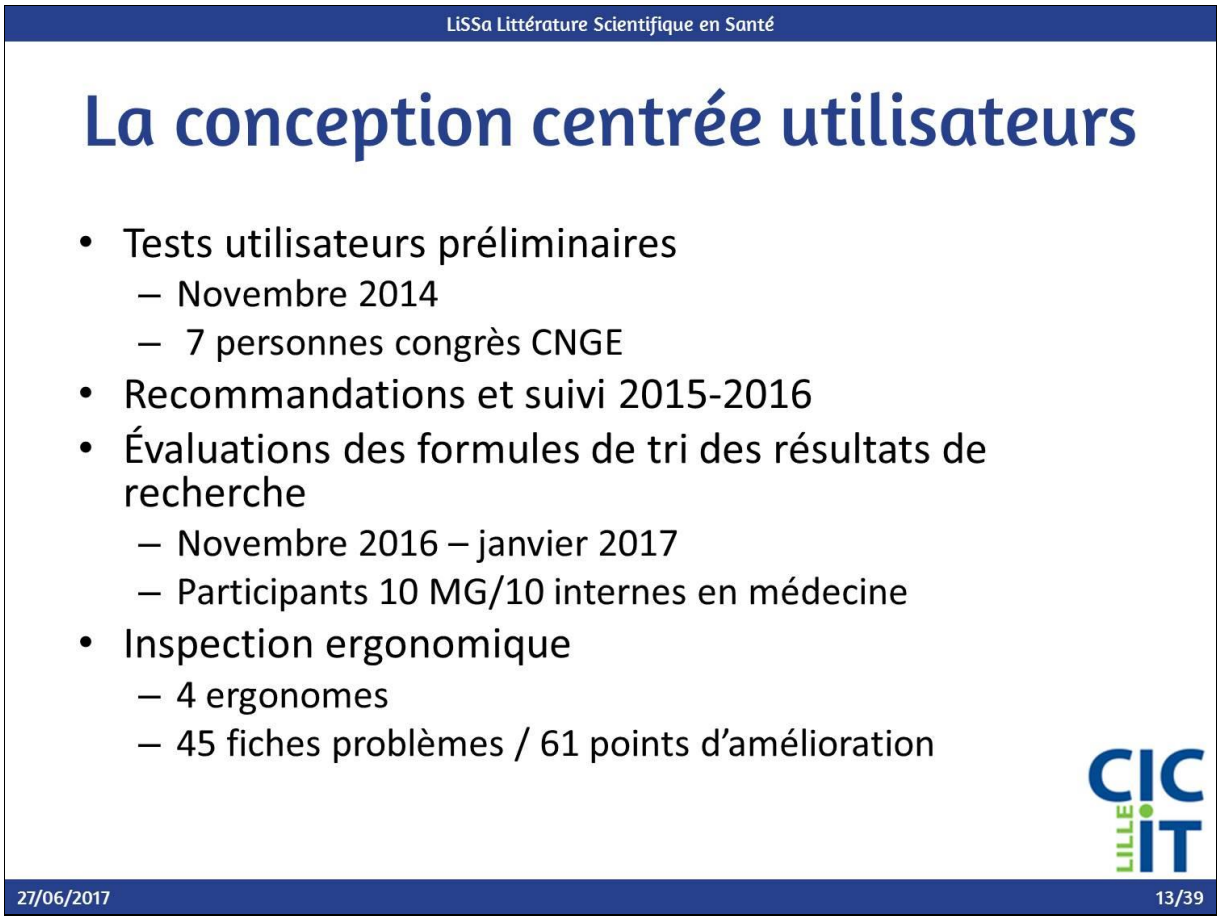
BDBfr
Base de Données Bibliographiques en français

LISSA

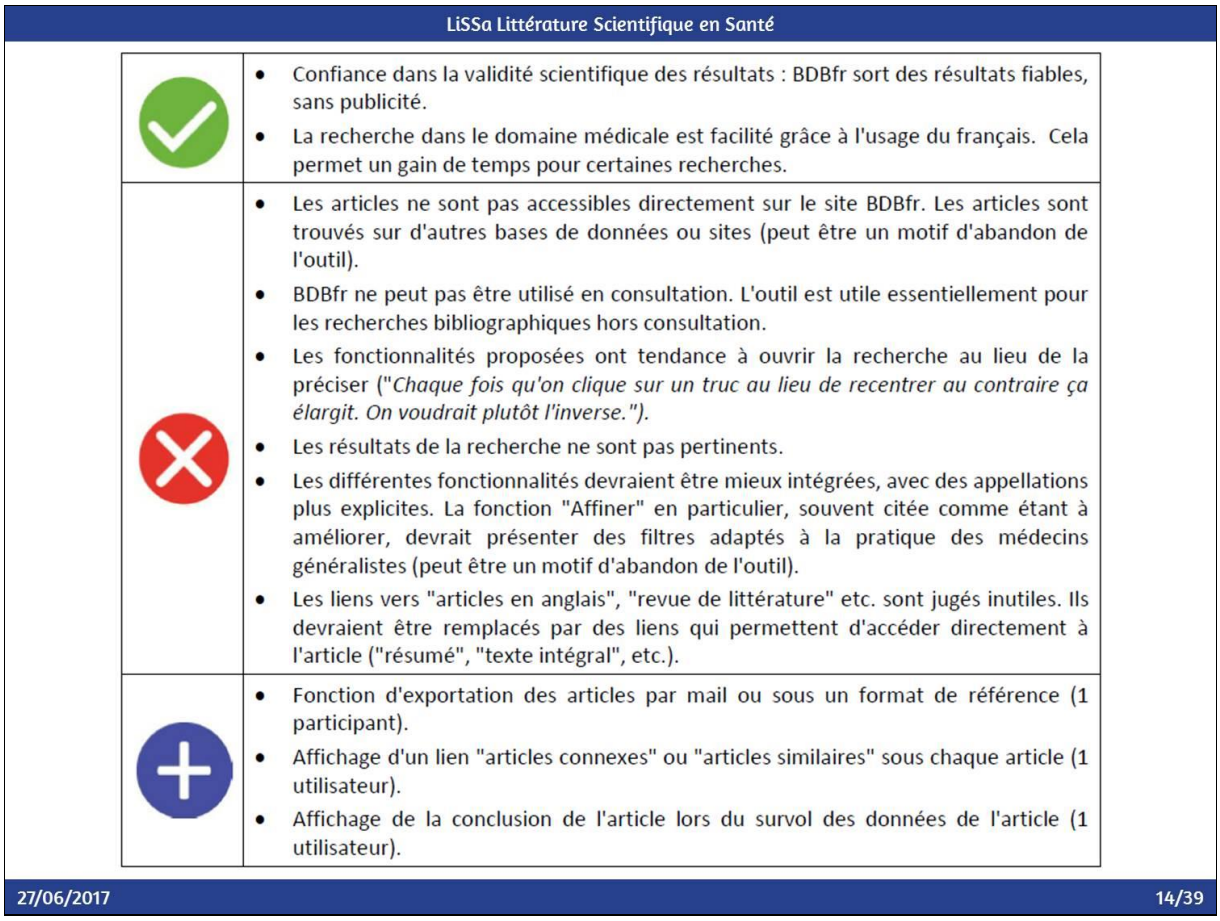
iSSa
Littérature Scientifique en Santé



Vous avez peut-être entendu parler du projet depuis pas mal de temps. Il s'est appelé de différentes manières. La toute première idée qu'on a eu c'était de faire une version multilingue de PubMed, c'est-à-dire en extrayant les données des différentes langues de PubMed et en les présentant chaque fois dans leur langue d'origine. Ça c'est le projet MLPubMed pour lequel on a commencé à publier, à présenter dans différentes conférences. C'était notre première ébauche et puis on a décidé de se focaliser sur le français. Et c'est là qu'on a réuni le consortium et déposé le projet auprès de l'ANR. En petit ici vous avez nos hésitations sur le logo et le panneau à droite c'était le moment, fort peu historique, du choix du nom. Pour la petite histoire, LiSSa a failli s'appeler PRECIS pour « Programme de Recherche de Connaissances et d'Informations en Santé ». Voilà pour l'historique.



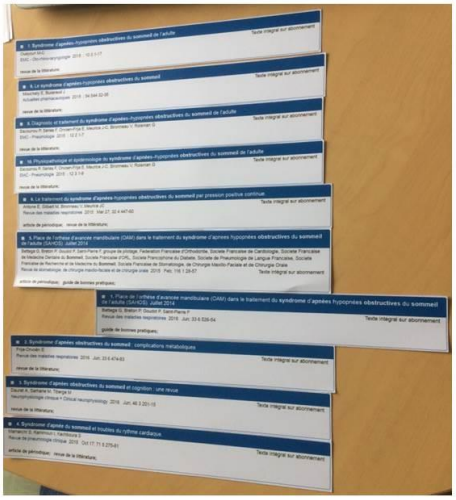
Je vais développer maintenant le rôle de notre partenaire du CIC-IT de Lille. Cela a été très important pour nous, la conception centrée utilisateurs. On apprend beaucoup en travaillant avec des ergonomes. On peut essayer de s'approprier leurs principes mais avoir des vrais professionnels avec nous ça a été très précieux. L'équipe est intervenue tout au long du projet et interviendra encore. Ce qui a été très intéressant c'est de réaliser des premiers tests utilisateurs très tôt dans le projet. Le projet a démarré en octobre 2014 et en novembre 2014 nous avions des retours qualitatifs sur notre outil. Cela s'est fait au congrès national des généralistes enseignants, le CNGE. 7 personnes ont pu être interviewées devant l'ordinateur en pratiquant l'outil, en verbalisant tout ce qu'ils avaient à dire sur notre base de données. C'était un point de départ très important et Evalab CIC-IT et intervenu tout du long en nous proposant des recommandations sur l'outil, en suivant les modifications que nous avons pu effectuer. Une fonctionnalité importante qu'on a implémenté c'est le tri des résultats. Encore une fois on s'est appuyé sur ce que disaient les utilisateurs. On a fait intervenir 10 médecins généralistes et 10 internes en médecine. Cela nous a permis de trancher entre plusieurs propositions qu'on avait en tête et ensuite de les affiner. Tout récemment, on a eu une inspection ergonomique de fond en comble de l'outil par quatre ergonomes qui nous ont relevé 45 fiches problèmes et 61 points d'améliorations. Tout cela on va l'améliorer tout du long. Ceux qui ont suivi le projet ont pu voir déjà plusieurs versions différentes de l'outil avec à chaque fois la prise en compte de ces recommandations et des améliorations sur le contenu et sur la forme.



Voici le premier retour qu'on a eu en novembre 2014 à ce congrès de médecins généralistes. Cela nous a donné de la confiance dans notre projet en validant quelques-uns de nos présupposés. Tout d'abord le rôle du français. Les médecins qui sont passé devant nous, qui ont testé l'outil ont confirmé le rôle et l'importance de la langue française. C'était nécessaire. Et aussi le rôle des publications scientifiques. Bien sûr on peut trouver de l'information francophone avec Google mais ce n'est pas suffisant. Le fait de présenter un moteur de recherche qui propose des références d'articles scientifiques a aussi été validé. On a su que cela répondait vraiment à des besoins. Ils nous ont aussi mis le nez sur les premiers défauts de l'outil : manque de temps de réaction, problème d'accès aux articles, manque de pertinence. Cela nous a donné des caps très tôt, des choses à améliorer pour la base de données.

System Usability Scale

	Pas du tout d'accord				Tout à fait d'accord
1. Je pense que j'aimerai utiliser cette application souvent.	1	2	3	4	5
2. J'ai trouvé l'application inutilement complexe.	1	2	3	4	5
3. J'ai trouvé l'application facile à utiliser.	1	2	3	4	5
4. Je pense que j'aurais besoin d'une assistance technique pour être capable d'utiliser l'application.	1	2	3	4	5
5. J'ai trouvé les différentes fonctions de l'application bien intégrées.	1	2	3	4	5
6. L'application n'était pas homogène.	1	2	3	4	5
7. J'imagine que la plupart des gens vont rapidement apprendre à utiliser cette application.	1	2	3	4	5
8. J'ai trouvé cette application fastidieuse à utiliser.	1	2	3	4	5
9. J'ai eu confiance lors de l'utilisation de cette application.	1	2	3	4	5
10. J'ai encore besoin d'apprendre avant de me sentir à l'aise avec l'application.	1	2	3	4	5

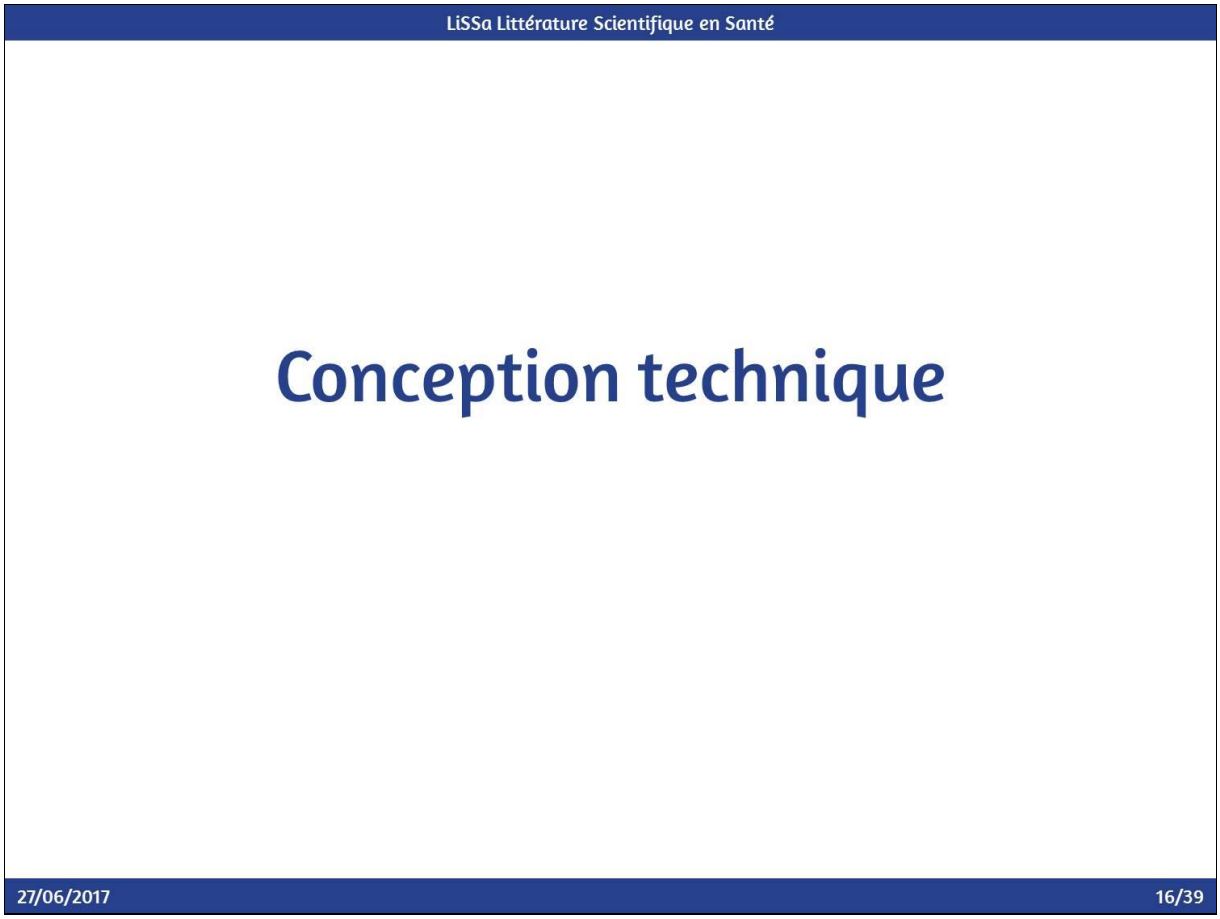


Et verbalisation !

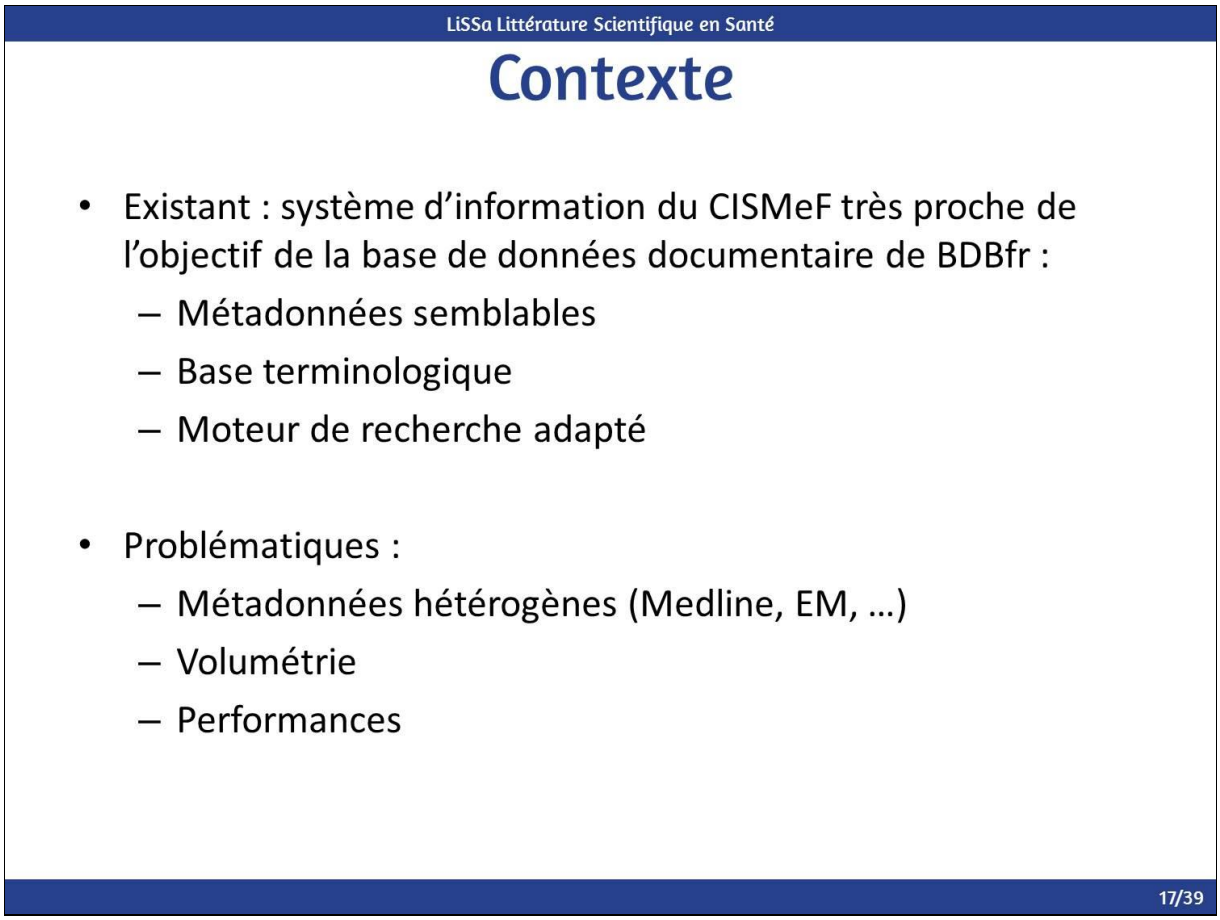
Ici quelques-uns des outils utilisés par les ergonomes. C'est intéressant de connaître leurs méthodes. Il y a questionnaires, tout bêtement, questionnaire SUS pour System Usability Scale. Et puis on découvre que ces questionnaires sont là aussi comme support de parole et que ce qui importe c'est de faire parler les gens. À droite on a un exemple sur le tri des résultats, les utilisateurs ont pu manier directement sur papier le tri des résultats et montré comment ils voyaient ce tri. Là c'est intéressant pour une technologie Web de repasser par des supports papier très concrets. On voit que cela peut apporter beaucoup.

Voilà pour l'aspect ergonomie, je vais passer la parole à Julien Grosjean qui est ingénieur informatique dans notre équipe, qui a suivi le projet depuis le début et qui va parler de la construction technique.

Intervention de Julien Grosjean, CHU de Rouen



Bonjour à tous, merci Gaétan. Le but n'est pas de vous assommer avec du jargon technique et d'autres algorithmes. Le but est de vous expliquer quelles sont nos méthodes, nos outils qu'on espère innovants. D'une part parce qu'on est historiquement un labo de recherche mais aussi parce qu'on s'inscrit dans un projet de recherche ANR technologies pour la santé. Donc notre but est aussi de s'appuyer sur des solutions techniques, un certain nombre d'outils qui permettent d'améliorer l'expérience utilisateurs. Avant tout, c'est ça notre intérêt.

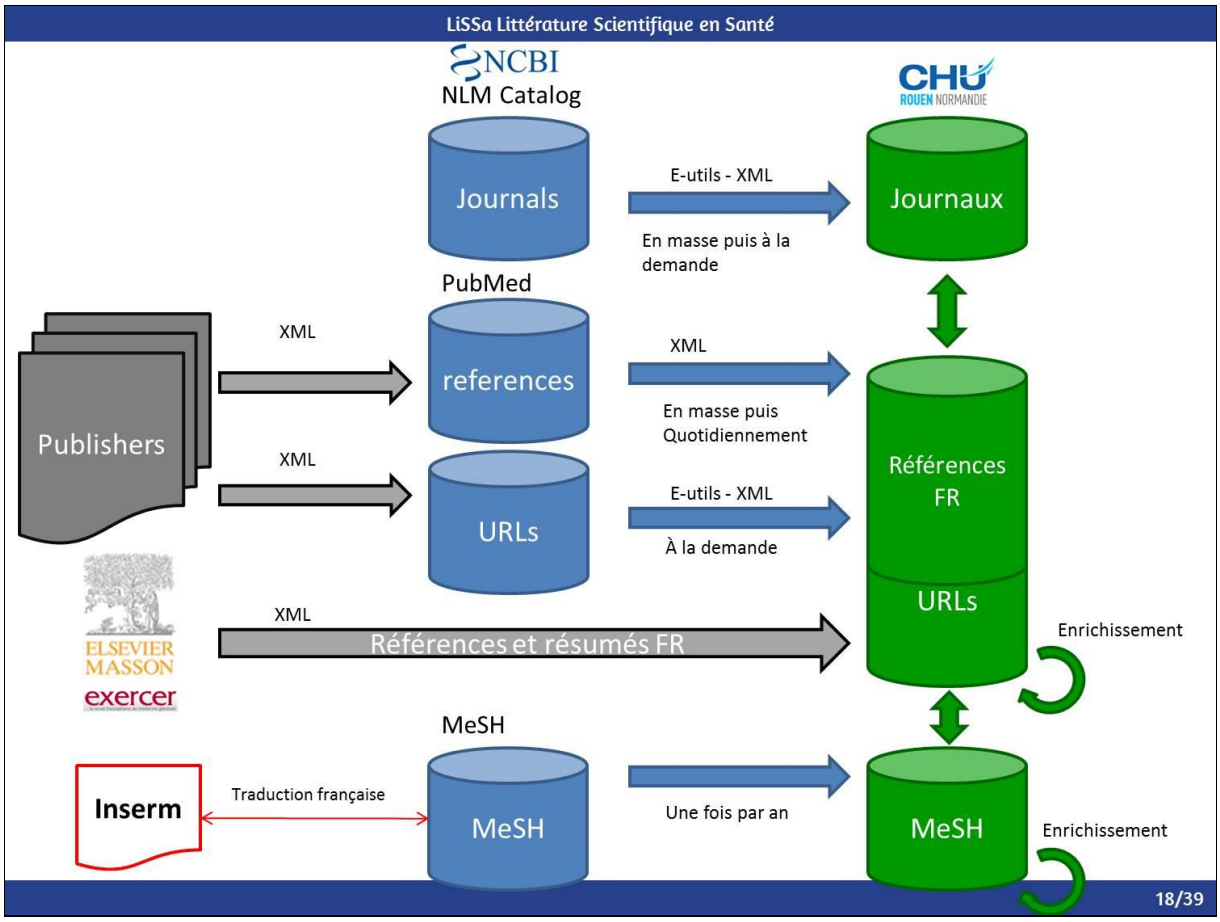


Quand on a commencé ce projet on était, bien sûr, pas en terrain complètement inconnu puisque, comme disait Gaétan, depuis 1995 maintenant, plus de vingt ans, on a utilisé le MeSH pour indexer le catalogue et index des sites médicaux de langue française avec un moteur de recherche préexistant, avec beaucoup de connaissances et d'outils déjà acquis dans le domaine. Vers 2005-2006 on s'est rendu compte que les outils avaient beaucoup évolué, que les technologies elles n'avaient pas beaucoup évolué et qu'on pouvait réutiliser une partie des acquis du moteur de recherche et des acquis de CISMeF. Mais, si on voulait aller plus loin pour BDBfr il fallait sans doute partir sur des méthodologies différentes et des technologies différentes.

On disposait de métadonnées semblables mais pour CISMeF, vous connaissez sans doute, on est sur des ressources sur Internet avec un certain nombre de métadonnées qui ne sont pas compatibles avec des métadonnées de citations classiques. On avait également, on a toujours, une base terminologique qui, à l'époque était sans doute un peu moins riche, qu'il fallait étendre on en reparlera pas vraiment mais, juste un petit mot sur l'indexation automatique avec la base de terminologies et puis vous parlez du moteur de recherche.

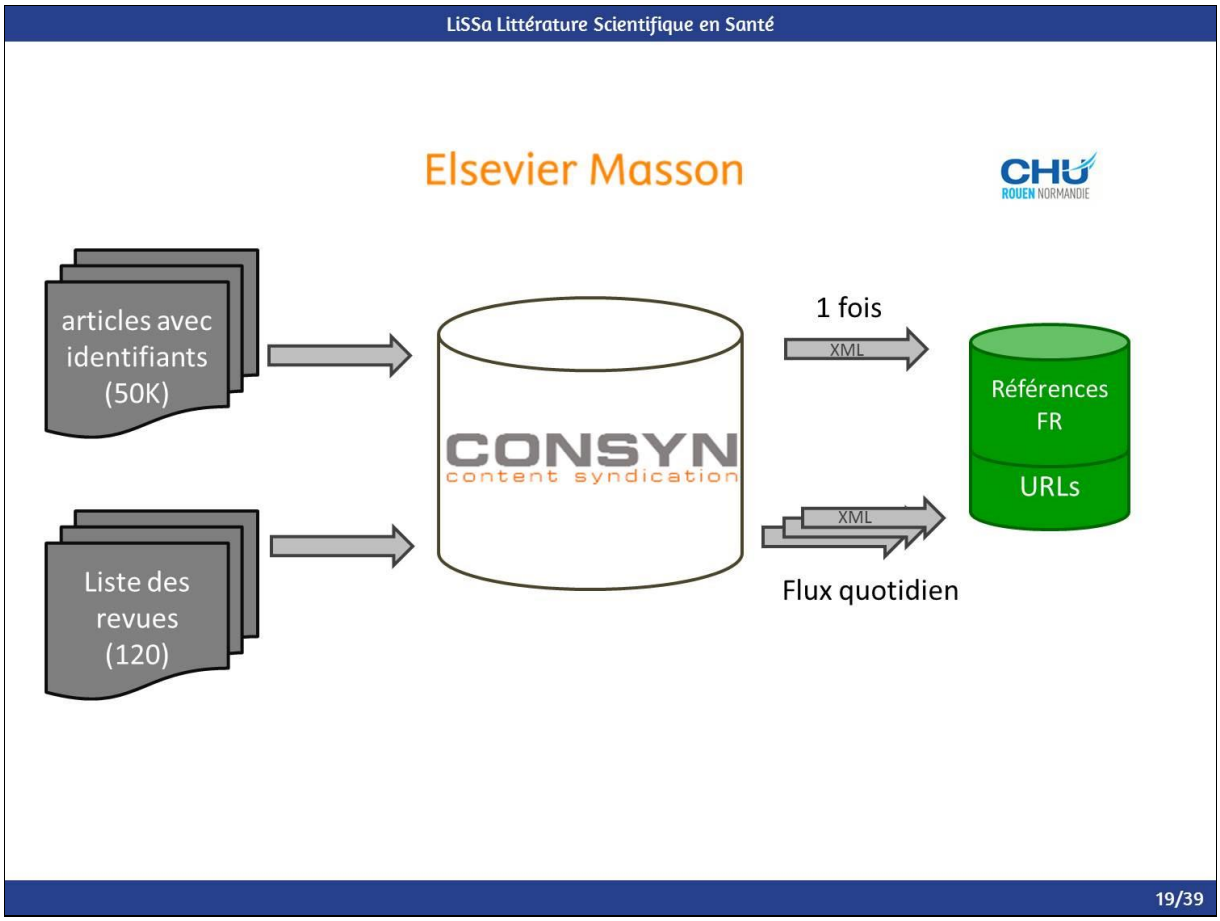
Quand on a commencé BDBfr il était environ de 100 000 références alors qu'on savait qu'a priori PubMed en langue française c'était minimum 650 000 citations, donc un passage à l'échelle pour une problématique technique. Les problématiques techniques étaient essentiellement celles-ci : essayer de fusionner des métadonnées qui étaient hétérogènes, qui pouvaient venir de plusieurs sources différentes, on a parlé de MEDLINE mais avec notre partenaire Elsevier Masson, il fallait qu'on puisse agréger ces ressources mais également des sources d'autres éditeurs. Et puis la question de la volumétrie avec un passage à l'échelle au moins d'un facteur 6,5 et on savait vouloir aller beaucoup

plus loin. Et puis des problèmes de performances, puisque le moteur proposé à ce moment-là on le trouvait un peu lourd, pas forcément adapté ni en termes de temps de réponse ni en termes d'utilisabilité pour les utilisateurs finaux.

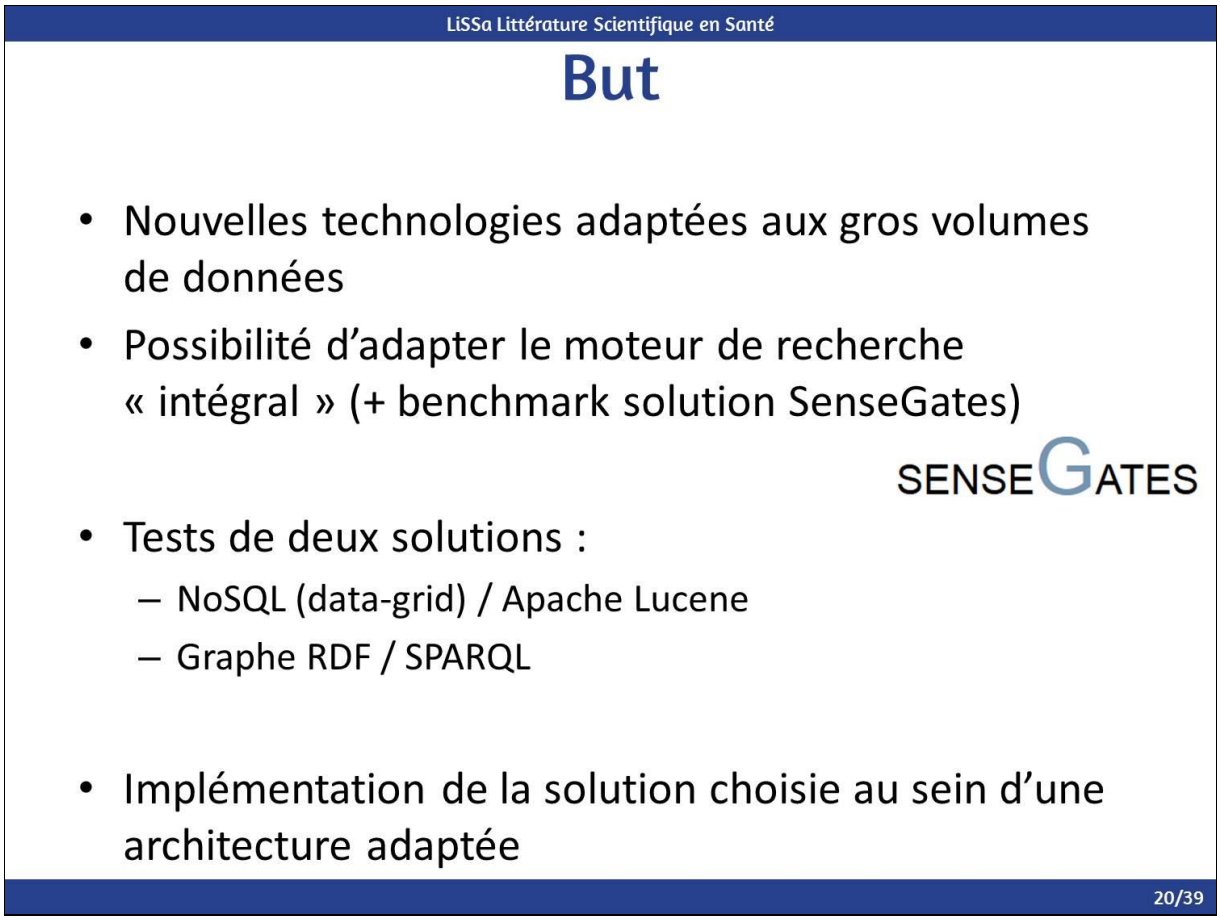


Voilà un schéma, un petit peu pompeux peut-être mais pas si compliqué. A gauche on a les sources d'informations. On s'est concentré tout d'abord fortement sur ce qui est issu du NCBI aux États-Unis, c'est-à-dire le catalogue de journaux et les références PubMed qui nous ont permis de faire un premier import dans la base : les journaux qui publient au moins en partie des articles médicaux en français. On a fait ça dans un premier temps en masse. C'était un challenge technique assez intéressant et aujourd'hui cet outil met à jour quotidiennement les ressources qui sont inclus dans PubMed ou qui sont mises à jour, indexées MeSH par exemple. Ces ressources sont directement versées dans notre base de données de façon quotidienne. On a fait ce travail dans le cadre de BDBfr avec les deux sources que vous voyez en dessous, Elsevier Masson et une plus petite contribution de la part d'Exercer qui propose peu de contenu mais qui a fait l'effort de proposer un format qui était compatible. C'est un problème bassement technique mais quand on parle de pouvoir indexer dans LiSSa de nouvelles ressources on va avoir ce problème technique de récupérer l'information chez les éditeurs et les fournisseurs d'articles scientifiques pour pouvoir les intégrer. Cela demande un travail pas forcément très complexe, fusionner les métadonnées, créer des programmes spécifiques qui permettent d'intégrer ça de façon intelligente. Parallèlement à ça, on a, tout en bas, la couche terminologique, ici le MeSH car c'est la terminologie la plus importante pour indexer mais, en vérité, on a beaucoup d'autres. C'est pareil il y a des mises à jour qui sont obligatoires. La mise à jour du MeSH c'est au moins une fois par an. Pour les autres terminologies qui peuvent nous servir, il y a un

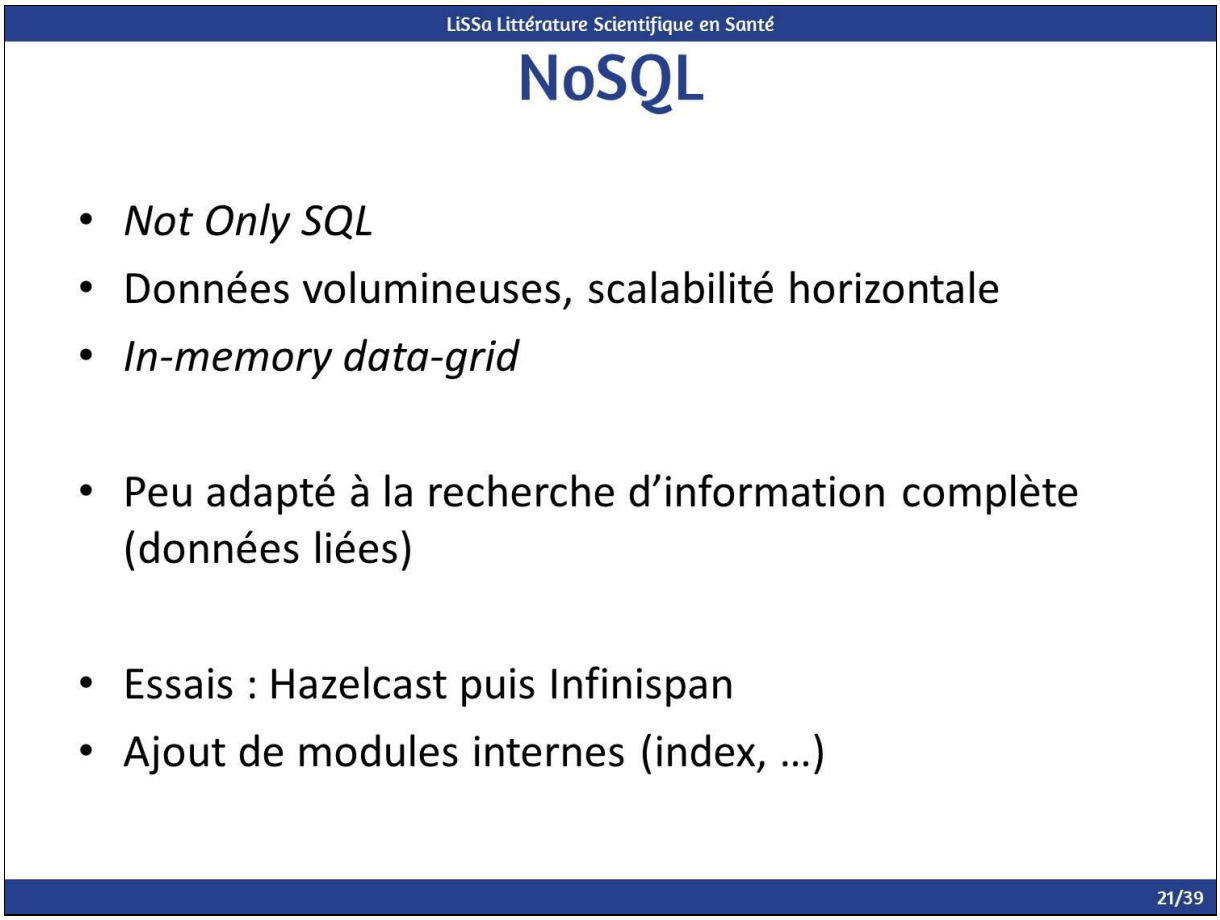
travail de notre équipe technique pour maintenir cette base terminologique qui contient des millions et des millions de termes et puis un travail des experts de l'équipe pour pouvoir maintenir ces terminologies quand cela n'est pas fait de façon officielle par d'autres sources. Je ne vais pas rentrer dans le détail.



Il y a eu un travail dans le cadre de BDBfr qui est fait avec notre partenaire Elsevier-Masson qui dispose déjà d'un entrepôt, d'une application Web qui nous permet de requêter sur leur base, sur la totalité de leurs articles. Ils nous mettent à disposition cet outil qui est un outil international qui est de leur côté surtout pour des usages internes, je crois. Il nous permet d'extraire automatiquement, sous forme de flux, d'abord en une fois et puis maintenant de façon quotidienne. C'est la même chose que pour PubMed, on extrait automatiquement les nouvelles ressources qui sont produites par Elsevier Masson. Là vous voyez qu'on a environ 120 revues dans notre base.



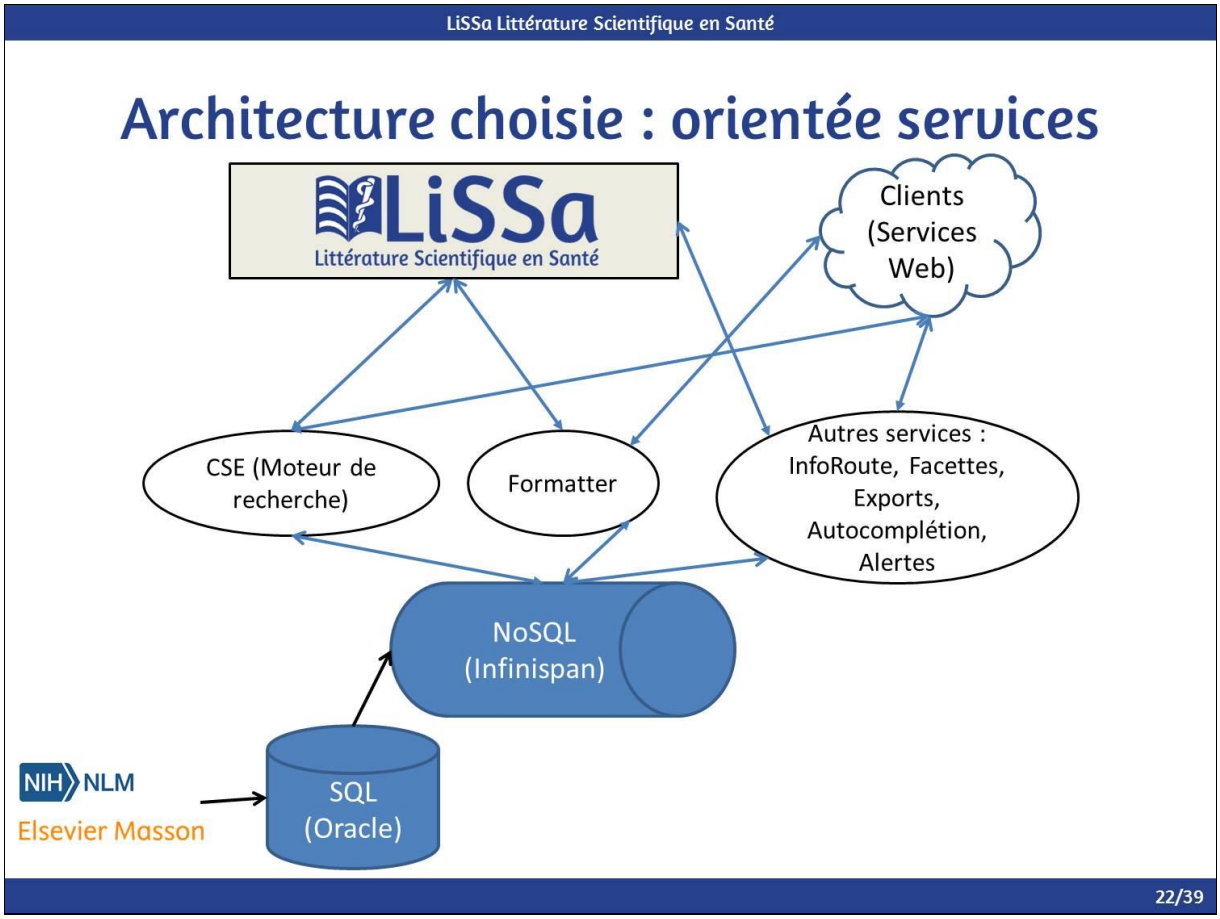
Sans renter trop dans les détails technologiques, j'ai parlé de notre challenge de créer des nouvelles méthodologies et des nouveaux outils qui permettaient de gérer des gros volumes de données. On a commencé un travail en 2013-14 sur la possibilité de changer de base de données. On était originellement en SQL, voilà un gros mot. Il existe aujourd'hui des nouveaux paradigmes en informatique, des nouvelles façons de faire, qui permettent de traiter des données beaucoup plus volumineuses et donc des moteurs de recherche associés qui permettent de requêter ces grosses bases de données de la façon la plus rapide possible, pour être en dessous de la seconde en recherche d'information. Ça c'est notamment l'objectif de notre partenariat avec SenseGates. C'était de pouvoir comparer un certain nombre de solutions qui permettent de requêter dans des millions de citations de façon vraiment très brève.



Parallèlement à ça, on a fait le test avec d'autres solutions qui sont vraiment des paradigmes en informatique et en recherche d'information : des bases de données qu'on appelle NoSQL et des bases de données plutôt sous forme de graphe avec des données liées, en RDF. Le constat était sans appel puisqu'on a implémenté la solution NoSQL qui est bien plus performante. NoSQL, sans rentrer dans le détail, c'est Not Only SQL, c'est-à-dire qu'on ne fait pas que du SQL. On a des données qui sont stockées plutôt à plat car le but est de gérer des données plutôt volumineuses. Scalabilité horizontale ça veut dire qu'on met beaucoup de machines. Et « In-memory data-grid », c'est des machines qui vont être très chers, très costaudes en mémoire vive. Donc ça demande un coût en infrastructure qui n'est pas négligeable mais qui permet d'améliorer significativement. On parle de minimum de diviser par 20, par 50, voire par 100 en fonction de certaines requêtes et donc avoir une expérience la plus instantanée possible.

Ces types de technologies c'est relativement récent, pas forcément toujours adaptées à toutes les problématiques qui existent, bien entendu et donc surtout pas à la recherche d'information complète, c'est-à-dire avec des données qui sont liées entre elles. Contrairement à des outils comme Google qui eux ont, bien sûr, énormément de données à gérer mais qui les gère souvent de façon à plat. Alors qu'ici on doit pouvoir chainer l'information. L'indexation des documents, les revues etc., il y a un certain nombre d'entités qui doivent être liées. Si on veut, c'était notre défi, avoir des temps de réponse correcte avec une interrogation complète où on peut utiliser vraiment de façon exhaustive la donnée et bien on se devait d'adapter nos solutions, de faire nos propres surcouches, de développer nous-même des modules qui n'existaient pas encore. On se basait sur un certain

nombre de frameworks nommés ici mais on a fait beaucoup de travail nous-mêmes. On a la chance de disposer d'une équipe de quatre ingénieurs et donc nous avons pu le faire.



Désolé pour ce schéma, on a pas eu le temps de le refondre, c'est un vieux schéma mais, en gros, ce qu'il faut voir c'est qu'on part sur une couche de données qui sont en partie en base données SQL et en partie NoSQL et qu'ensuite on dispose de toute un ensemble de services qui vont interroger cette base, qui vont être à la fois, tout en haut, utilisé dans le cadre du site Web LiSSa que vous connaissez mais aussi tous ces services peuvent être disponibles pour les machines. Ça c'est très important aujourd'hui, c'est de proposer des API. Si on compare par exemple à PubMed, on sait qu'aujourd'hui les appels par les machines à PubMed ont dépassé les appels par les êtres humains. C'est donc qu'aujourd'hui il y a énormément d'outils qui vont récupérer l'information pour pouvoir agréger, pour pouvoir présenter autrement l'information, pour faire un tas de choses. Et donc LiSSa se devait de proposer également les mêmes fonctionnalités, c'est-à-dire des services Web qui sont utilisables par les machines. Et ainsi, que les gens puissent développer des surcouches à l'envie.

Gaétan, je te laisse la parole.

Gaétan Kerdelhué (suite)

Merci beaucoup Julien. Pour résumer ça va beaucoup plus vite !

>> JG

C'est ça on a gagné au moins un facteur 20.

>>Gaétan Kerdelhué

Si vous testez LiSSa, ou même CISMeF aujourd'hui…

>>Stéfan Darmoni

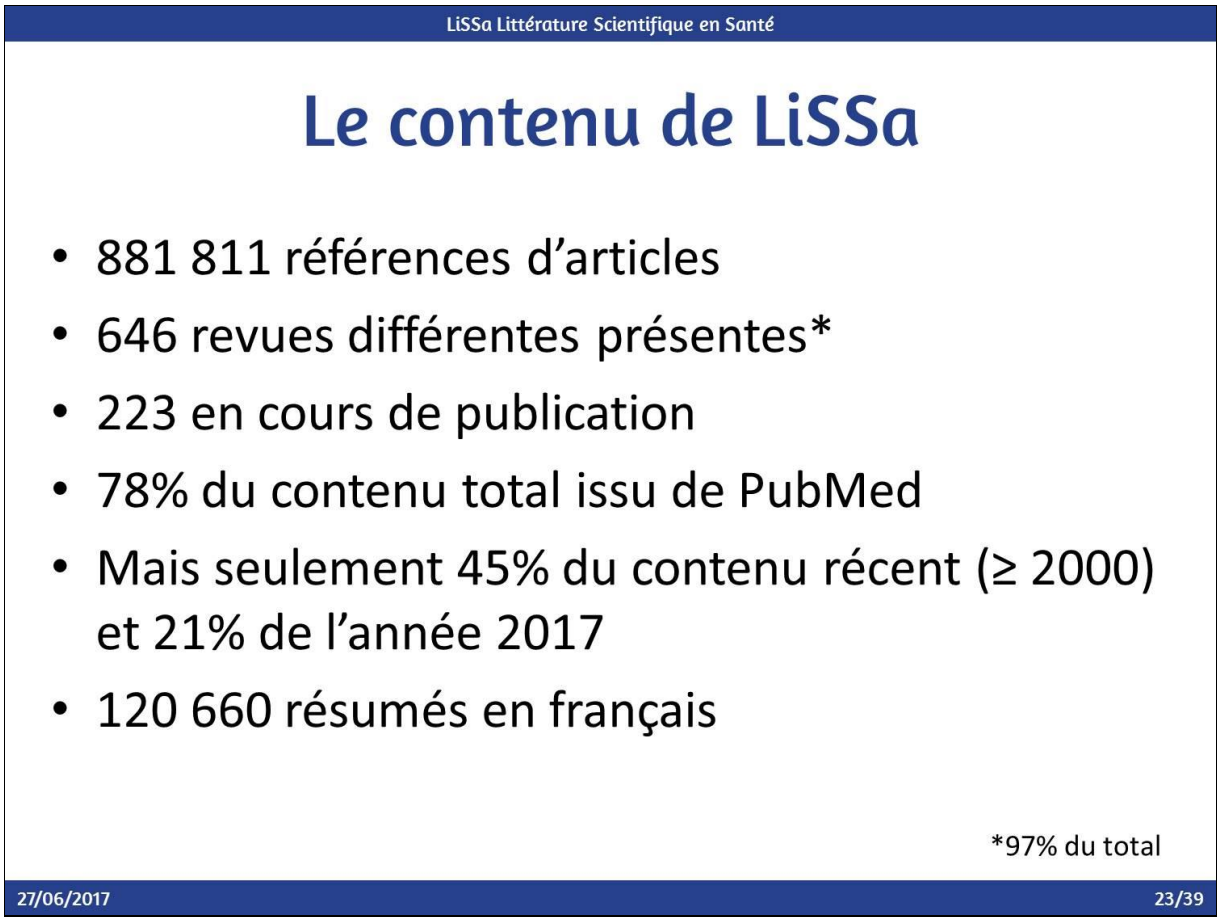
Pas encore CISMeF !

>> Gaétan Kerdelhué

Si vous comparez à ce que qu'on faisait il y a quelques années, le saut est vraiment très important.

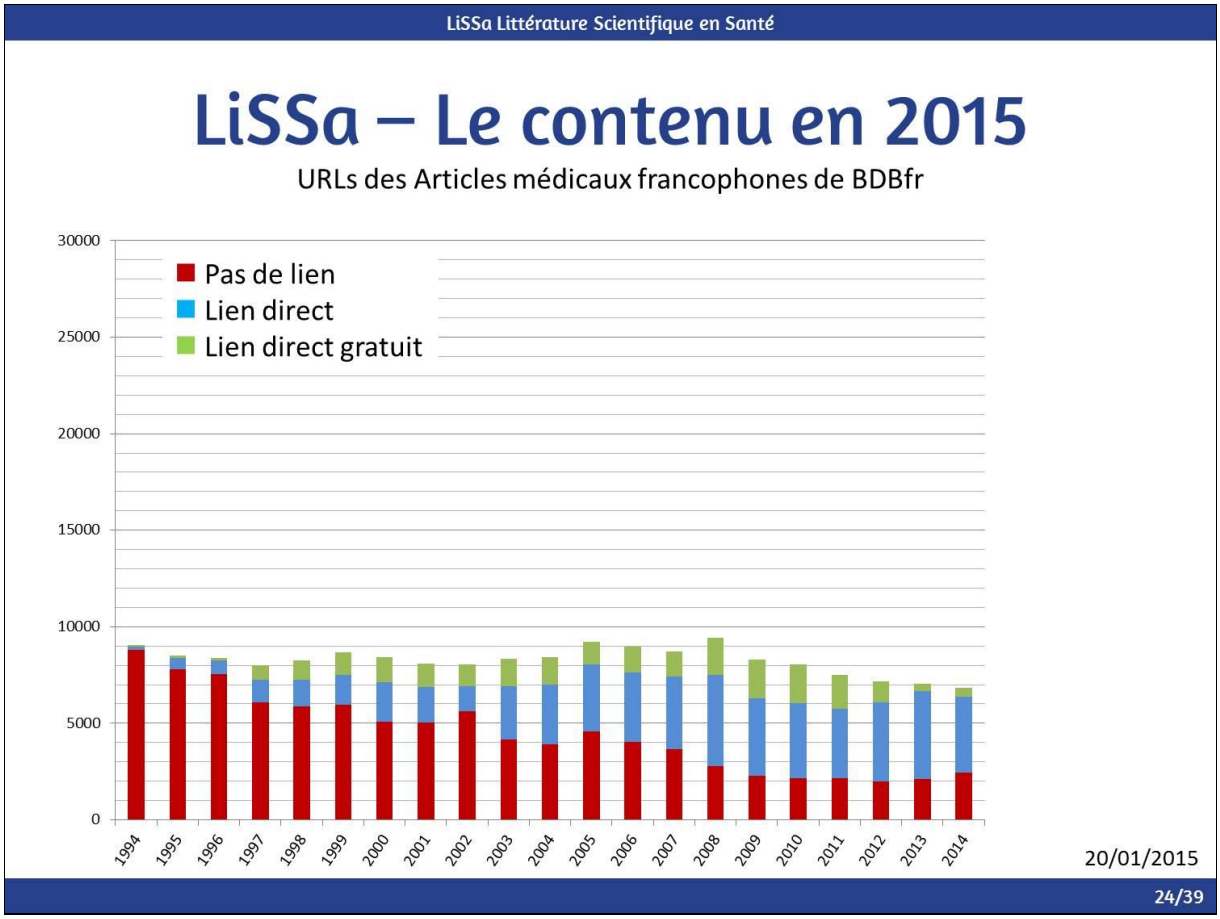
>>JG

Vous pouvez tester CISMeF qui a 8 fois moins de contenu que LiSSa et qui est plus long que LiSSa



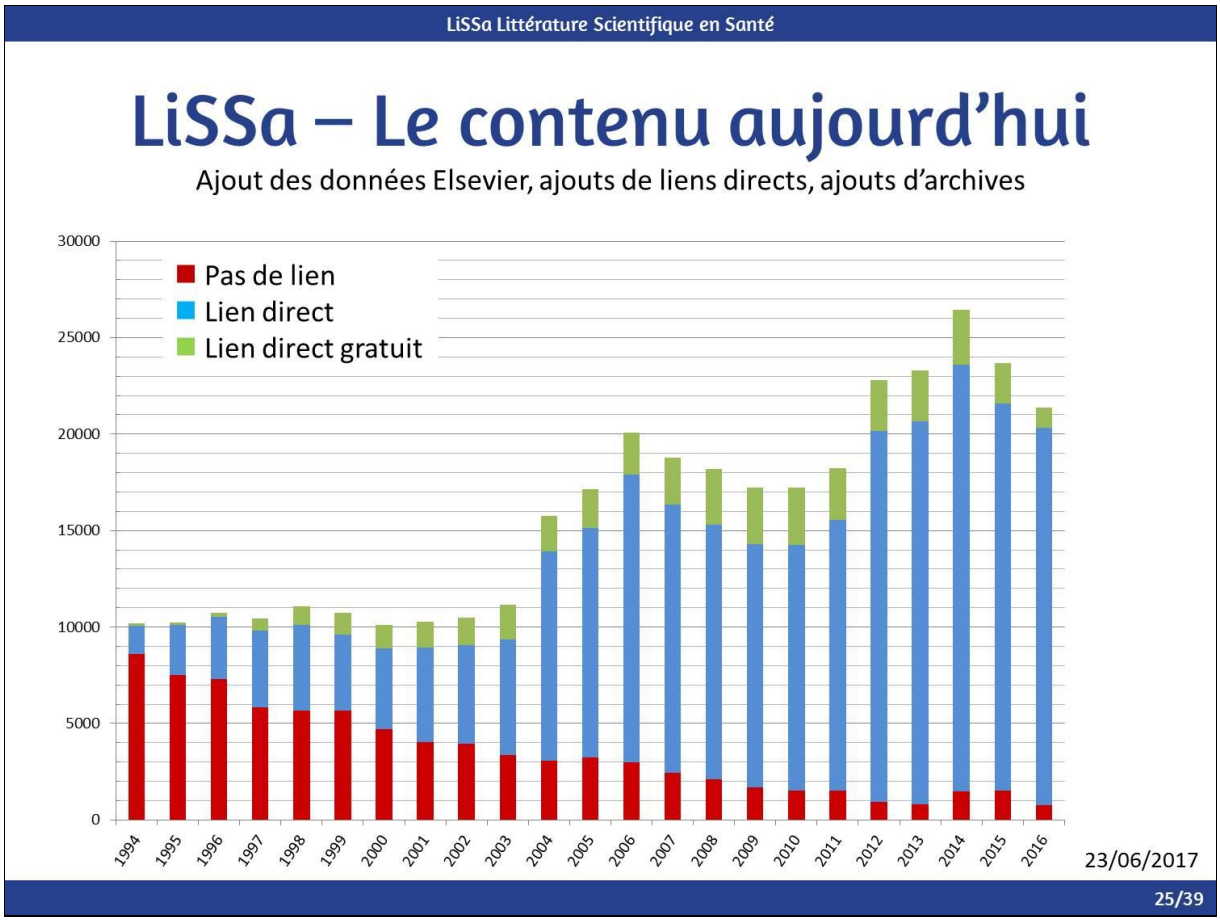
Ces chiffres datent d'hier. Qu'est-ce qu'on va trouver dans notre base de données ? Nous avons aujourd'hui 881 000 références d'articles. Tout ça provient de 646 revues différentes. Ce sont les 646 revues qui proposent pus de 100 articles ou qui sont en cours de publications. Ca fait plus de 97% du total. On a encore plusieurs milliers de revues mais qui ne proposent qu'une poignée de références uniquement, qui sont moins intéressantes. Il y a aujourd'hui 223 revues en cours de publications. Si on regarde nos informations on les a pris de PubMed et on travaille avec les éditeurs. Si on regarde sur l'ensemble, 78%, trois quarts de notre contenu vient de PubMed. Mais si on regarde plus précisément les nouveautés, on verra que ça offre un visage assez différent. Si on regarde ce qui est supérieur à 2000, on verra que ce qui vient des éditeurs est devenu majoritaire. On n'a que 45% qui viennent de PubMed. Et encore plus, si on regarde sur une année en cours, sur l'année 2017, là on voit que 80% du contenu vient des éditeurs. Grosso modo, PubMed est très lent sur le contenu non anglophone. Nous on a le contenu en direct des éditeurs et on peut proposer les choses beaucoup plus vite.

Parmi les choses qui nous distingue de PubMed et où il y a une vraie valeur ajoutée, au-delà de l'interface en français, des titres en français c'est aussi la présence de résumés en français. Aujourd'hui on en a 120 000 qui sont quasiment tous récents au sens large, qui sont des articles qui ont moins de 20 ans. Là il y a une vraie plus-value au niveau de l'information. On a beaucoup de résumés structurés, sur le modèle IMRAD, introduction, méthodes, résultats, discussion et conclusion.



Ça c'est l'image de LiSSa au début du projet, vous voyez que par tranche d'une année on avait à peu près 10 000 articles. Ce sont des données qui viennent seulement de PubMed et on peut voir qu'il y

a, en rouge, beaucoup d'articles qui n'avaient pas de liens électroniques. On trouve la référence mais on n'a pas de lien, on n'a pas de bouton, on ne peut pas poursuivre sur le site de l'éditeur ou autre pour obtenir l'article. Et puis, il y a une petite portion d'articles gratuits qui reste très minoritaire.



Ça c'est la situation au début du projet et voilà où on en est aujourd'hui. On peut voir que quasiment une grande majorité de nos articles proposent des liens électroniques. On va pouvoir aller sur le site de l'éditeur, on va pouvoir obtenir l'article soit parce qu'on y est abonné, soit parce qu'on arrive sur le paiement de l'article.

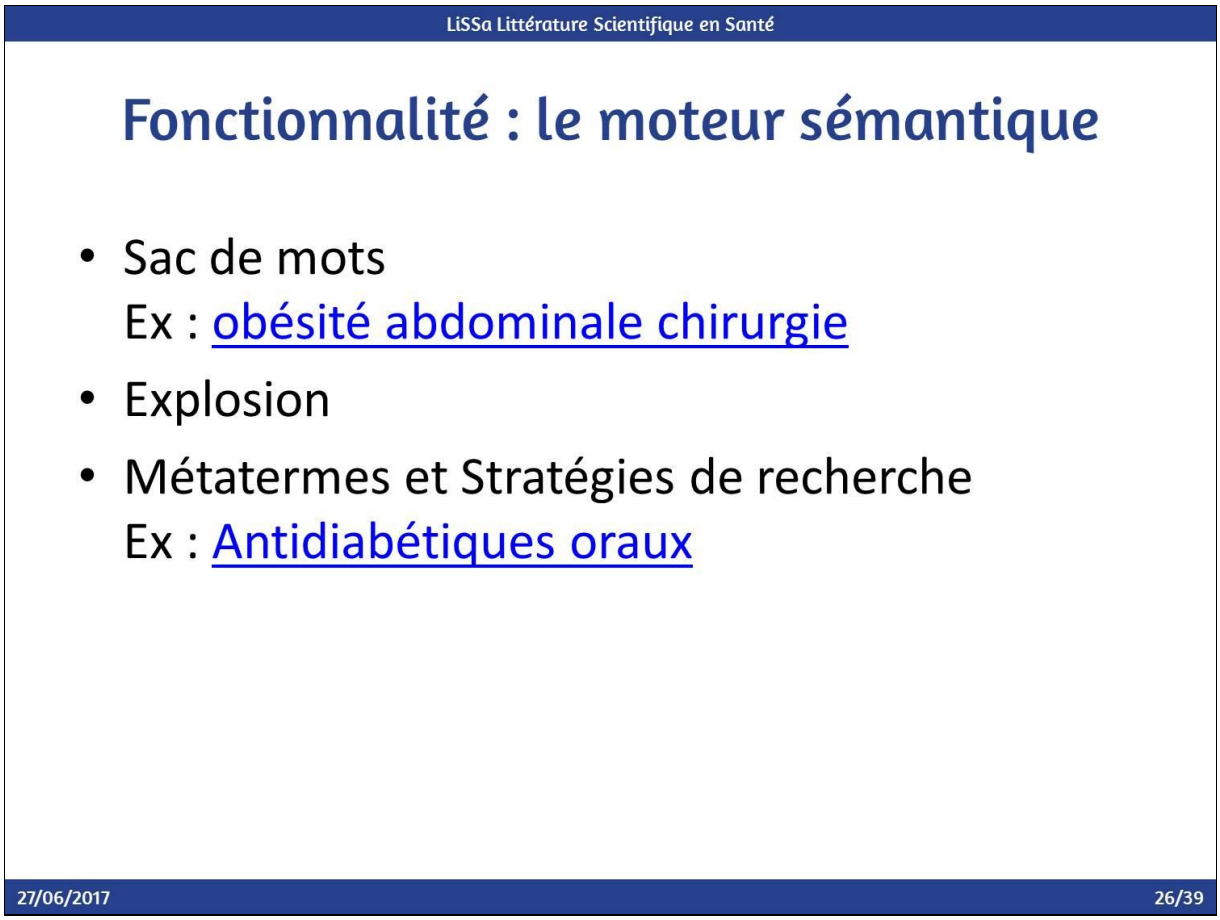
On a aussi, ce n'est peut-être pas très visible sur le schéma, bien augmenté tout ce qui était liens vers des articles gratuits. On est allé voir si les chercheurs en biomédecine déposent leurs articles sur des archives ouvertes. On en a trouvé très peu, quelques centaines à vrai dire, on continue à regarder. Mais là où on a trouvé des sources d'informations, des archives ouvertes intéressantes, c'est de la part de revues, comme la Revue Médicale de Bruxelles, des revues indépendantes qui ont fait le choix de mettre tout leur contenu gratuitement en ligne. Là on va trouver pas mal de liens gratuits. Aujourd'hui sur la totalité de la base on a 45 000 articles gratuits. Si on compare à d'autres disciplines la biomédecine, la biomédecine en français, c'est un endroit où le passage à l'open access ne s'est pas fait. Il y a d'autres disciplines où le changement de paradigme est plus avancé.

Je vais essayer de faire des démos, c'est toujours le moment un peu délicat, on croise les doigts pour que ça marche. Je vais vous présenter différentes fonctionnalités de l'outil. Tout d'abord notre moteur sémantique qu'on a développé à partir de notre savoir-faire avec CISMeF et la capacité de

notre moteur de recherche de savoir ce dont lui parle l'usager et notamment ce qu'on appelle un algorithme sac de mots. Par exemple…

>>JG

Le temps que Gaétan fasse sa requête je pense que ce qui est important c'est de bien percevoir que notre approche est antagoniste d'une approche comme celle de Google. C'est-à-dire qu'on essaye d'apporter du sens, de comprendre ce que veut dire l'utilisateur quand il tape une requête pour pouvoir avoir une précision maximale. On sait qu'on contrôle suffisamment les données, qu'elles sont en grande partie indexées pour pouvoir essayer d'avoir un « raisonnement » intelligent, essayer que la machine raisonne un petit peu sur les requêtes qui sont tapées pour pouvoir comprendre ce que veut dire l'utilisateur et donc lui donner des résultats qui sont le plus pertinents possibles.



Tout simplement ça veut dire savoir gérer la synonymie. Est-ce qu'on parle de « Syndrome de Down » ou de « trisomie 21 » ? Si on tape l'une ou l'autre de ces expressions, le moteur saura la reconnaitre. Et puis, comme ici, si l'expression n'est pas dans le bon ordre on va savoir la retrouver. J'ai tapé « chirurgie abdominale chirurgie » et on peut voir ici, si vous cliquez sur « Voir la requête effectuée », deux termes qui viennent du vocabulaire contrôlé MeSH Medical Subject Headings en français, à la fois l'obésité abdominale, là les mots étaient côte à côte c'était facile mais on reconnait aussi le terme « Chirurgie bariatrique », ça veut dire qu'il a réussi à prendre le dernier mot de ma requête utilisateur « chirurgie » et à l'associer avec le premier mot « obésité ». Et donc « Chirurgie de l'obésité » égale « Chirurgie bariatrique » et le système retourne toutes les références qui sont indexées avec ce descripteur MeSH. Ça c'est vraiment une force du moteur, repérer dans les

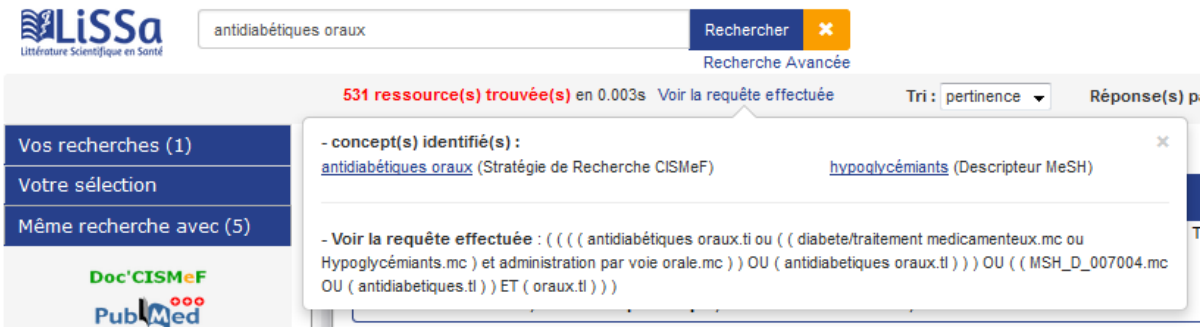
requêtes en langage naturel ce qui fait partie du vocabulaire contrôlé et donc ramener beaucoup de résultats. Bien sûr on a aussi un rappel plus large en cherchant dans le titre, le résumé les mots ensemble, les mots séparément. Cela nous permet d'avoir une relative confiance dans ce que nous rapporte le moteur.

L'explosion est un principe qui existe dans PubMed, qui est sûrement très familier aux documentalistes. Si on reconnait un terme MeSH de haut niveau, le système va automatiquement aller chercher tout ce qui est plus bas dans la hiérarchie MeSH. Si je tape « maladies pulmonaires » on va chercher ce qui parle des maladies pulmonaires en général mais on va aussi chercher chaque document sur chaque type de maladies pulmonaires. Entre « maladies pulmonaires » et « asthme » il y a peu de lettres en commun mais on aura quand même les articles sur l'asthme qui vont nous être retournés. Pour que cela ne perturbe pas les utilisateurs, on se sert aussi du tri pour que ces résultats, s'ils sont un peu moins intéressants, n'arrivent pas en premier. Ça c'est ce que sait faire le MeSH, c'est le langage de PubMed.

Dans l'équipe CISMeF on a aussi beaucoup travaillé sur le vocabulaire contrôlé et apporter notre propre gestion des connaissances avec deux notions importantes.

Tout d'abord les métatermes qui sont les disciplines médicales. On s'est assuré d'avoir un bien meilleur rappel pour tout ce qui est spécialité médicale. Dans PubMed si vous tapez « cardiology » vous aurez les articles qui parlent de la discipline, comment elle se pratique etc. Vous n'aurez pas tous les articles du domaine. Dans LiSSa, si vous tapez « cardiologie », cela vous ramènera ce qui parle de la discipline en tant que telle mais aussi tous les articles sur les maladies du cœur, tous les articles sur les méthodes de diagnostic dans cette discipline etc. On est sûr d'avoir un bien meilleur rappel pour les spécialités médicales.

Et puis on a créé ce qu'on appelle les stratégies de recherche. Là c'est pour des notions bien particulières qui n'existent pas dans le MeSH ou qui existent dans le MeSH au travers de nombreux mots-clés. Ici on a thésaurisé, on a compilé le savoir-faire de nos documentalistes. Par exemple, si vous tapez « antidiabétiques oraux » dans LiSSa – on va faire la démo – ce sera automatiquement remplacé par une requête « savante » en quelque sorte « experte ».



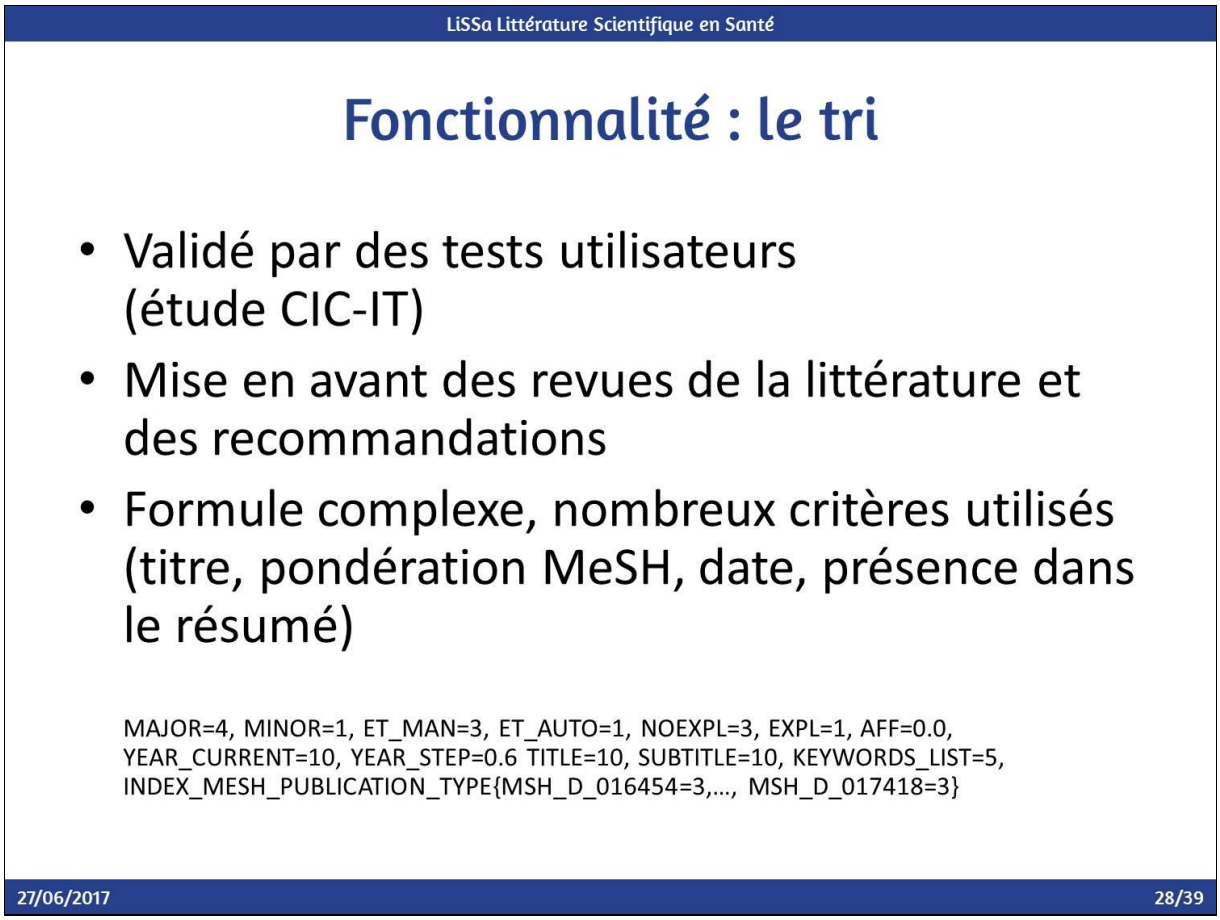
On peut voir que le moteur de recherche va chercher « antidiabétiques oraux » dans le titre mais il va aussi savoir que des mots-clés MeSH peuvent correspondre, comme « diabète/traitement médicamenteux » ou le mot-clé « hypoglycémiant », croisés avec le mot-clé « administration par voie orale ». Ici, en l'occurrence le savoir-faire d'une documentaliste, une vraie requête experte qui sera transparente pour l'utilisateur lambda. Donc tout ce savoir-faire, développé pour CISMeF, se retrouve dans LiSSa.

Une autre fonctionnalité : dans notre jargon on appelle ça des facettes ou tout simplement des filtres. C'est tout ce qu'on va trouver à la gauche de l'écran et qui va permettre d'affiner une recherche. Cela permet à un utilisateur non-averti, qui ne connaît pas nos métadonnées, tous les champs d'informations qu'on peut proposer, d'avoir un accès assez facile à ces informations pour filtrer sa requête, tout simplement sur l'année, sur la revue, sur les autres mots-clés qui peuvent être intéressants sur le sujet… [Démo] On va choisir une revue en particulier

Les progrès techniques font qu'aujourd'hui si on coche une option, les chiffres des facettes sont recalculés à la volée. [Démo]. Pour les habitués de PubMed, là on est assez différents. Dans PubMed les filtres proposés sont des filtres statiques. On rajoute des bouts de requêtes, ce n'est pas recalculé comme ici.

>>Stéfan Darmoni

À leur décharge, ils ont 27 millions de références. Bien plus que nous on va être honnêtes…

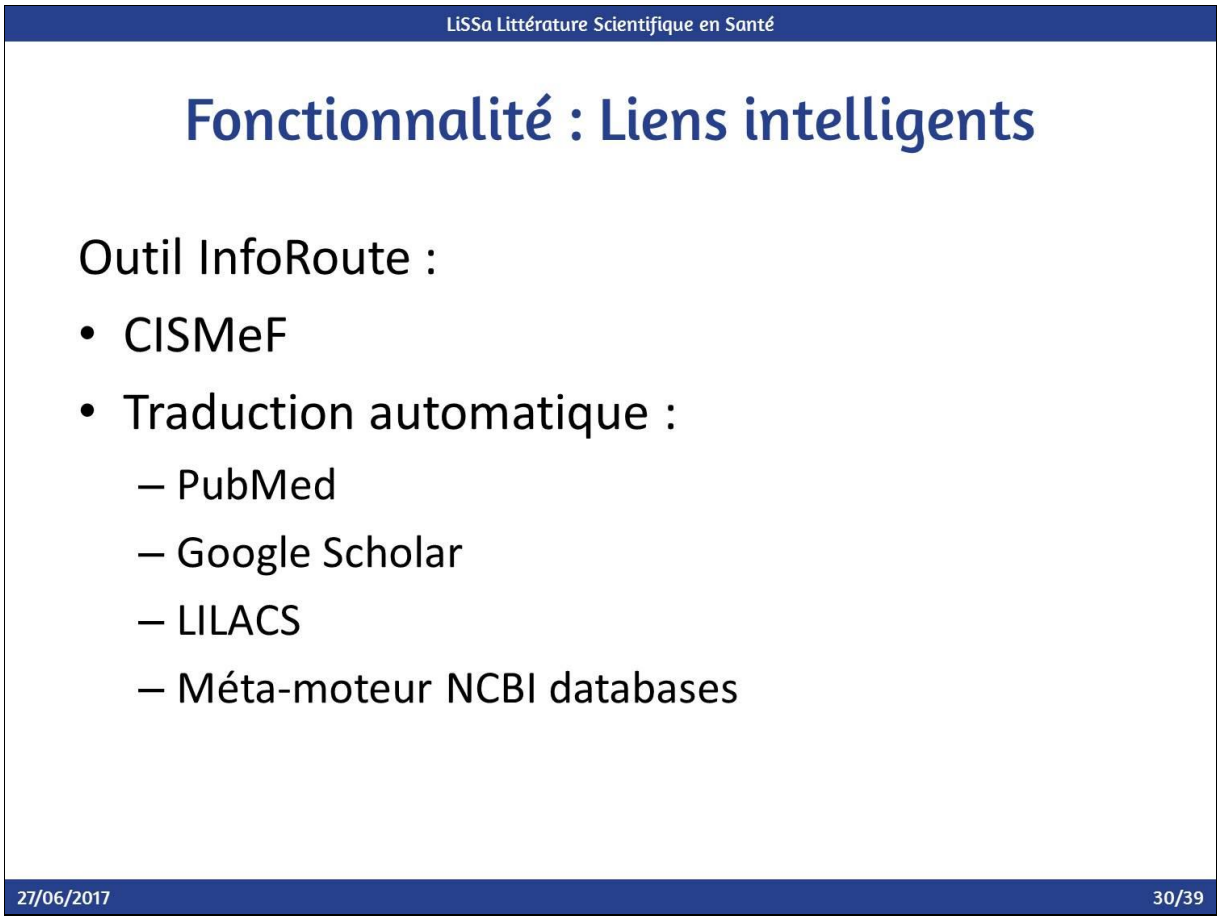


Ce qu'on a implémenté aussi c'est un tri qu'on espère pertinent et ça on l'a fait valider avec des utilisateurs comme je le disais tout à l'heure. Ce qu'on a choisi de faire c'est de mettre en avant ce qui nous semblait le plus utile dans ce qui était francophone. On a vu que les besoins étaient autour de … à la fois ce que disaient les revues et ce que nous disaient les utilisateurs, c'est-à-dire que le français est vraiment très utile comme véhicule de la formation continue et de la synthèse des connaissances et donc on a voulu mettre ça en avant dans notre moteur de recherche. Donc tout ce qui est revue de la littérature ou recommandations, tout ça va avoir un petit plus pour être proposé plus haut dans les résultats. Si il y a une reco' de l'année sur mon sujet, je l'aurai en premier, je n'aurai pas besoin de la chercher. Ça c'est quelque chose qu'on a vraiment voulu implémenter.

On a créé une formule assez complexe qui nous permet de faire des réglages très précis sur tout un tas de critères. La formule est assez peu lisible pour le commun des mortels mais cela nous permet de faire plein de réglages fins. Aujourd'hui, avec les ergonomes, on discute encore d'affiner ce tri par pertinence. C'est un sujet de recherche sur lequel on espère publier.

Le moteur de recherche nous permet de faire des exports. On a voulu être compatible avec les logiciels de gestion bibliographique. Le but de LiSSa est d'aider à ce que les articles soient cités, il faut participer à l'automatisation du processus au maximum. On a choisi des exports au format RIS qui est un format compatible avec les logiciels de gestion bibliographique les plus connus. L'offre commerciale la plus importante c'est EndNote et l'offre gratuite alternative c'est Zotero. Ils sont bien connus des chercheurs et des étudiants désormais. En 1 ou 2 clics maximum on pourra importer ses références bibliographiques et ensuite les citer, les mettre en forme proprement dans un document, dans une thèse, un mémoire, un article… On a aussi écouté les demandes autour de nous pour ces exports. Des documentalistes nous ont demandé des exports sous forme de tableau compatible avec Excel ou d'autres logiciels de type tableur. Et puis la possibilité de créer des pages web simplifiée qui pourront servir de « mode rapport » ou qui pourront être copiées-collées facilement dans Word. [Démo] On a voulu créer une page Web à la volée avec une URL pérenne. Ce lien peut être copié, envoyé à quelqu'un, il retrouvera les documents que vous avez sélectionnés.

Donc le plus possible ne pas être coincé sur l'outil, pouvoir faire quelque chose des articles, les faire suivre, les passer dans un logiciel, les exporter, les envoyer…



Quelque chose qu'on vient de rétablir dans l'outil, c'est un service qu'on appelle Inforoute qui va permettre de lier notre base de données avec plein d'autres sources d'information. On parle beaucoup de l'important du français depuis tout à l'heure, on ne voit pas le français comme un univers fermé. On le voit comme un premier pas, peut-être, vers la littérature internationale et donc on essaie de faire le pont vers plein d'autres bases de données. [Démo] C'est une zone qu'on appelle « même recherche avec ». On va se lier avec Doc'CISMeF, le moteur de recherche de CISMeF, pour avoir les ressources francophones en accès libre, type recommandations de la HAS ou autre.

[Démo] Ici la requête utilisateur en langage naturel en français va être compris par notre moteur de recherche, qui va la raccrocher à du vocabulaire contrôlé qui lui-même est multilingue et donc on va pouvoir proposer un lien en anglais vers PubMed. Ici on peut voir que tous les termes de recherche sont traduits. Non seulement on utilise les termes MeSH reconnus mais on a mis ici un peu de notre savoir-faire : on va ajouter tout un tas de synonymes, tout un tas de termes pertinents qui seront cherchés dans le titre et le résumé pour avoir des résultats de recherche qui seront justes.

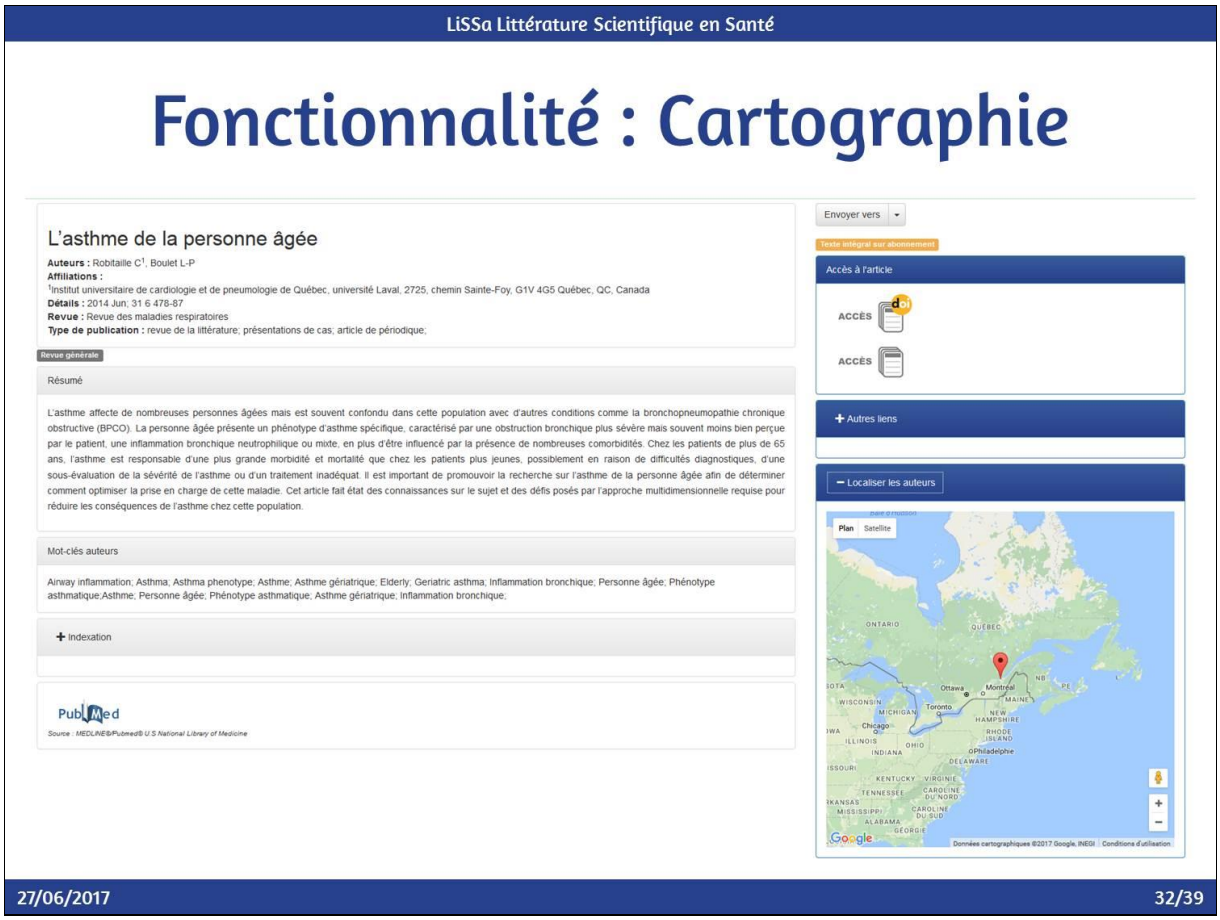
On le fait pour PubMed, on ne se veut pas en concurrence de PubMed, on se veut en complément et même en point d'accès si besoin. On propose un lien vers Google Scholar, qui est une version de Google limitée à des ressources sélectionnées. Une autre base de données qu'il nous a semblé très important d'inclure, c'est la base de données LILACS qui est plus ancienne que LiSSa et qui repose sur la même logique. C'est une base de données qui référence des ressources en langue espagnole et en portugais. Ils ont fait des choix très semblable : se raccrocher au MeSH et à des vocabulaires contrôlés traduits internationalement. On peut les interroger à partir de notre base. C'est tout un pas de la littérature internationale qui s'ouvre auquel on n'avait pas accès a priori.

Au sein du NCBI, l'organisme qui chapote PubMed, il y a tout un tas de bases de données et donc il propose un méta-moteur de livres de génomiques etc. On donne accès à ce moteur.

On propose un formulaire de recherche avancée. On a mis beaucoup d'efforts sur la recherche en langage naturel pour l'utilisateur lambda mais on essaie aussi de satisfaire les documentalistes, les bibliothécaires ou même des utilisateurs pointus. On en connait dans toutes les disciplines.

Avec ce formulaire on va faciliter l'usage des opérateurs booléens. « ET » « OU » « SAUF », on les connaît bien, mais utiliser les parenthèses c'est parfois assez fastidieux donc on a un formulaire qui rend les choses assez simples. L'usager n'aura pas besoin de gérer les parenthèses. On aura les « ou » en colonnes et les « ET » par lignes. Cela permettra d'exploiter les métadonnées assez finement. Dans les listes déroulantes, on aura tous nos champs d'information.

Si on a besoin de faire une recherche précise sur un journal, sur un nom d'auteur, on pourra le faire ici et, également, utiliser les opérateurs booléens. [Démo] J'ai tapé « sport » et on a vu que plusieurs mots-clés pouvaient correspondre « sports » « athlètes » etc. Si je veux englober tout ça, « faire des OU », je pourrai le faire facilement ici. On peut aussi utiliser des troncatures, avec l'étoile.

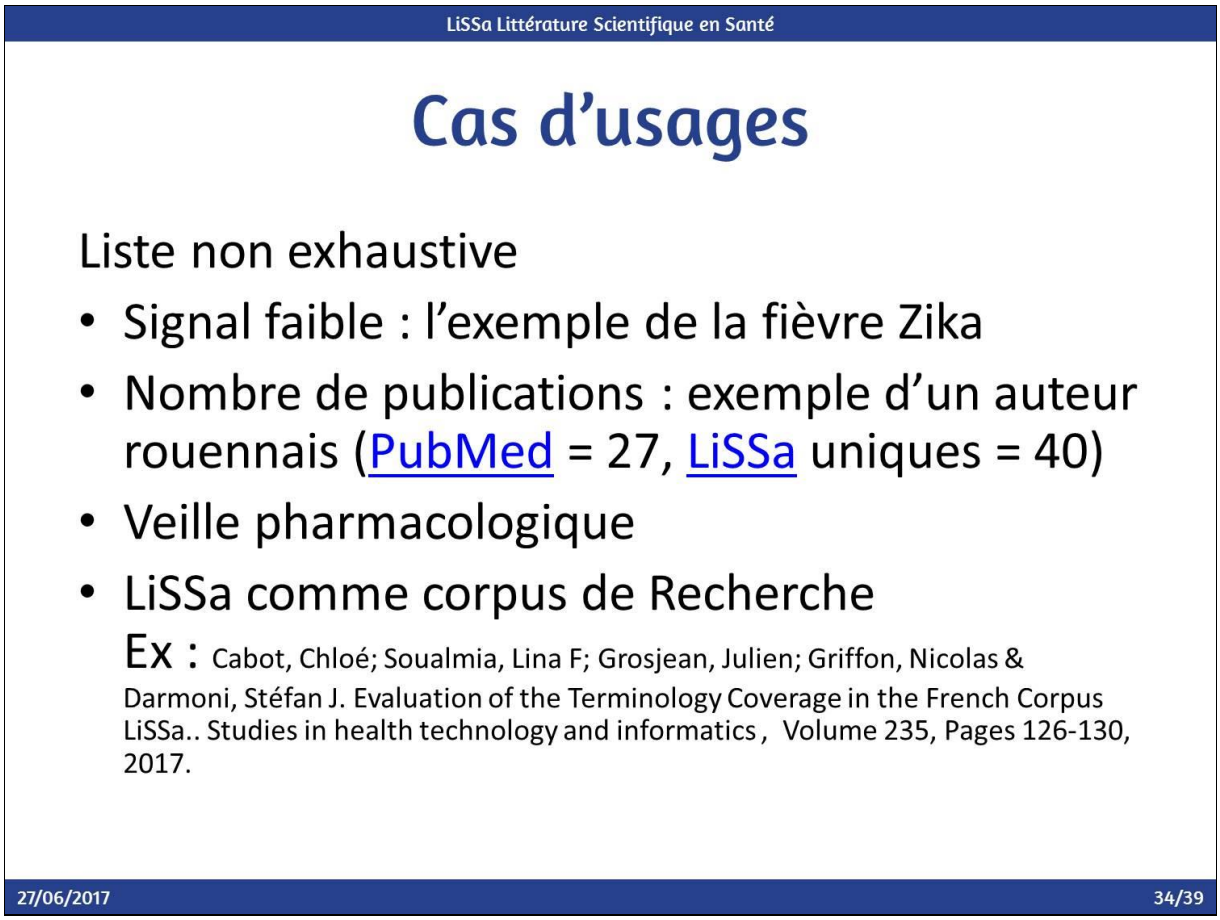


Pour chaque référence, on peut voir les résumés dont je parlais, les différents accès électroniques, quand ils existent. On a aussi implémenté une fonctionnalité : une carte qui affiche l'origine des résultats, l'origine des auteurs. On parle de la francophonie, donc il n'y aura pas que des auteurs français. S'il y a plusieurs auteurs avec plusieurs affiliations, elles apparaissent toutes sur la carte.

Fonctionnalité : Score bibliométrique

- SCImago Journal Rank (SJR indicator) est une mesure de l'influence scientifique de revues académiques qui tient compte à la fois du nombre de citations obtenues par la revue et de l'importance des journaux d'où proviennent ces citations.

Sur cette même page, on a aussi commencé à donner quelques informations bibliométriques. Pour cela on a utilisé une alternative au facteur d'impact qui s'appelle Scimago JournalRank OU SJR Indicator. C'est une mesure de l'influence de chaque revue qui prend en compte le nombre de citations et qui repose sur une réflexion plus poussée que l'IF dans le sens ou cela prend en compte la notoriété de la revue qui émet la citation. Si votre article reçoit une citation de « Nature », cela n'aura pas le même poids que si c'est une citation qui provient d'un autre journal de moindre influence. Donc c'est un score un peu différent. Il y a des limites, ce score n'est pas encore bien connu donc il parle beaucoup moins que le facteur d'impact et on ne peut pas le proposer encore pour toutes les revues francophones car il y a beaucoup de revues qui ne sont pas connues encore de ce service. Néanmoins, on a 200 000 articles qui proposent ce score.



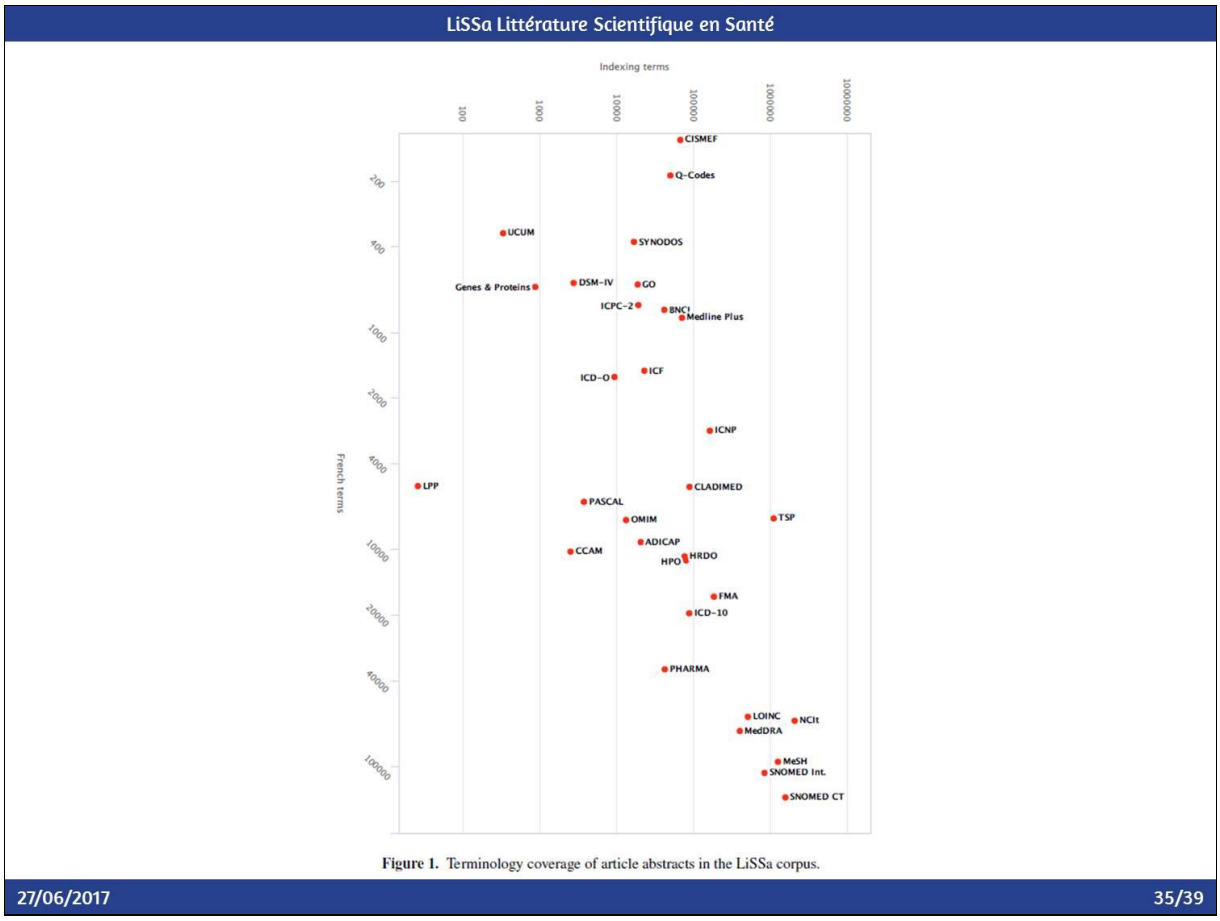
Voilà pour le tour d'horizon des fonctionnalités qu'on propose aujourd'hui. Je vais vous donner quelques cas d'usages, quelques utilisations qu'on a rencontrées. Et je donnerai la parole à des bibliothécaires de la salle qui ont des exemples à nous montrer de comment LiSSa est utilisé.

Je ne vais pas tout détailler. Sur certains cas, des choses paraissent en français avant d'être publié dans d'autres langues. Par exemple la fièvre Zika, l'épidémie en Polynésie, on a une référence francophone qui précède d'un mois celle qu'on peut trouver dans PubMed. Pour détecter des signaux faibles, cela peut être intéressant.

Si on regarde l'évaluation des chercheurs, on peut voir que prendre en compte le français pour certaines disciplines peut changer beaucoup de choses.

Par exemple, un chercheur rouennais qui travaille en éducation thérapeutique a 27 citations connues dans PubMed. Si on regarde ce qu'il y a dans LiSSa (hors PubMed), il y a 40 citations en plus. Pour certaines disciplines, pour certains auteurs, si on prenait en compte ces publications francophones cela changerait énormément de choses dans leur visibilité et, peut-être, dans la façon dont on les considère.

Des choses qu'on n'avait pas anticipés mais qui apparaissent désormais et que LiSSa peut avoir un rôle dans la veille pharmacologique. On a eu des contacts avec des industriels. Savoir ce qui est publié sur tel médicament, sur telle substance ça peut être important. Ils doivent surveiller ce qui est publié. Et donc LiSSa peut avoir un rôle.

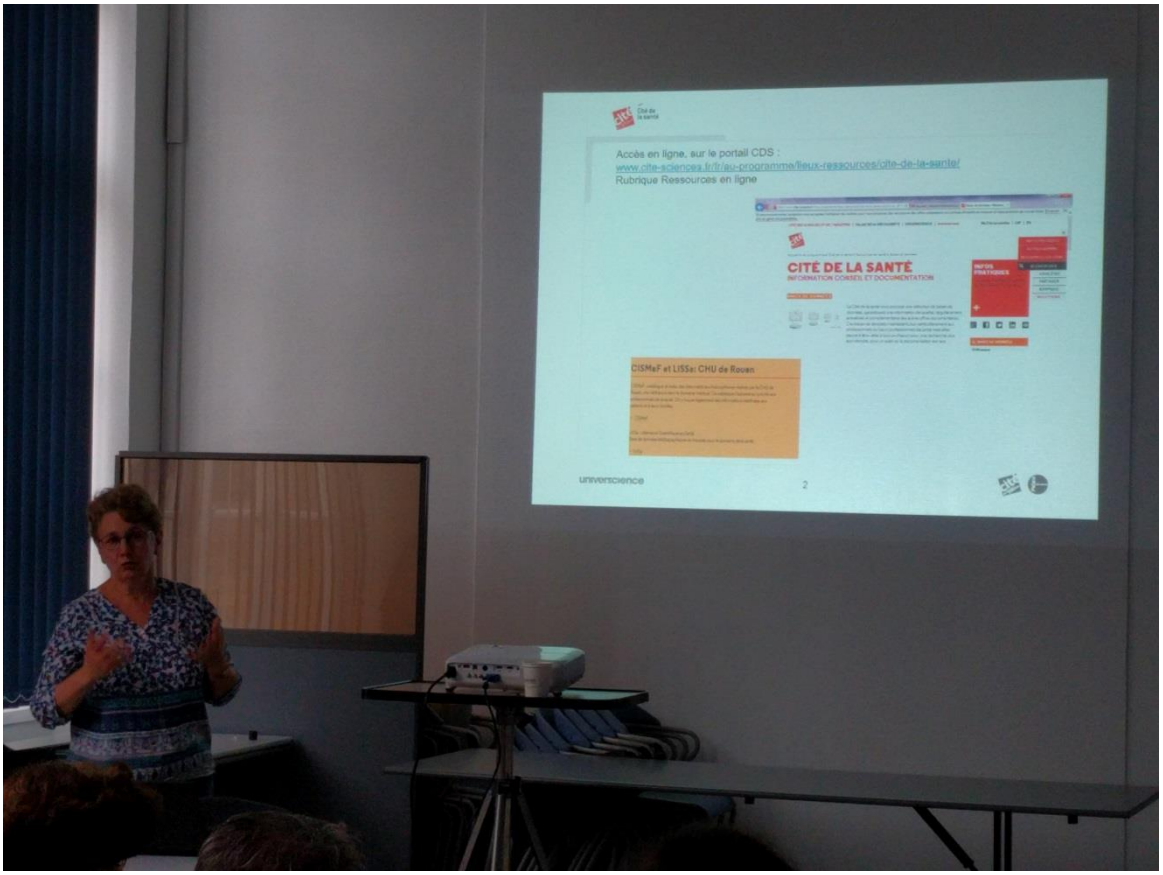


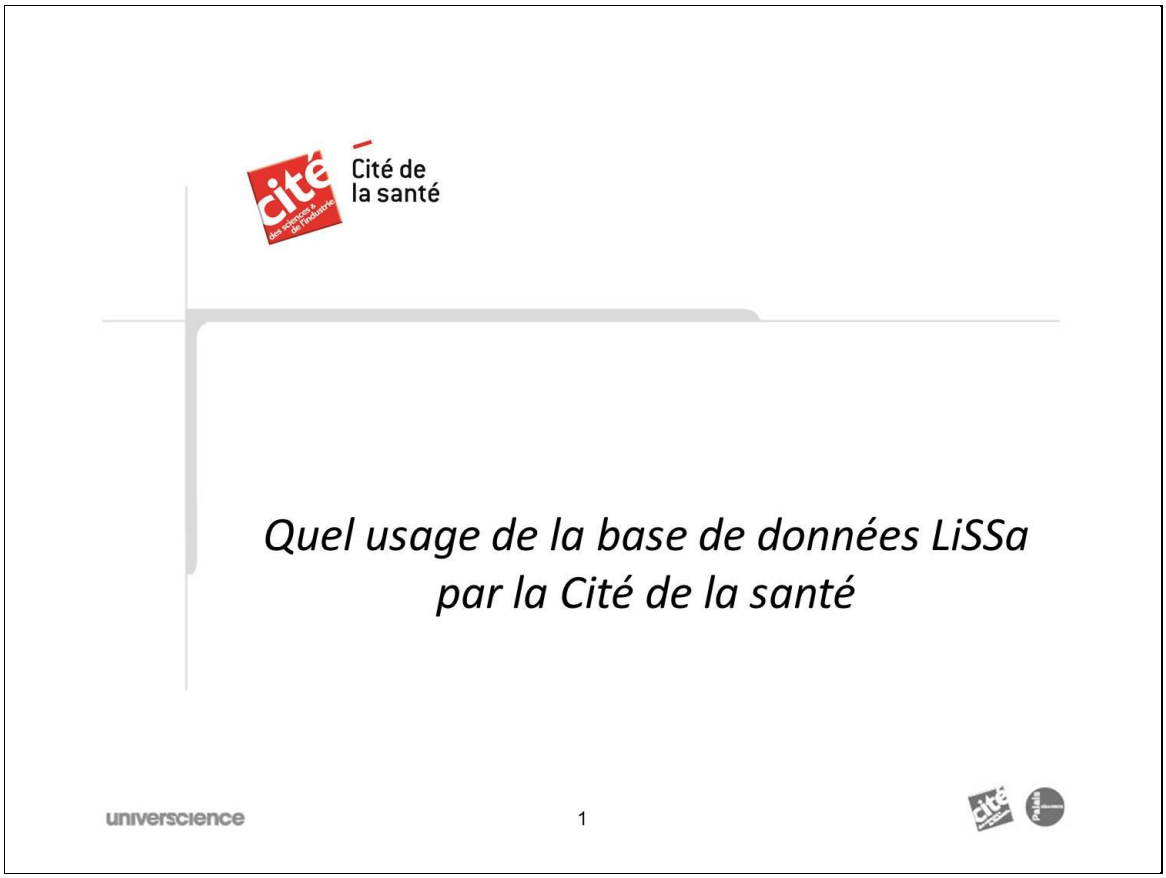
Un dernier exemple pour terminer. On a dit tout à l'heure que LiSSa pouvait être un support pour la Recherche, on a une première publication qui a été présenté au congrès MIE[1]. On a pris LiSSa comme corpus, on a testé les différentes terminologies médicales et on a essayé de voir lesquels fonctionnaient le mieux pour une indexation automatique. On attendait le MeSH mais ce n'était pas le MeSH la meilleure terminologie, on se rend compte que c'est le NCIT, le thésaurus du National Cancer Institute, suivi par la SNOMED CT. Donc on peut découvrir des choses grâce à LiSSa. J'ai un autre exemple, on a été contacté par un étudiant, qui est à Belfast en ce moment, et qui voudrait savoir tous les essais cliniques qui ont été publiés en français en 2014 et 2015, la liste la plus exhaustive possible. On peut utiliser LiSSa pour trouver la meilleure liste possible de ces publications.

Voilà pour nos cas d'usages, je voudrai donner la parole à Mme Lebastard de la cité de la Santé si elle le veut bien.

[1] Cabot, Chloé; Soualmia, Lina F; Grosjean, Julien; Griffon, Nicolas & Darmoni, Stéfan J. Evaluation of the Terminology Coverage in the French Corpus LiSSa.. Studies in health technology and informatics , Volume 235, Pages 126-130, 2017.

Intervention de Sandrine Lebastard, documentaliste à la Cité de la santé





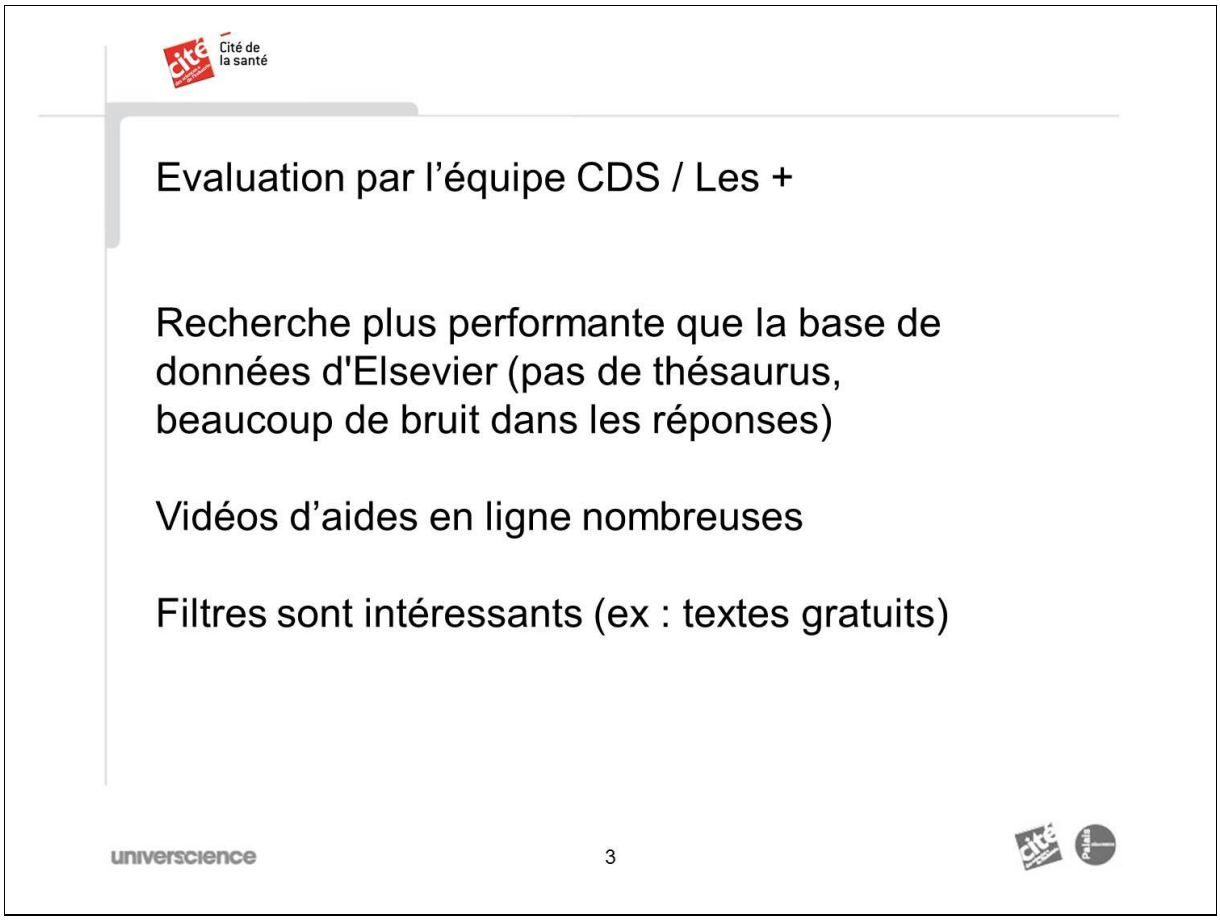
Bonjour. Je vais vous présenter ce que c'est La Cité de la Santé car vous ne connaissez peut-être pas. C'est un lieu d'information et de documentation en Santé qui se trouve dans la cité des sciences. La Cité de la Santé a été créée en 2002 sur le modèle de la cité des métiers. Vous avez des ressources, parce que nous sommes dans la bibliothèque, et également des espaces de conseil.

Nous utilisons LiSSa pour un autre service, pour répondre au public qui vient nous interroger directement dans la bibliothèque, mais également pour un service de questions-réponses en ligne, qui a été créé en 2008, qui s'appelle Questions Santé. Nous répondons au public, à toute question documentaire qui est posée sur la Santé.

J'ai interrogé mes collègues, c'est une petite présentation de notre ressenti à l'usage de cette base de données.

Là vous avez la présentation de notre portail, vous le trouvez sur Google, il suffit de taper Cité de la Santé pour tomber sur notre service.

LiSSa est présente dans les bases de données, ainsi que CISMeF, qu'on utilise aussi beaucoup je dois dire. On a regroupé les deux dans les bases de données.



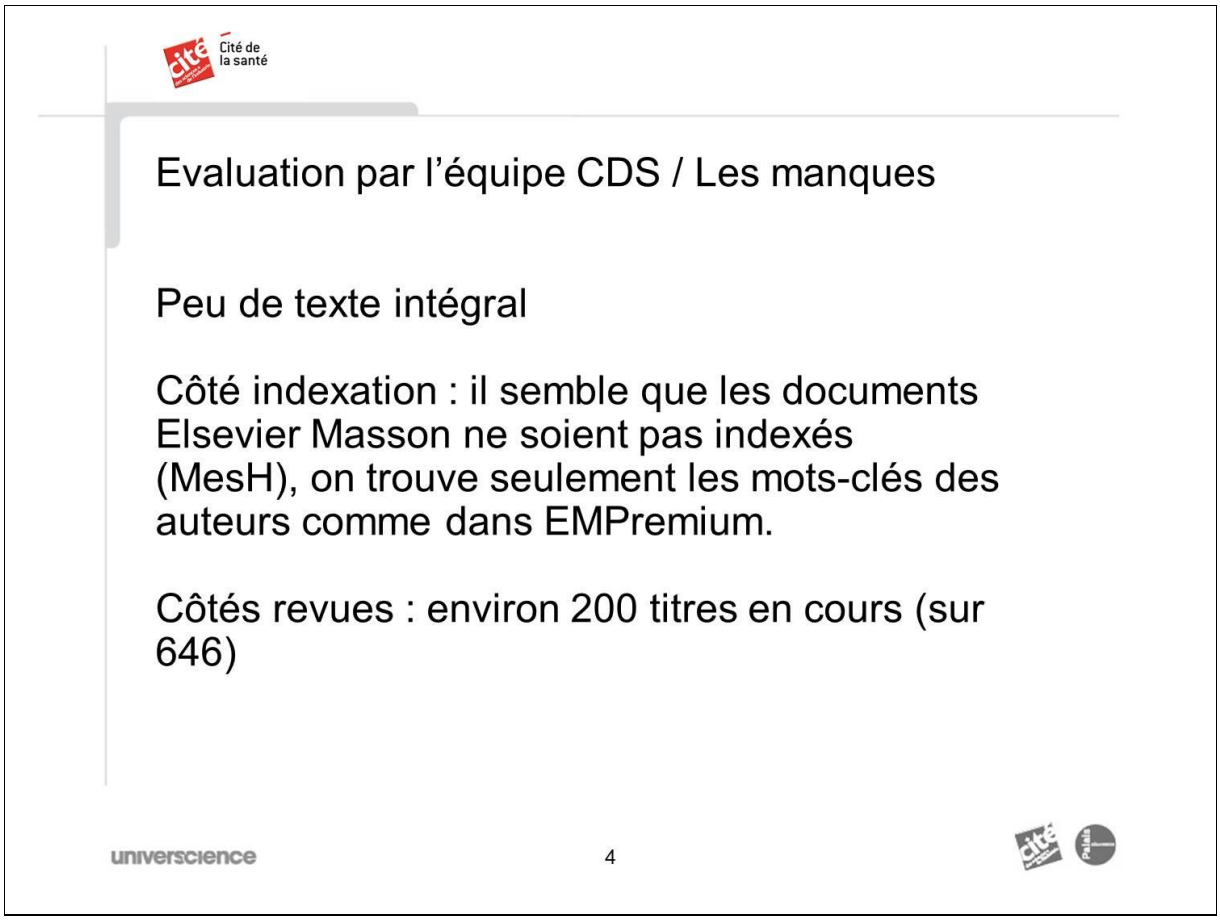
Les « plus » que nous avons vu dans LiSSa.

Nous nous sommes abonnés à EM-Premium et nous avons les packs infirmiers car nous recevons beaucoup d'élèves infirmières qui viennent pour leur mémoire de fin d'études. Nous avons aussi l'ensemble des traités de l'Encyclopédie Médico-chirurgicale ainsi qu'un certain nombre de revues en ligne. Donc on utilise aussi pas mal Elsevier.

La recherche est plus performante car sur EM Premium la recherche est assez capricieuse, on a des tonnes de résultats, on a beaucoup de bruit. C'est vrai que LiSSa réduit les résultats et c'est tout de suite plus facile.

Il y a beaucoup de vidéos d'aide en ligne, de tutoriels sur LiSSa.

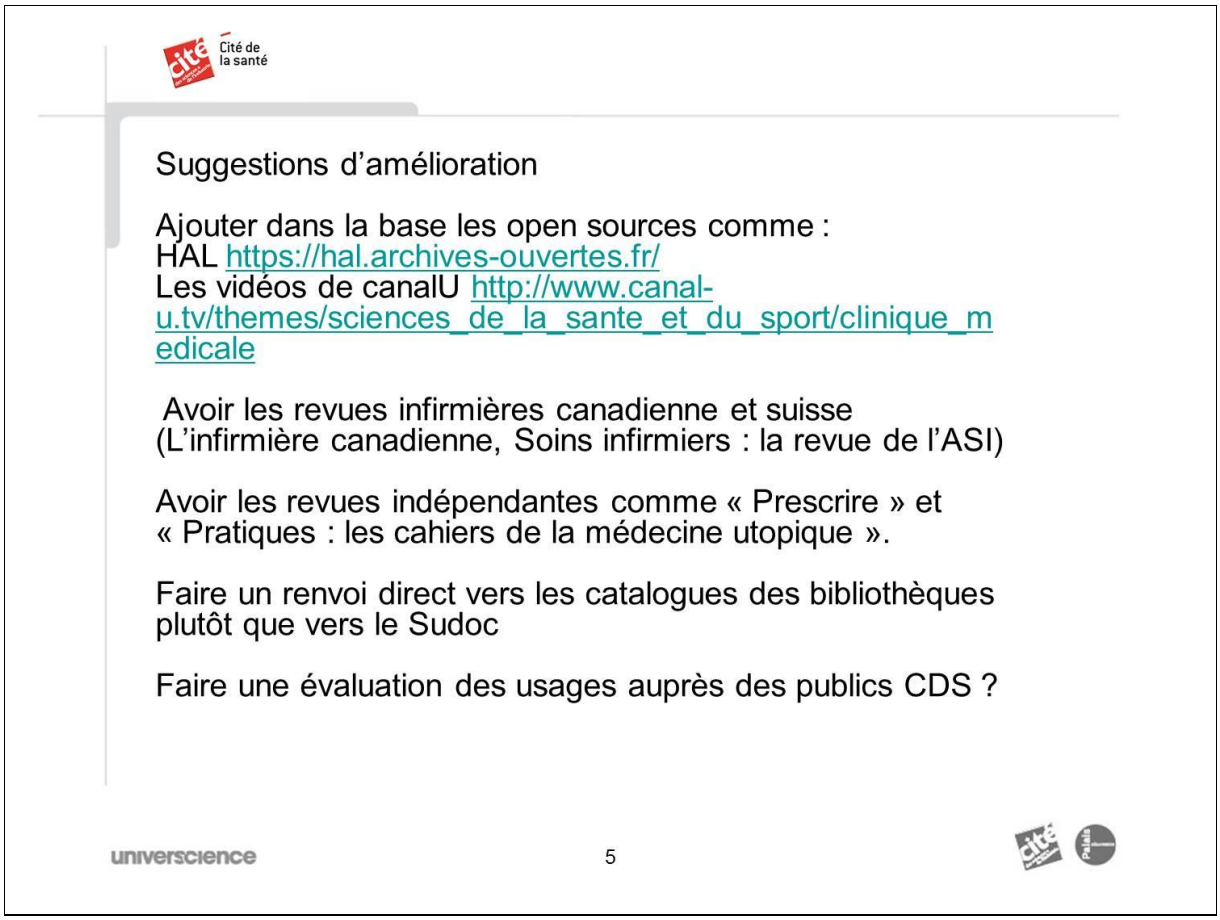
Les filtres sont intéressants, notamment de savoir qu'il y a les textes gratuits puisque nous recherchons principalement des textes pour envoyer à nos internautes pour qu'ils aient la réponse à leur question. C'est intéressant de pouvoir sélectionner ce qui est payant et ce qui est accessible gratuitement.



Les manques :

- Il y a peu de texte intégral. Quand on fait une recherche on trouve beaucoup d'articles payants. C'est Elsevier Masson principalement donc c'est sûr qu'il y a la banque de données qui est à acheter. On arrive à s'en sortir parce qu'on donne des résultats issus de EM consulte, on met un peu de contenu de EM Premium donc on arrive à faire des réponses. Mais s'il y avait plus de texte intégral, on utiliserait plus LiSSa.
- Côté indexation, les documents Elsevier Masson ne sont pas forcément indexés en MeSH, on ne trouve que les mots clés des auteurs, comme dans EM Premium. Donc ce n'est pas un plus.

Au niveau des titres aussi, on a regardé les titres et on a vu que sur les 646, il n'y a que 200 titres en cours. Je pense que c'est intéressant pour des personnes qui font des recherches en Histoire mais quand on veut avoir des articles, ça se résume tout de même à 200 titres. Donc il faut nuancer les 646.



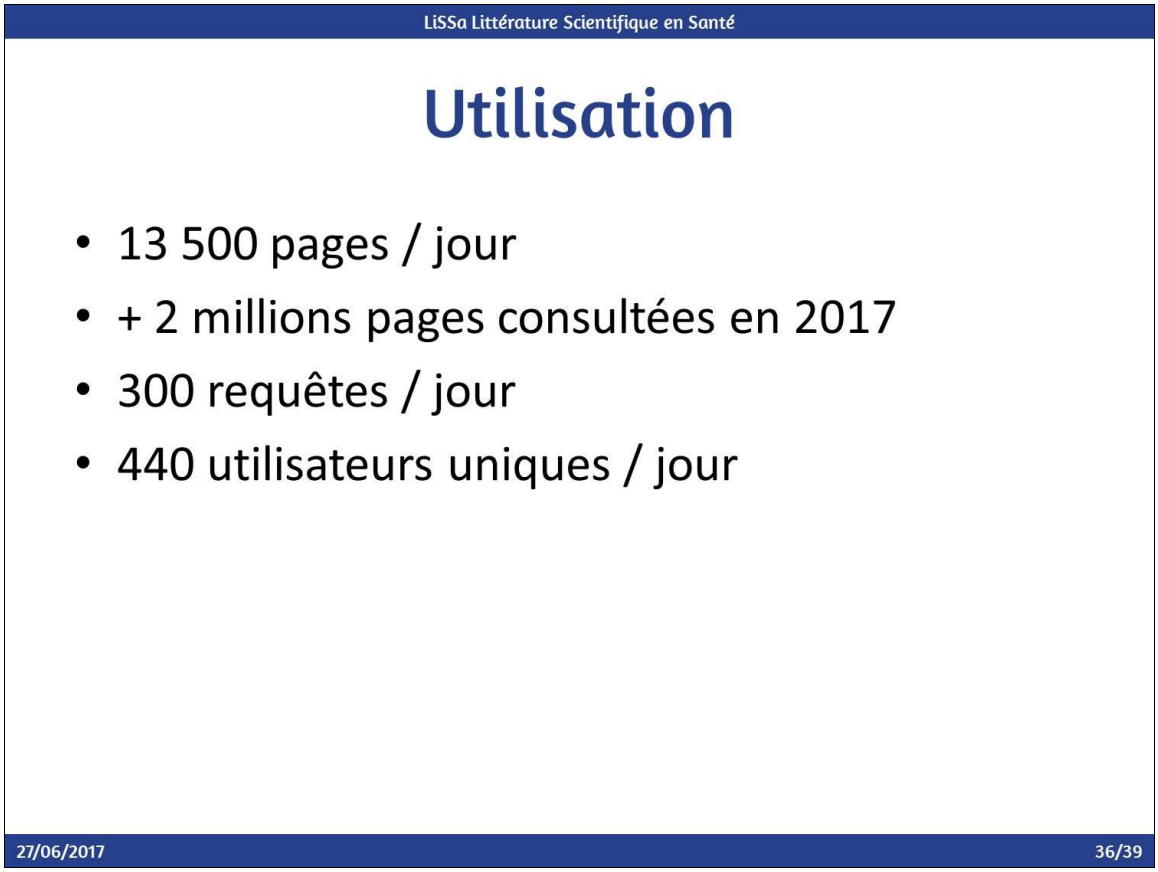
Les suggestions d'améliorations, ce seraient d'intégrer les open-sources comme par exemple les archives ouvertes, quand on fait la recherche de tomber aussi sur ce corpus-là. Une collègue parlait aussi des vidéos, on a parlé que d'articles, ce serait un nouveau support…

Cette même collègue suggérait d'inclure des revues autres, comme L'Infirmière Canadienne, La revue suisse des soins infirmiers. On est souvent tourné vers des revues pour les élèves infirmières. On en a beaucoup donc on a ce point de vue un peu spécifique. Et puis aussi avoir des revues indépendantes, comme Prescrire mais aussi Pratiques, les cahiers de la médecine utopique.

Une petite remarque aussi sur le renvoi. Quand on est sur un article on peut voir la disponibilité en bibliothèques. En cliquant on arrive sur le Sudoc, ce serait bien si on arrivait directement sur le catalogue de la bibliothèque.

Pour finir, nous proposons de faire une évaluation auprès de nos publics à la Cité de la Santé.

Gaétan Kerdelhué (suite)



Merci beaucoup. Rapidement pour finir, où en sommes-nous en termes d'utilisation de la base ?

Tout d'abord, le plus précieux pour nous ce sont ces retours directs et savoir qu'on commence à être connu de la communauté des bibliothécaires et des documentalistes car on sait que ce seront eux nos porte-paroles et que, si on a pas vocation tout de suite à être connu du citoyen lambda, on peut compter sur ces relais qui sont très importants pour nous faire connaître auprès des professionnels de santé et des étudiants. Aujourd'hui nous avons 13 500 pages consultées par jour. Pour chaque référence d'article on a une page web, c'est ça qui fait vivre ce trafic. 300 requêtes dans le moteur par jour. On sait qu'il y a environ 440 utilisateurs uniques par jour qui consulte le moteur.

On essaye de diffuser :

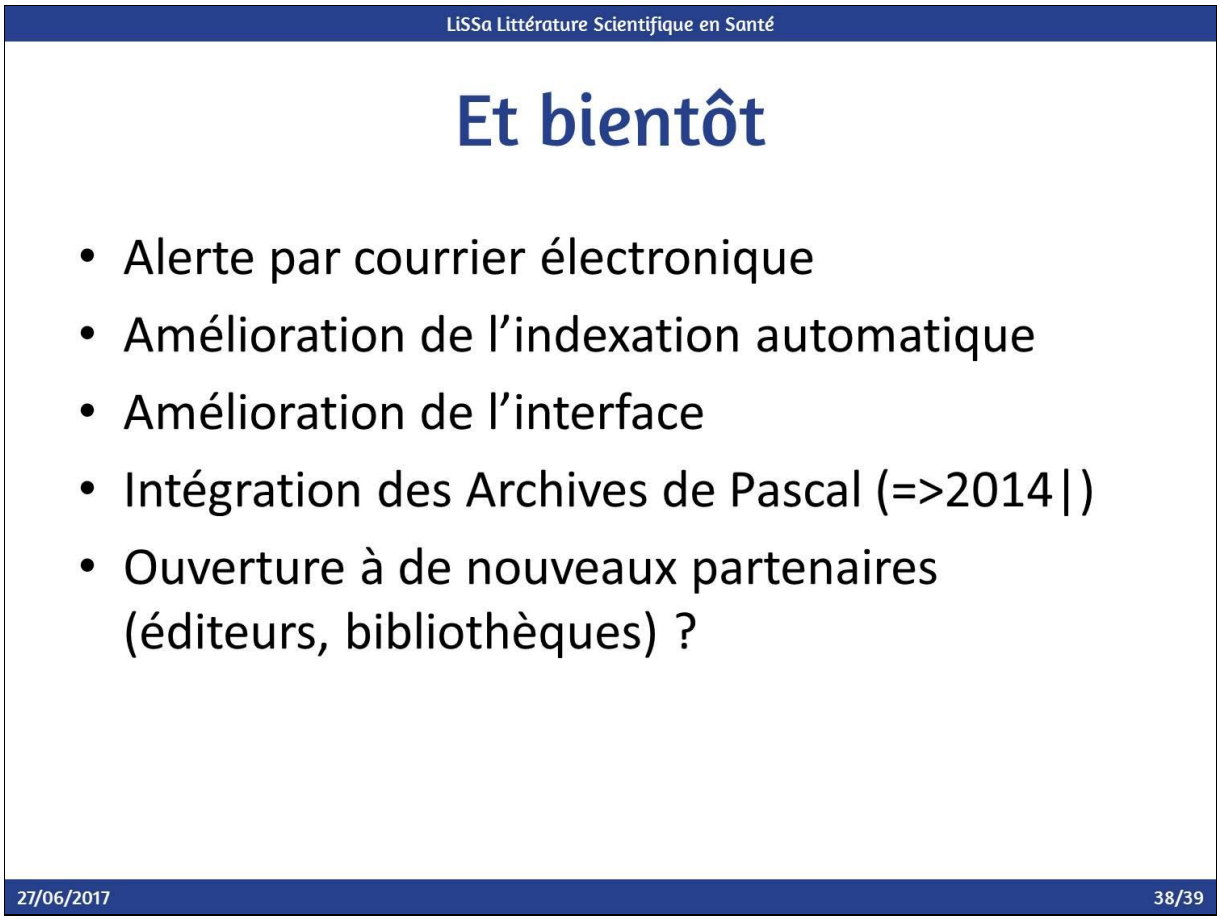
- Aujourd'hui est un jalon important pour nous.
- On a aussi des publications scientifiques, en langue française et un peu en anglais bien sûr.
- Des présentations lors de journées ou conférences. Dans le monde des bibliothèques. JNBDS c'est pour Journée nationale des bibliothécaires et documentalistes en santé à laquelle on a présenté le poster qui est là et vous avez peut-être pu avoir le flyer. En octobre on aura le réseau national des documentalistes hospitaliers. Le Pr Gehanno a présenté LiSSa auprès de la société française de médecine du travail.
- On a aussi des formations. On s'est déplacé à Lille à l'Université Catholique de Lille où on a reçu différents professionnels de Santé. On en parle à l'ADBS qui est une association de documentalistes. Bien sûr dans nos cours à l'université de Rouen. On a un contact pour Lyon très bientôt
- On essaye de multiplier les supports. On a parlé des vidéos. On a un écho, très limité, sur les réseaux sociaux. Notamment Elsevier Masson a mentionné LiSSa sur son blog, sur sa page Facebook.

>>Stéfan Darmoni

On attend aussi des bandeaux publicitaires dans les revues Elsevier.

>>Phil Woodall (Elsevier Masson

Oui je peux confirmer les encarts ont été publiés dans, au minimum, une vingtaine de revues



Dans les choses qu'on voudrait pour bientôt :

- des alertes par courrier électronique. Un utilisateur pourra s'inscrire sur le site, taper ses mots-clés, avoir ses résultats tout de suite et puis avoir des résultats qui lui seront envoyés. Jusqu'à présent on faisait du « Pull » et maintenant on poussera aussi l'information, faire du « Push ».
- On va améliorer l'indexation automatique. On en a parlé.
- On a toutes ces remarques des ergonomes qu'on est en train d'implémenter au fur et à mesure. Il faut s'attendre à pas mal de petites améliorations, un peu partout, de l'interface.

En termes de contenu, un fonds important vient d'arriver. Je pense que certains sont familiers de la base de données PASCAL qui n'est plus alimenté depuis quelques années, depuis 2014. PASCAL, financé par l'INIST a fait le choix de mettre toutes ces archives en accès libre à travers un site mais aussi en fournissant des données. Donc nous allons avoir plusieurs centaines de milliers d'articles en français, en biomédecine qui vont arriver par ce biais-là.[2]

>>Stéfan Darmoni

C'est ancien mais nous avons tout de même pas mal de gens qui cherchent des informations pour des maladies assez rares où les articles sont assez anciens. Ce n'est pas neutre.

>>Gaétan Kerdelhué

[2] Réalisé en septembre2017

Selon le sujet une information de 2010 pourra être périmée ou encore très valable.

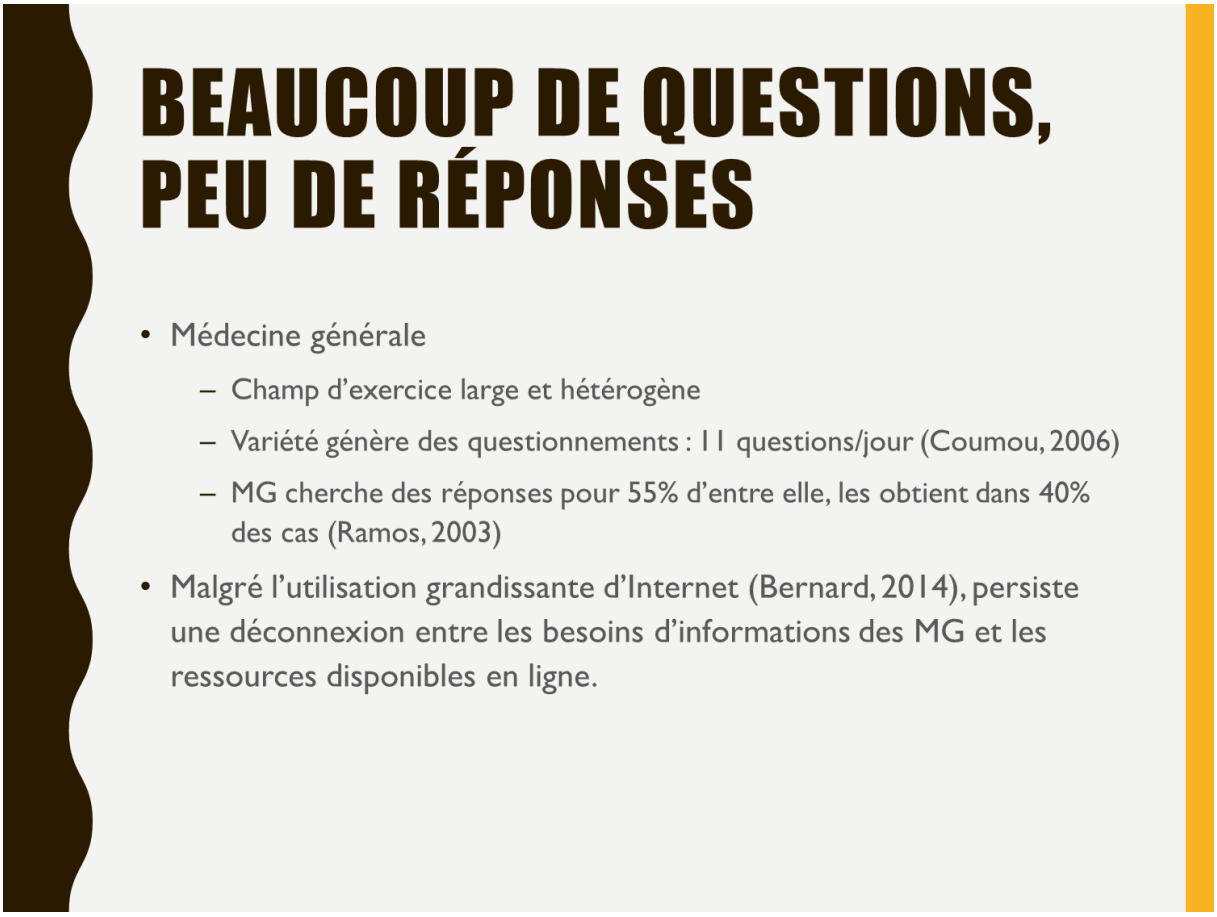
Enfin ce qu'on attend, ce qui a été exprimé ici, c'est qu'on veut s'ouvrir à de nouveaux partenaires, à de nouveaux éditeurs pour recevoir des données. C'est vrai qu'il faut des éditeurs capables de nous fournir des données structurées. Nous on peut s'adapter à pas mal de formats différents. On a l'exemple de la revue du CNGE, Exercer qui est une revue indépendante, qui a ses propres moyens et qui a été capable de nous fournir ses données. Fournir des données structurées, c'est une difficulté à dépasser mais, même pour une revue indépendante, c'est tout à fait faisable.

Je vous remercie et je vais passer la parole à Matthieu Schuers qui est médecin généraliste qui exerce en maison médicale à Neufchâtel-en-Bray à côté de Rouen, qui est aussi enseignant à l'Université de Rouen et qui a travaillé sur les besoins en Informations des médecins.

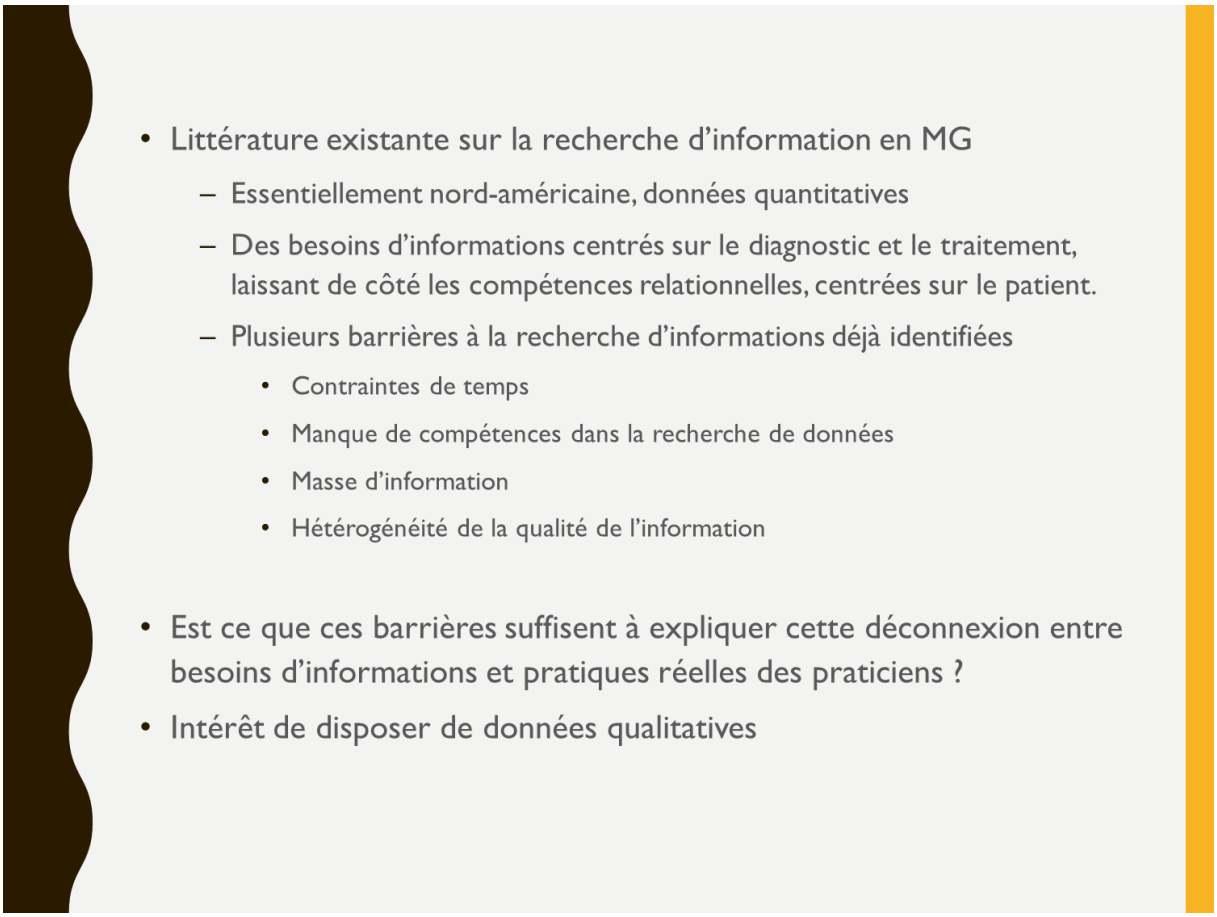
Intervention du Dr Matthieu Schuers

Bonjour à tous. Effectivement c'est quelque chose qui vient en amont de ces réflexions-là. C'est essayer de comprendre comment font les médecins généralistes et les internes de médecine

générale - c'est le public auquel on s'est concentré – quand ils vont chercher de l'information sur Internet.



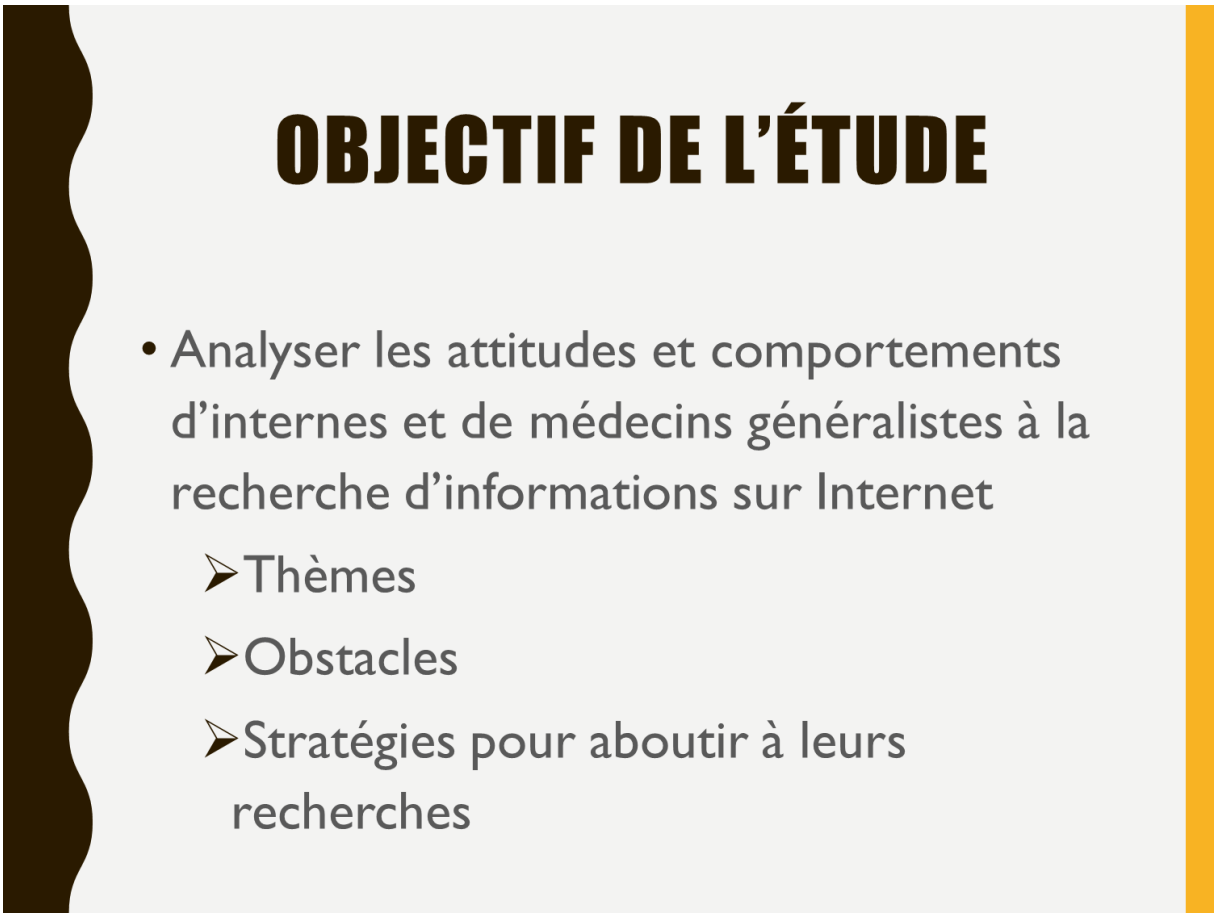
La médecine générale c'est un champ d'exercice qui est large, qui est très varié et qui implique un investissement important. Dans la littérature on considère qu'un médecin, au cours d'une journée de consultation, va se poser un peu plus d'une dizaine de questions. Il va essayer de répondre à un peu plus de la moitié d'entre elles et trouver une réponse dans à peu près 40% des cas. Forcément tous les médecins généralistes utilisent de plus en plus Internet et, malgré cette utilisation grandissante, il y a toujours des difficultés à trouver des réponses. Il y a une déconnexion entre leurs besoins en informations et les ressources qui sont disponibles en ligne.



Il y a déjà beaucoup de littérature sur le sujet et notamment sur la façon dont les généralistes vont chercher l'information sur Internet. Essentiellement de la littérature qui vient des États-Unis et c'est beaucoup de données quantitatives, par questionnaires. Ce qu'on trouve c'est que les besoins d'informations, les besoins qui se posent sont centrés essentiellement autour du diagnostic et du traitement. Ça évacue beaucoup toutes les questions autour des problématiques autour du patient, autour de la communication, autour de la relation médecin-malade. Tout ce qui est centré autour du patient ressort assez peu dans ces études-là.

Beaucoup d'études ont cherché aussi à identifier les barrières à la recherche d'information médicale sur Internet qui pouvaient exister. Comme toujours la contrainte de temps ressort en premier et qui ne signifie pas toujours grand-chose a priori. Il y a le manque de compétence dans la recherche de données, dans la construction des requêtes. Il y a la masse d'information qui est difficile à utiliser et l'hétérogénéité dans la qualité de l'information.

La question qu'on s'est posée est : finalement, est-ce que ces barrières qu'on identifie dans la littérature arrivent à expliquer complètement les difficultés que peuvent rencontrer les professionnels à la recherche d'information ? On s'est dit qu'il serait intéressant d'aller plus loin et c'est pour cela qu'une approche qualitative nous semblait intéressante et c'est donc ce qu'on a essayé de faire.

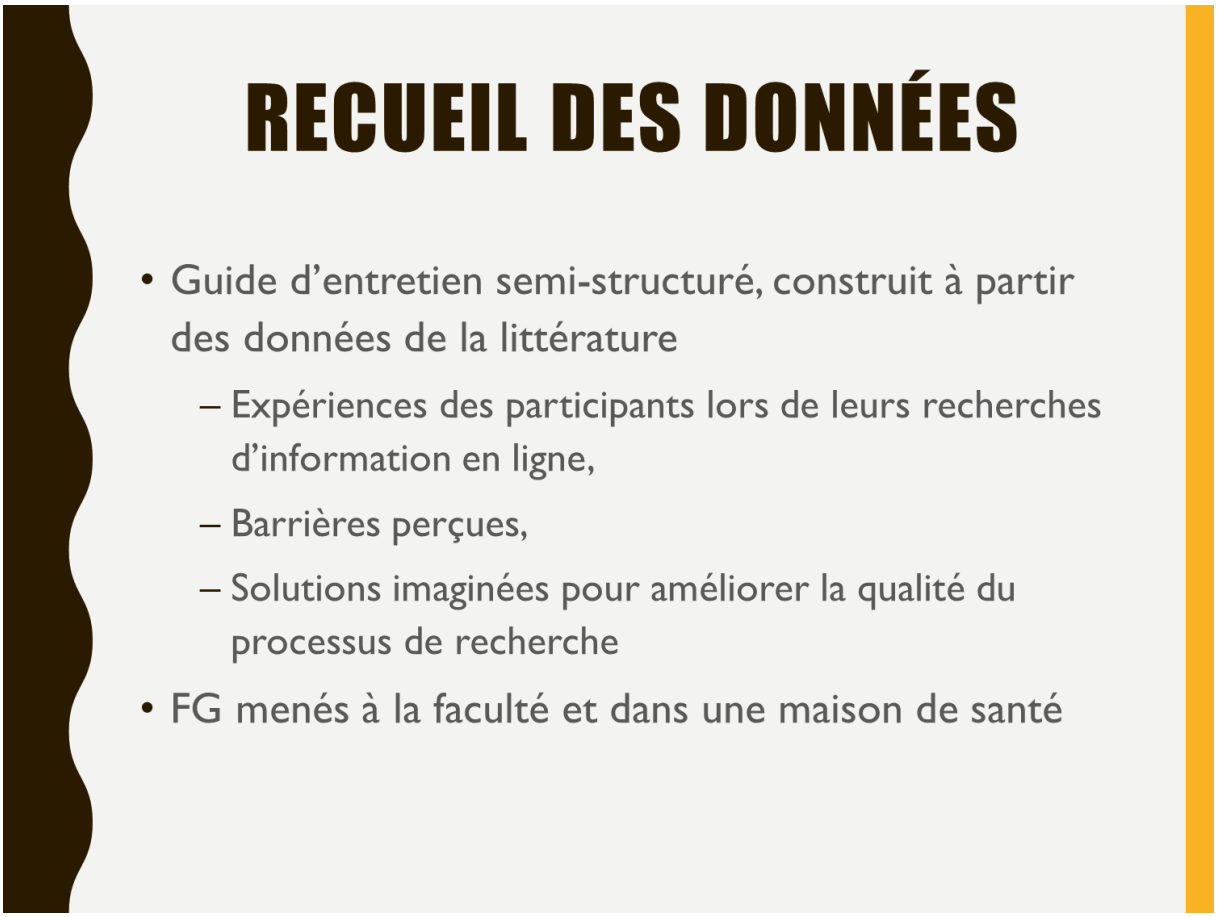


L'objectif de notre étude était :

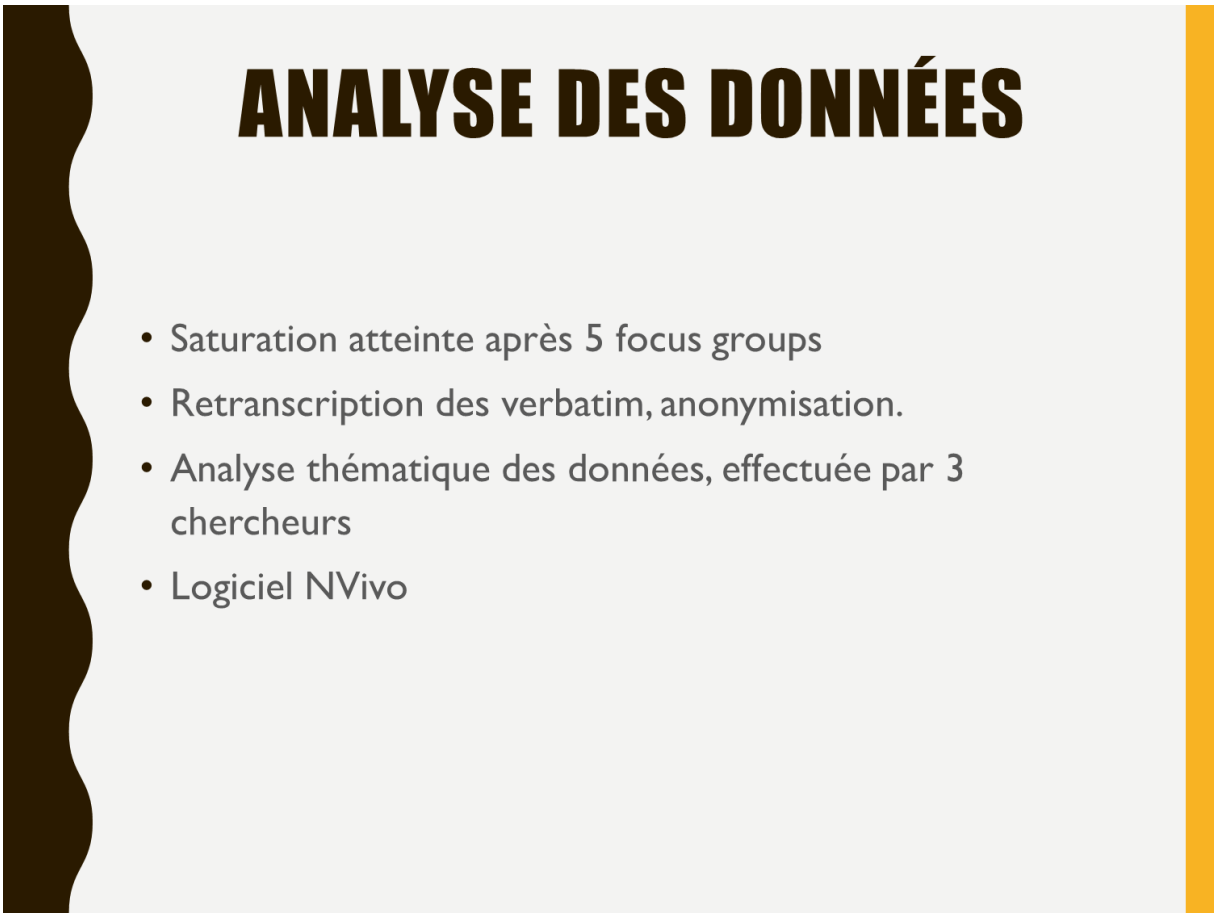
- d'analyser les attitudes et les comportements des internes de médecine générale et les médecins généralistes lorsqu'ils étaient à la recherche d'information sur Internet
- essayer d'identifier leurs thèmes de recherche
- les obstacles rencontrés dans leurs recherches

Et les stratégies qu'ils élaboraient pour contourner ces obstacles

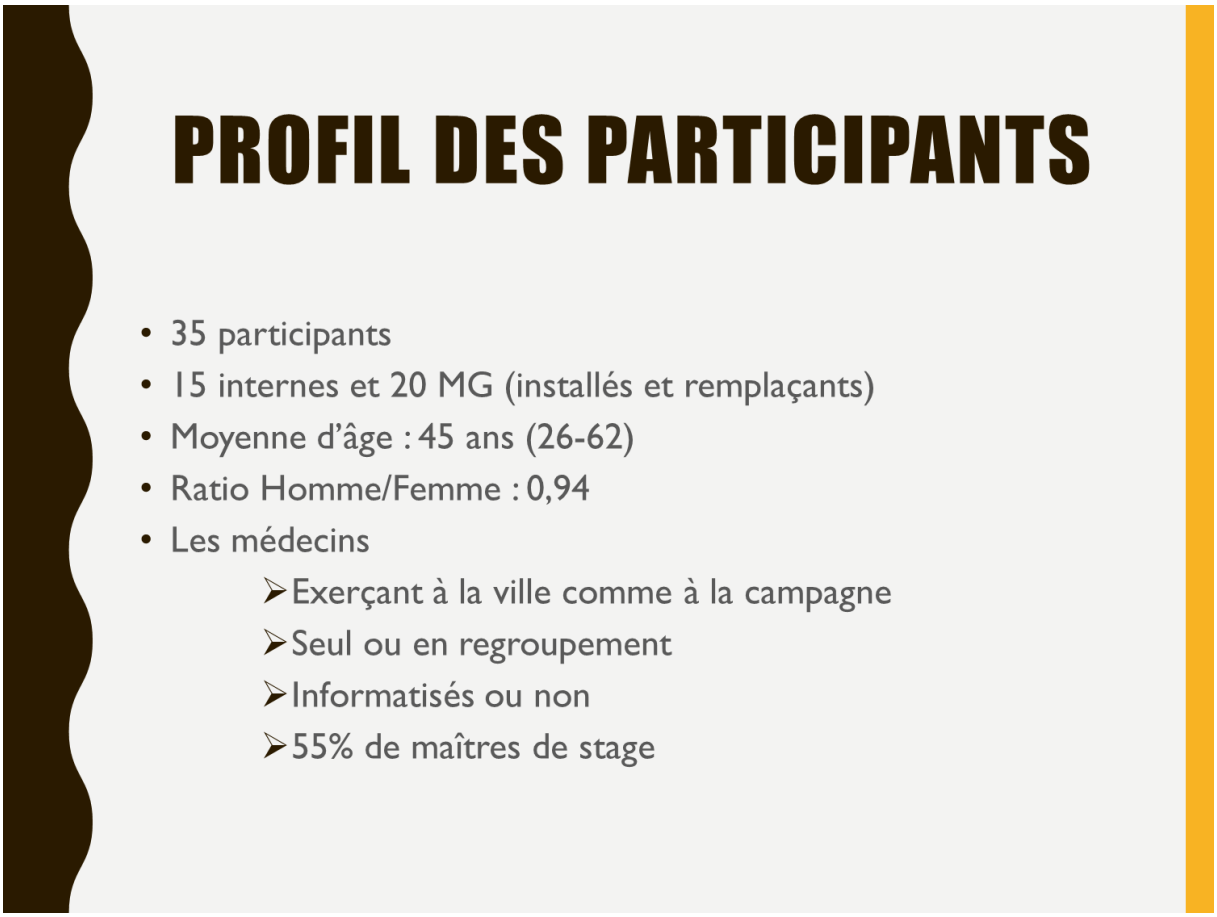
Je passe rapidement sur la méthode. C'est une étude qualitative. On a fait des Focus group, c'est-à-dire des petits groupes de participants qu'on a interrogés ensemble. C'est une étude qu'on a mené entre la fin 2013 et le début 2014 auprès d'internes de médecine générale et de médecins généralistes avec un échantillonnage qu'on appelle raisonné. C'est différent d'un échantillonnage aléatoire, l'objectif n'est pas d'être représentatif, l'objectif est d'avoir toutes les représentations, tous les sentiments qui soient représentés. On a effectué l'étude en Normandie, autour de Rouen et en région parisienne et on a obtenu les accords éthiques et de consentement nécessaires.



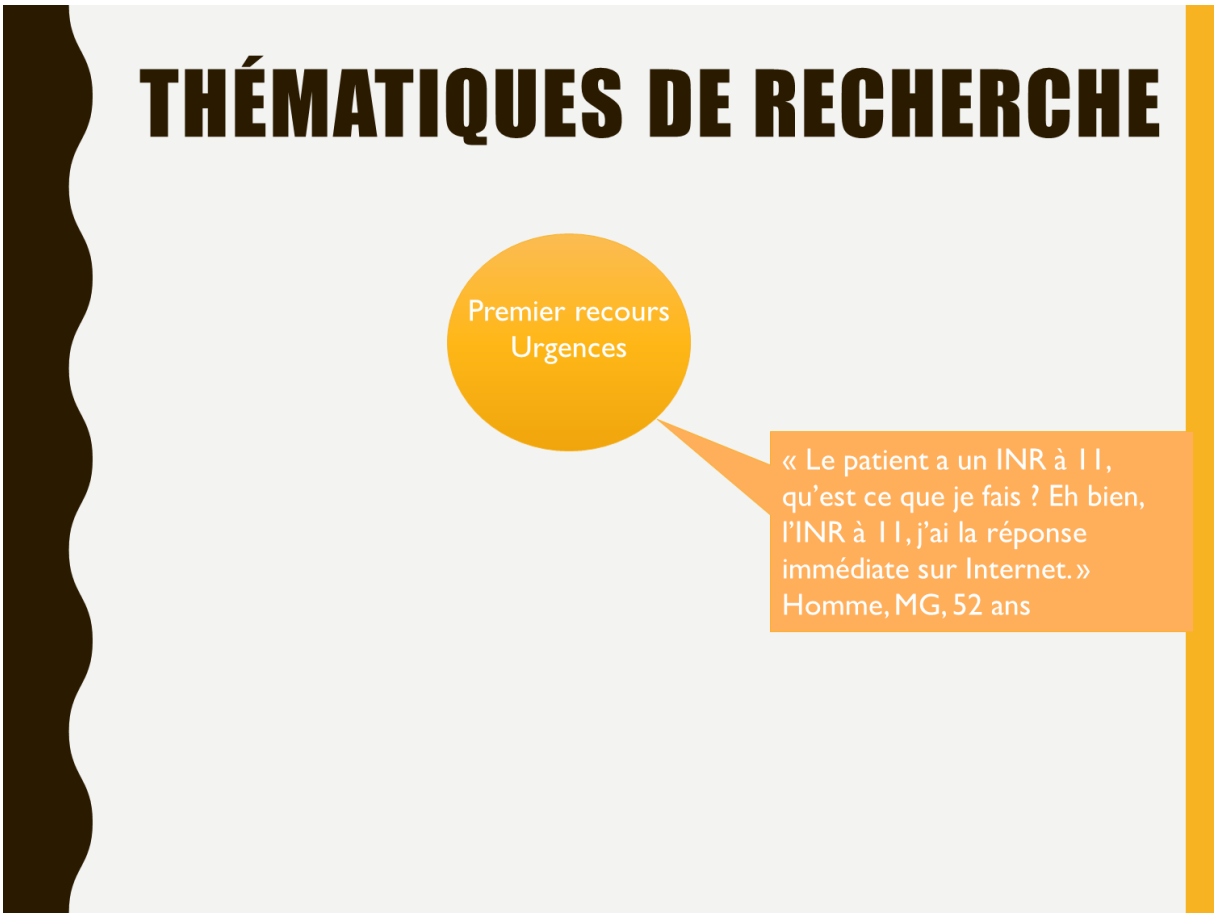
Pour animer ces focus groups on a construit un guide d'entretien semi-structuré à partir de la littérature. L'objectif était de recueillir l'expérience des participants lors de leurs recherches, les obstacles qu'ils rencontraient et les solutions qu'ils imaginaient pour améliorer la qualité de leurs requêtes, de leur processus de recherche et la qualité des résultats. La majorité des focus groups étaient menés au sein de facultés, soit à Rouen, soit à Paris VII et 1 dans une maison de santé, à Neufchâtel.



Après cinq focus groups, on a plus eu de nouvelles données donc on s'est arrêté de les recueillir. On a retranscris toutes les informations, on a tout anonymisé. Tout a été analysé par trois chercheurs avec un logiciel qui s'appelle NVivo.

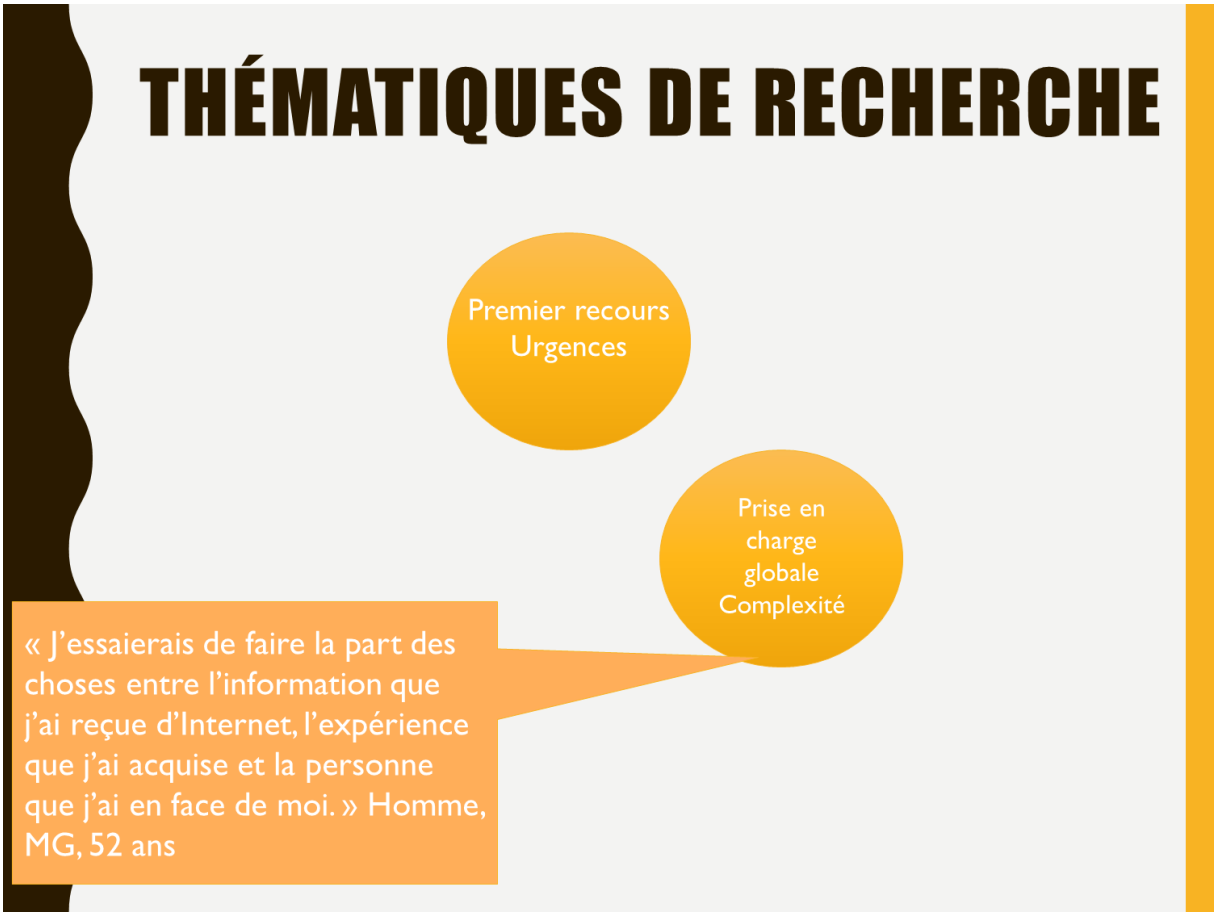


On avait 35 participants, (15 internes, 20 médecins), des remplaçants et des installés ; 45 ans de moyenne d'âge ; avec un ratio hommes-femmes autour de 1 ; des médecins exerçant aussi bien en ville qu'à la campagne, aussi bien seul qu'en groupe. La plupart étaient informatisés mais pas tous et un peu plus de la moitié étaient maîtres de stage.

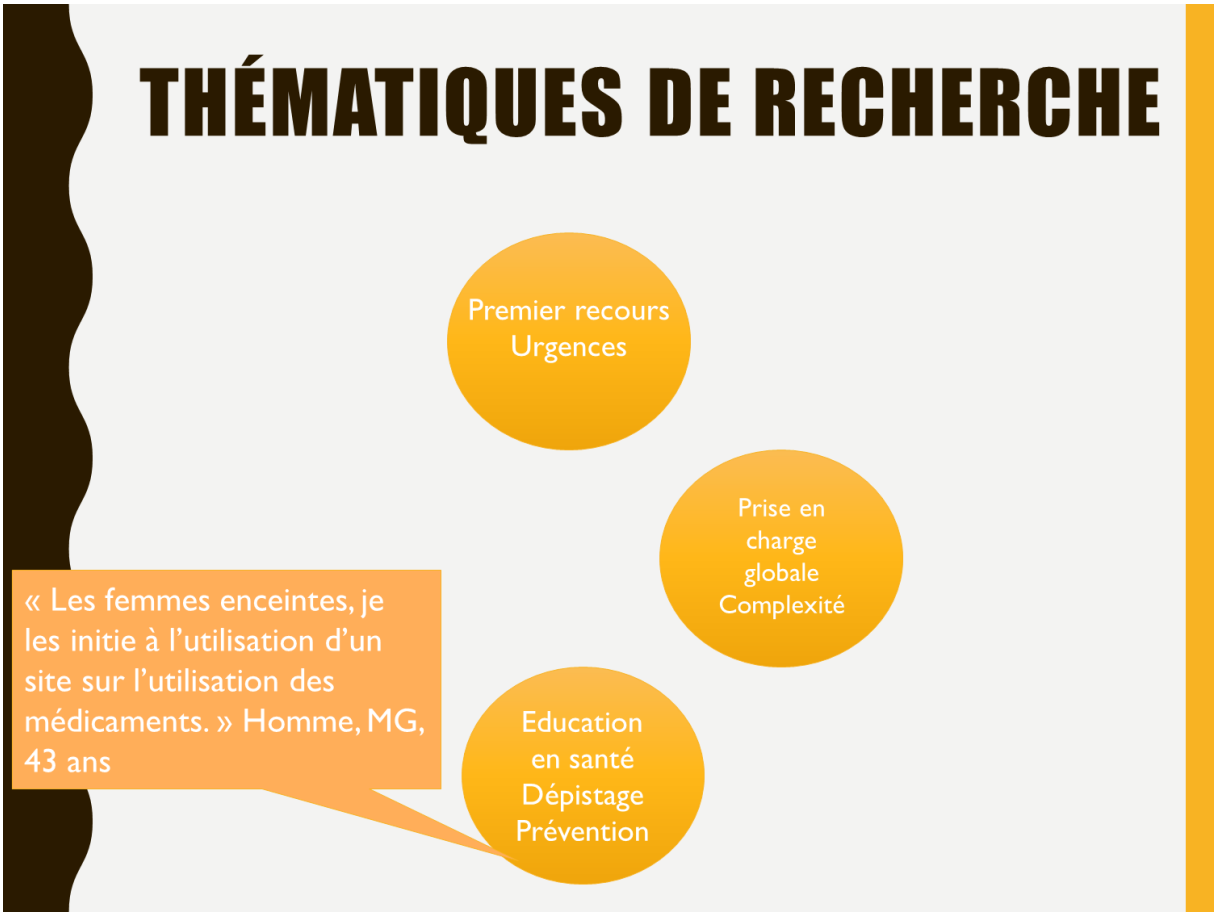


La première question qu'on leur a posé c'est leur thématique de recherche. On s'est rendu compte qu'elles étaient assez vastes. Il y avait des questions concernant le diagnostic et le traitement des maladies. Je vous ai mis pour chaque idée, un verbatim, une phrase d'un participant pour illustrer.

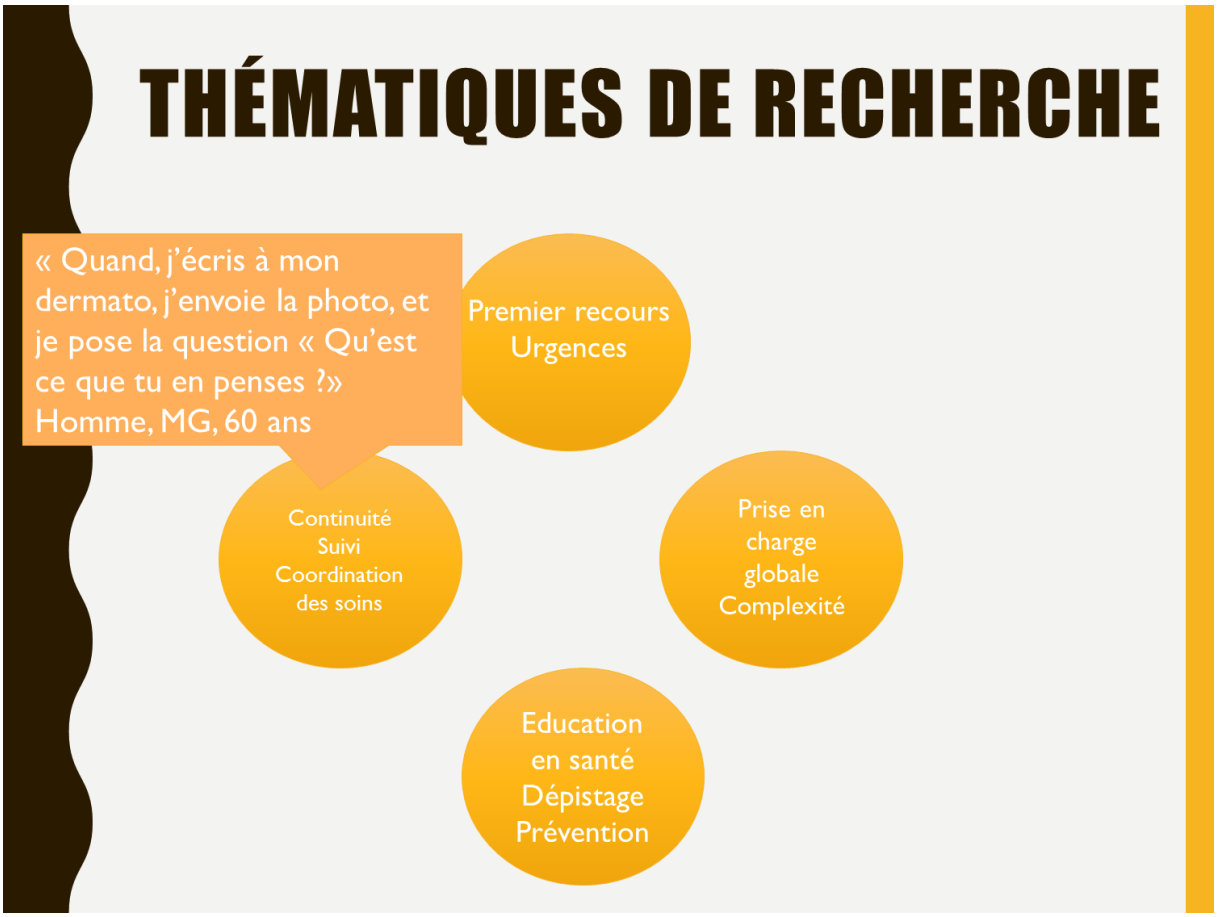
Tout ce qui va concerner la gestion du premier recours, des urgences, le diagnostic, le traitement.



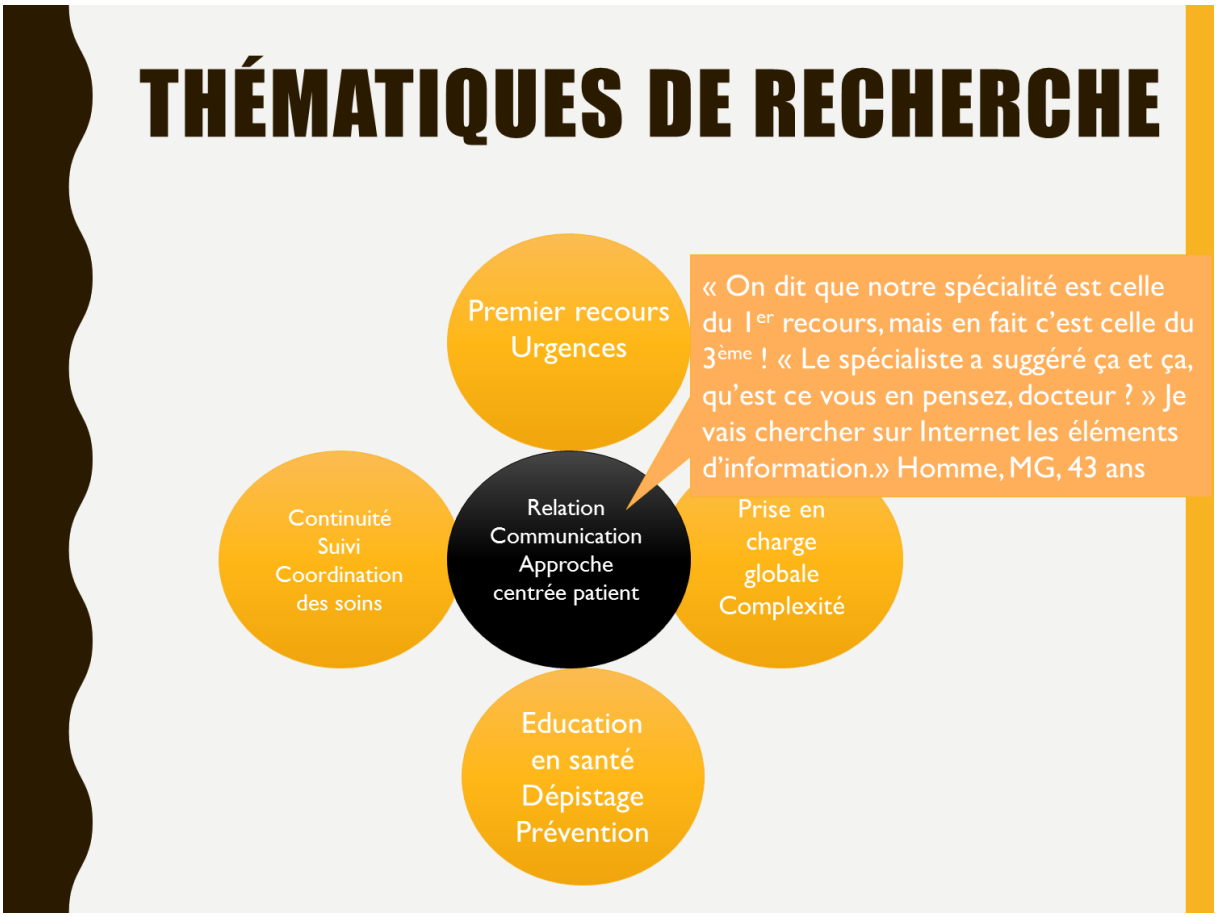
Il y avait aussi d'autres types de questionnement qui prenaient en compte tout ce qui est prise en charge globale, la complexité de la consultation.



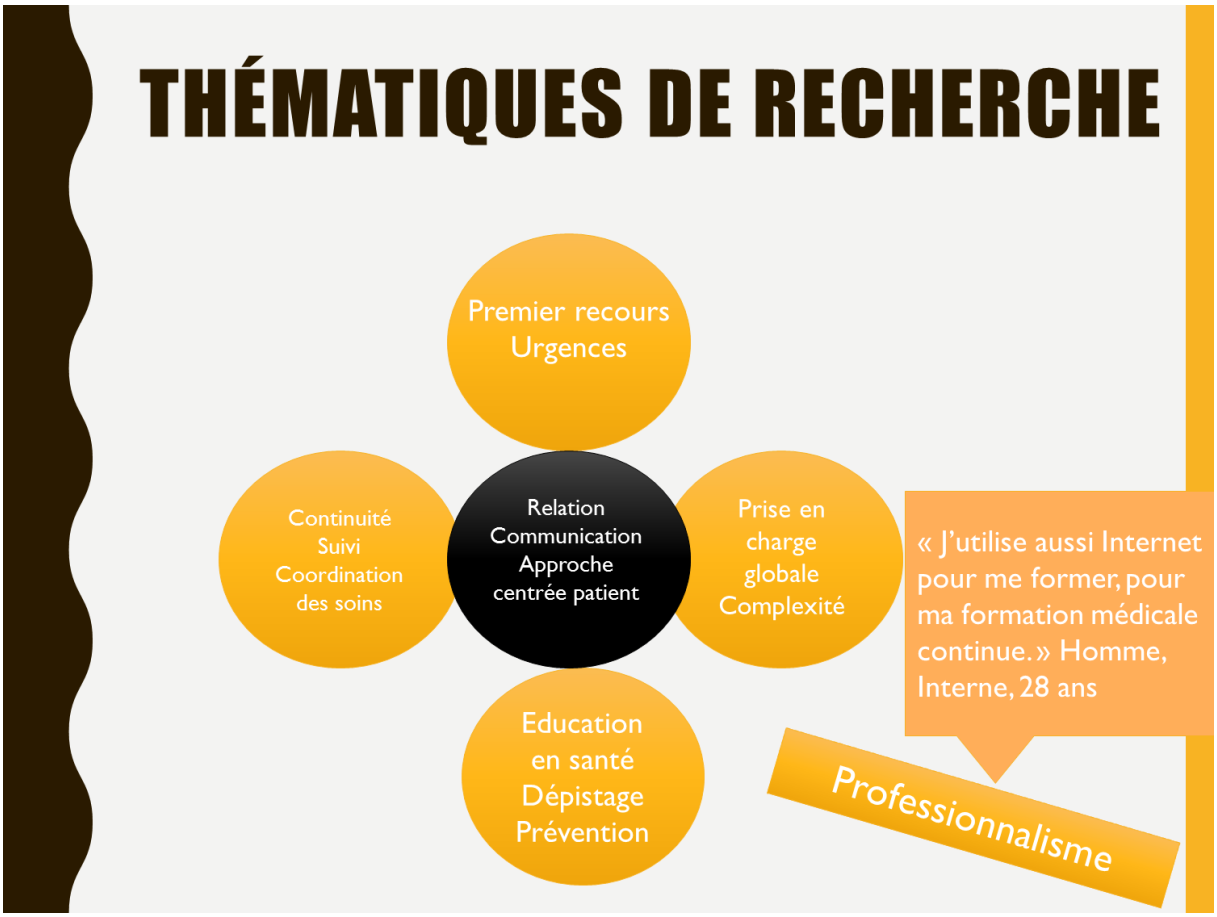
Il y avait aussi des problématiques de prévention, d'éducation à la santé, de dépistage.



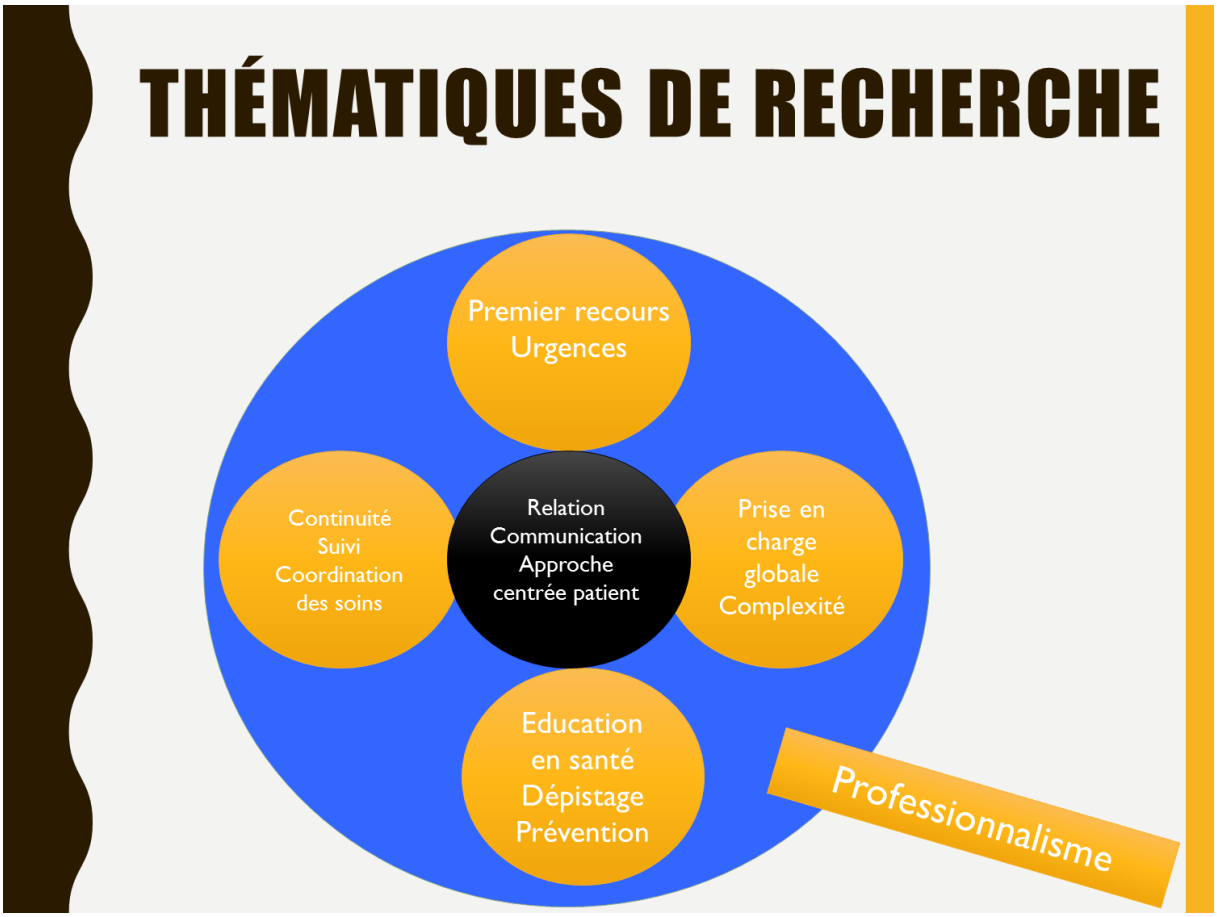
Des problématiques – là on dépasse un peu la recherche d'information, on est plus dans le partage d'information mais c'était intéressant donc on l'a mis quand même – qui vise à faciliter le suivi des patients, la coordination des soins.



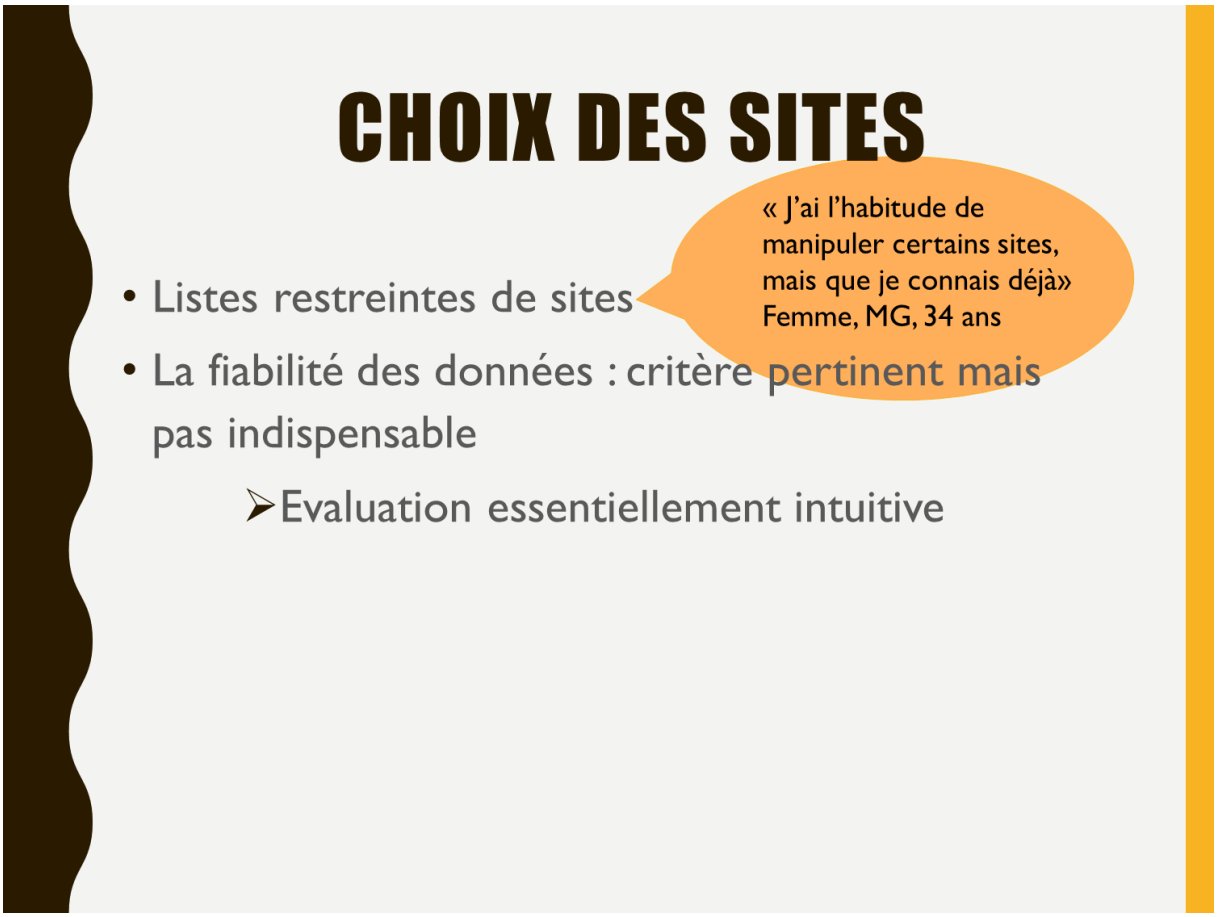
Et puis beaucoup de questionnements autour du patient et d'une approche centrée patient justement. « On dit que notre spécialité est celle du 1er recours, mais en fait c'est celle du 3ème ! [avec un patient qui vient et qui dit :] « Le spécialiste a suggéré ça et ça, qu'est ce vous en pensez, docteur ? » Je vais chercher sur Internet les éléments d'information.» [pour en discuter avec le patient.



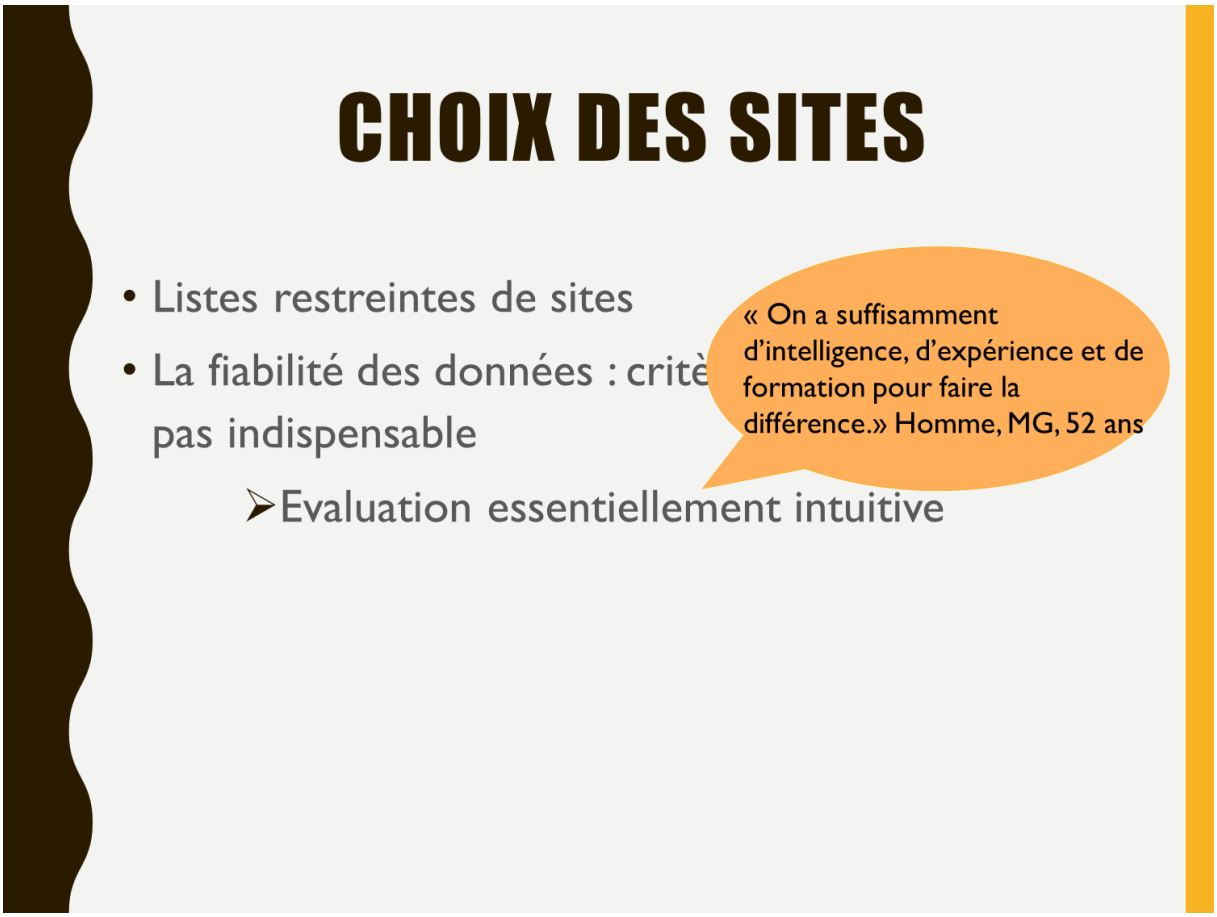
Beaucoup de médecins expliquaient qu'ils utilisent Internet dans la formation médicale continue.



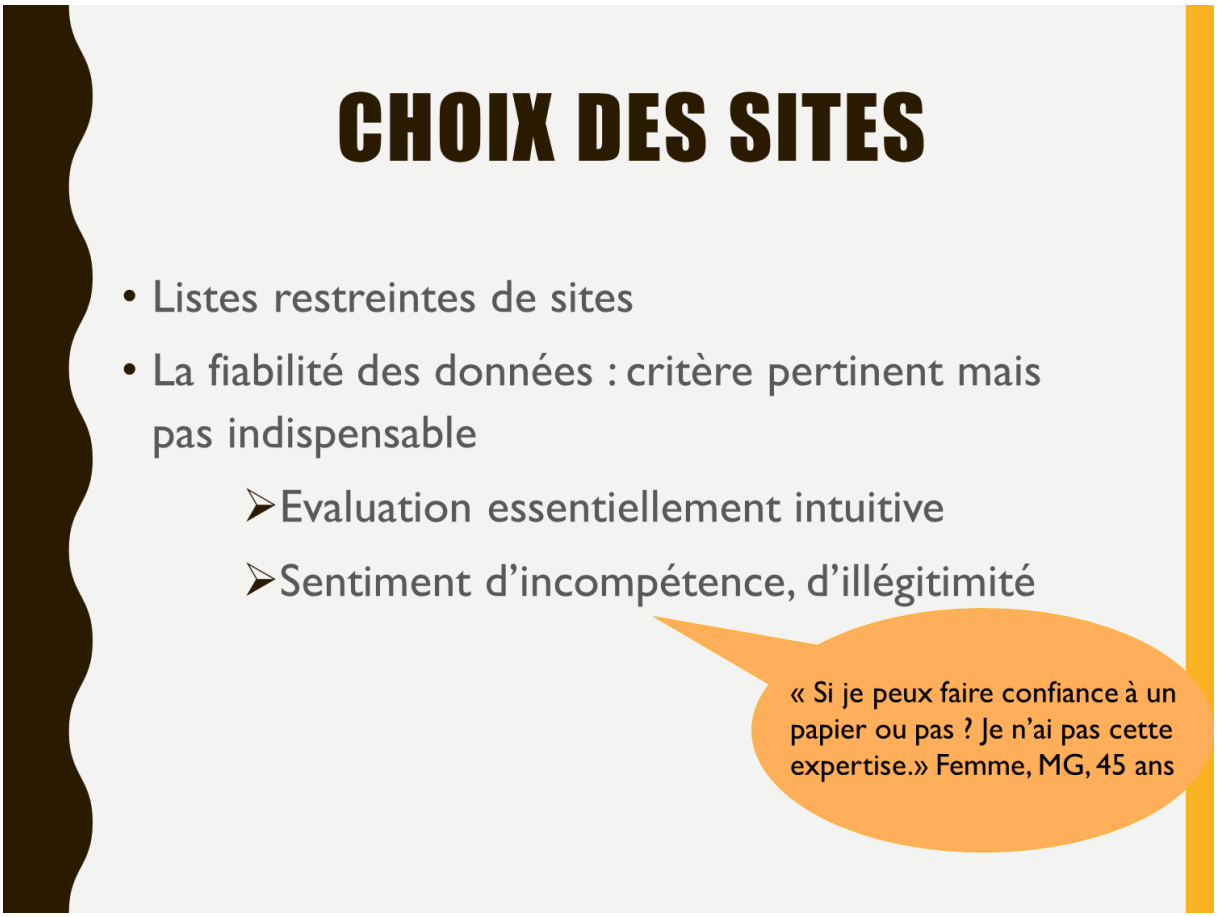
Au fur et à mesure, on a un ensemble des compétences de la médecine générale qui se dessinent et qui dépasse le diagnostic et le traitement des maladies.



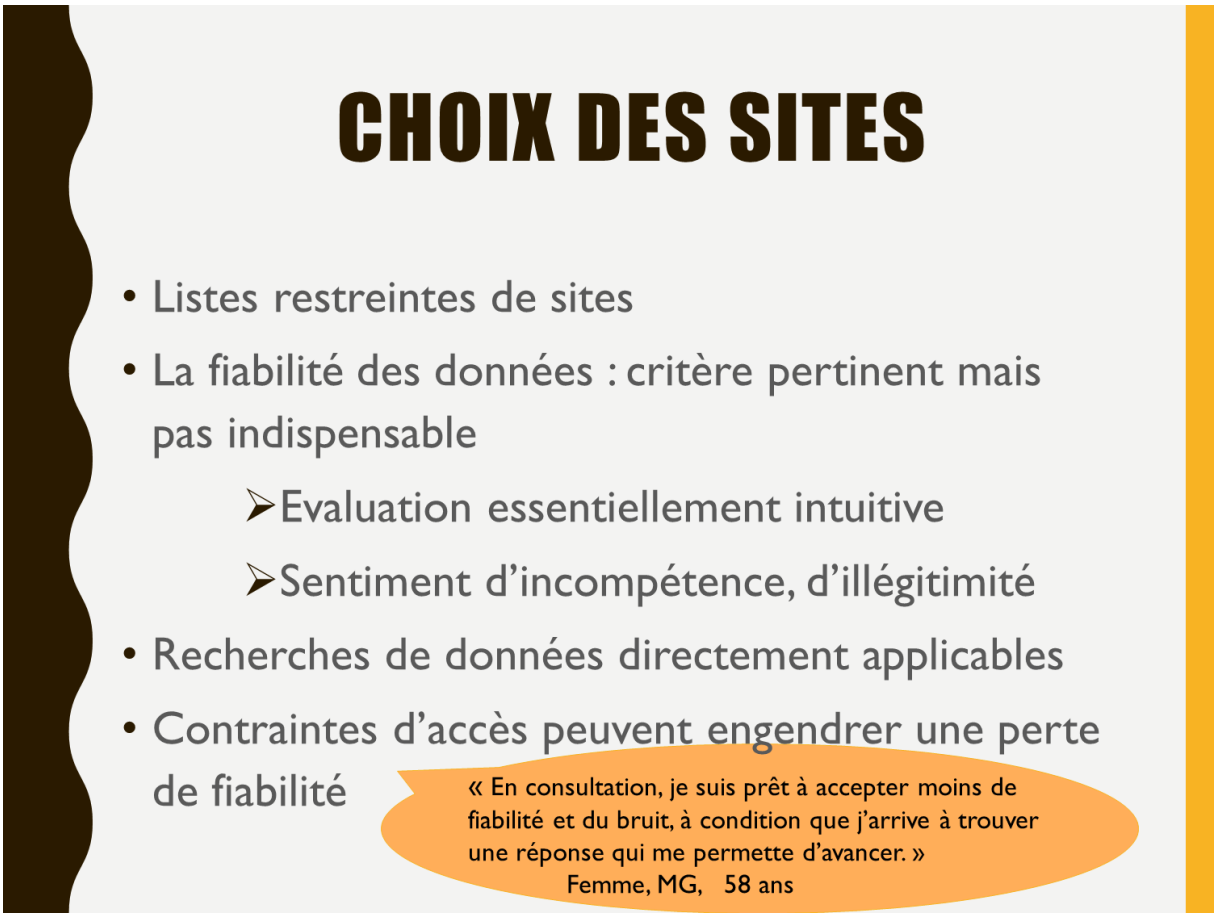
On a essayé de les interroger sur la manière dont ils recherchaient de l'information et sur le choix des sites utilisés. Globalement on avait pour chaque utilisateur une liste restreinte de sites, des sites peu nombreux, qu'ils connaissent bien et qu'ils utilisent souvent.



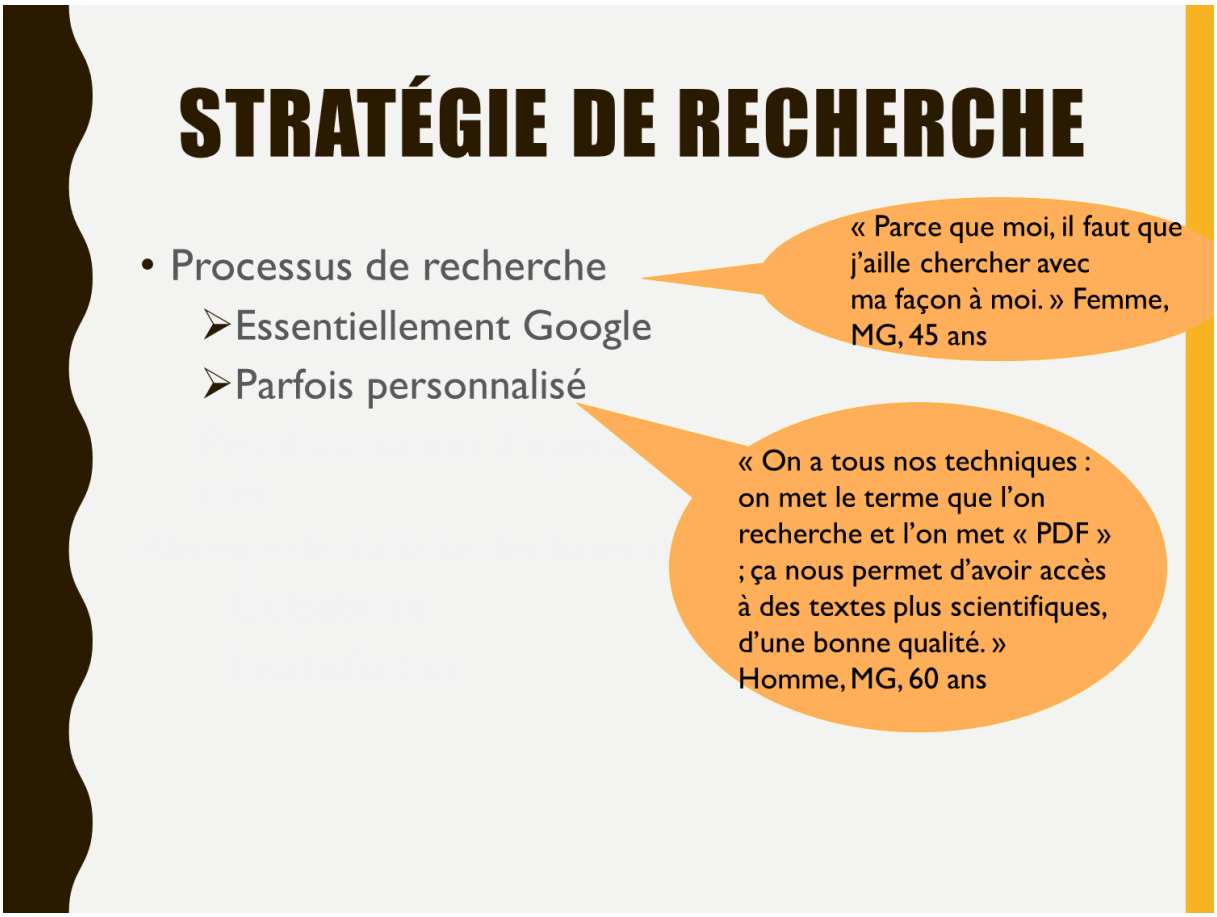
Pour choisir les sites, ils ont parlé de fiabilité, c'était un critère pertinent mais qui n'était pas fondamental. Pour évaluer cette fiabilité, l'évaluation est essentiellement intuitive.



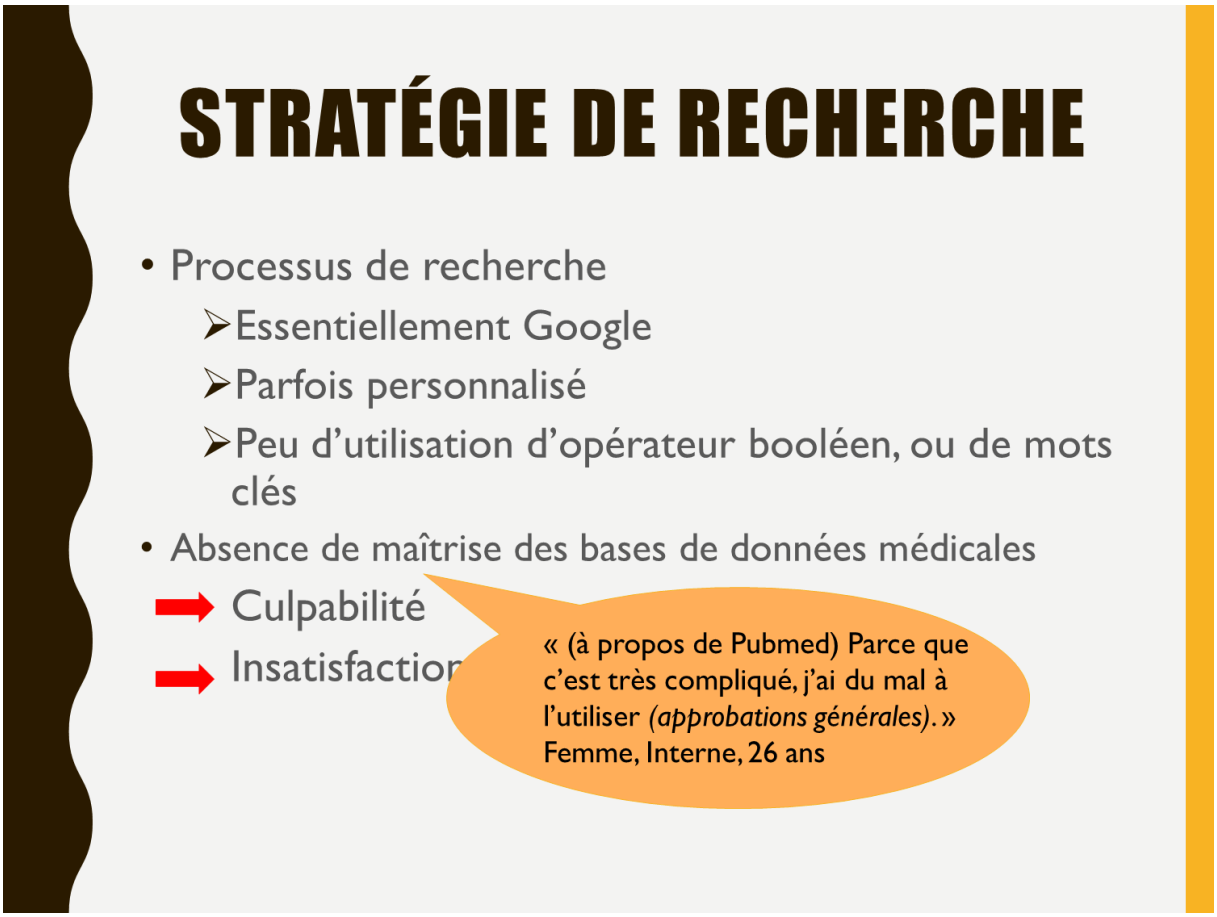
En allant un peu plus loin, on se rend compte qu'il y avait un certains nombres de critères qu'ils utilisent pour évaluer la qualité des sites : Ils tiennent compte de la personne qui les a informés de l'existence du site, de l'institution qui est derrière le site, des modalités de financement. Ils tiennent compte aussi de l'apparence du site, du caractère scientifique ou non du vocabulaire. Malgré cela, la plupart du temps, l'évaluation est assez intuitive. La majorité des participants se sentaient incompétents voire illégitimes pour évaluer cette fiabilité.



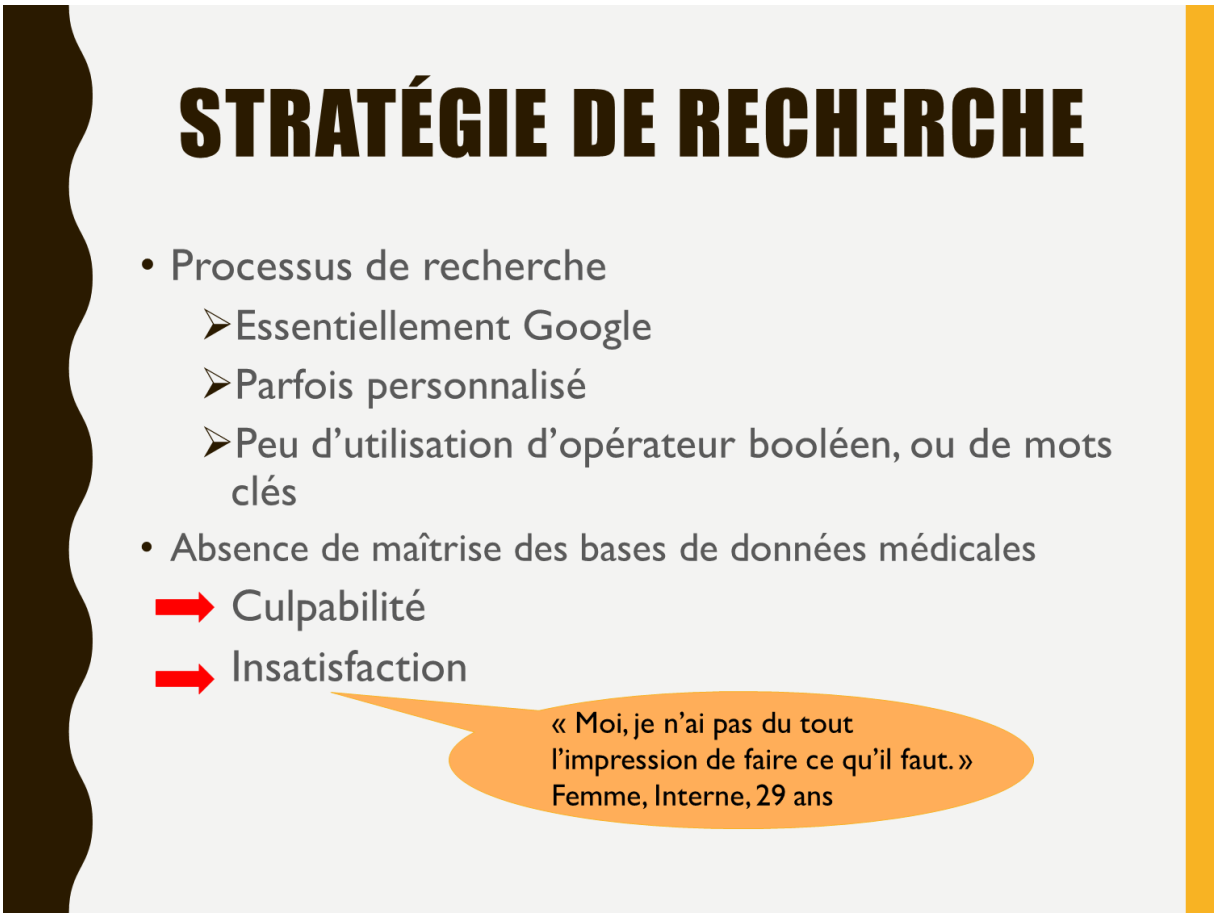
Ils cherchaient des données directement applicables à leur exercice et ils étaient conscients que les contraintes d'accès, notamment le temps à leur disposition, allaient engendrer une perte de fiabilité et ils acceptaient cela.



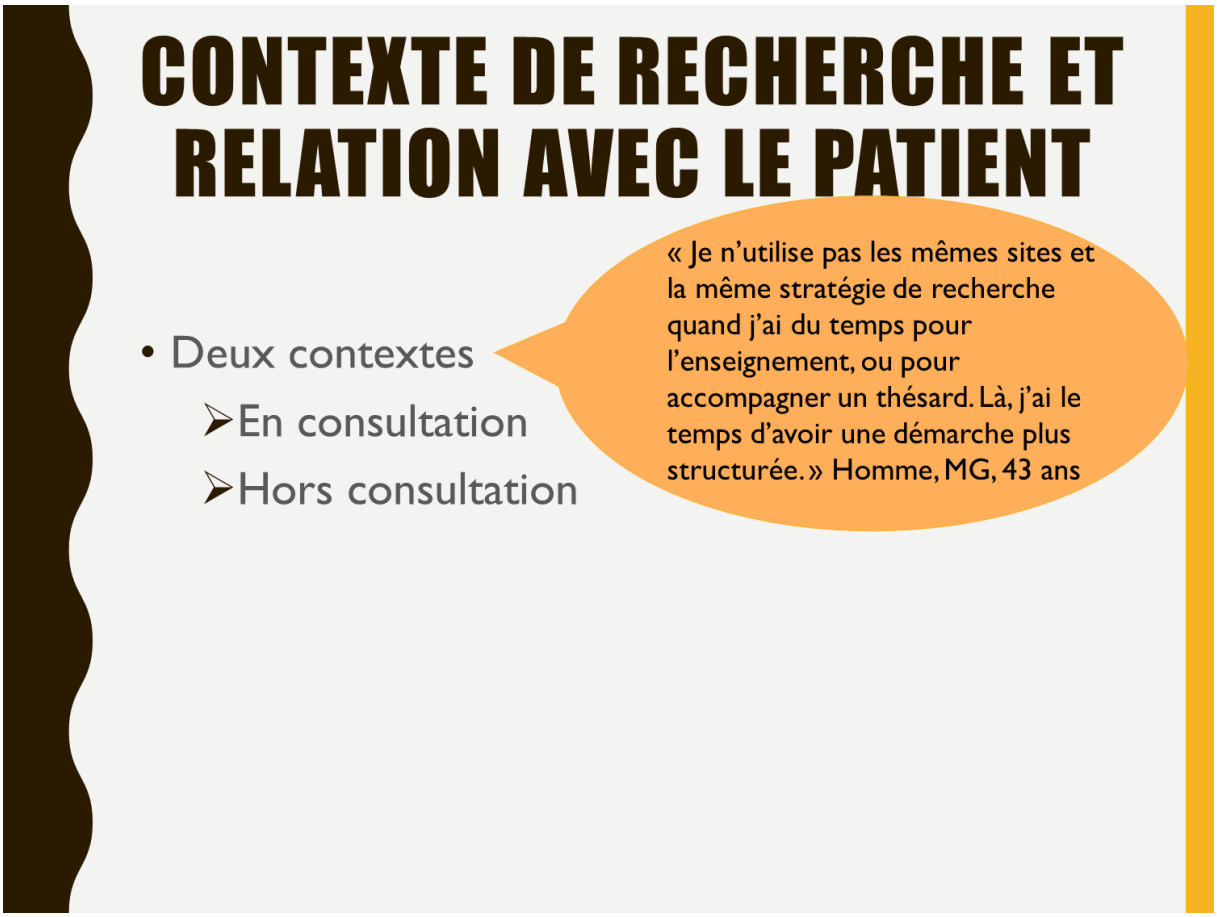
En ce qui concerne leurs stratégies de recherche, de la même manière que pour leur liste de sites, ils avaient des processus de recherche bien à eux qu'ils avaient construits au fur et à mesure. « Parce que moi, il faut que j'aille chercher avec ma façon à moi. ». Ils utilisaient essentiellement Google mais avec des manières personnalisées, par exemple mettre « PDF » dans la recherche pour obtenir des documents différents, d'autres utilisaient « univ » pour tomber sur des sites d'ECN ou de formation en deuxième cycle. Ils utilisaient très peu d'opérateurs booléens et très peu de mots-clés.



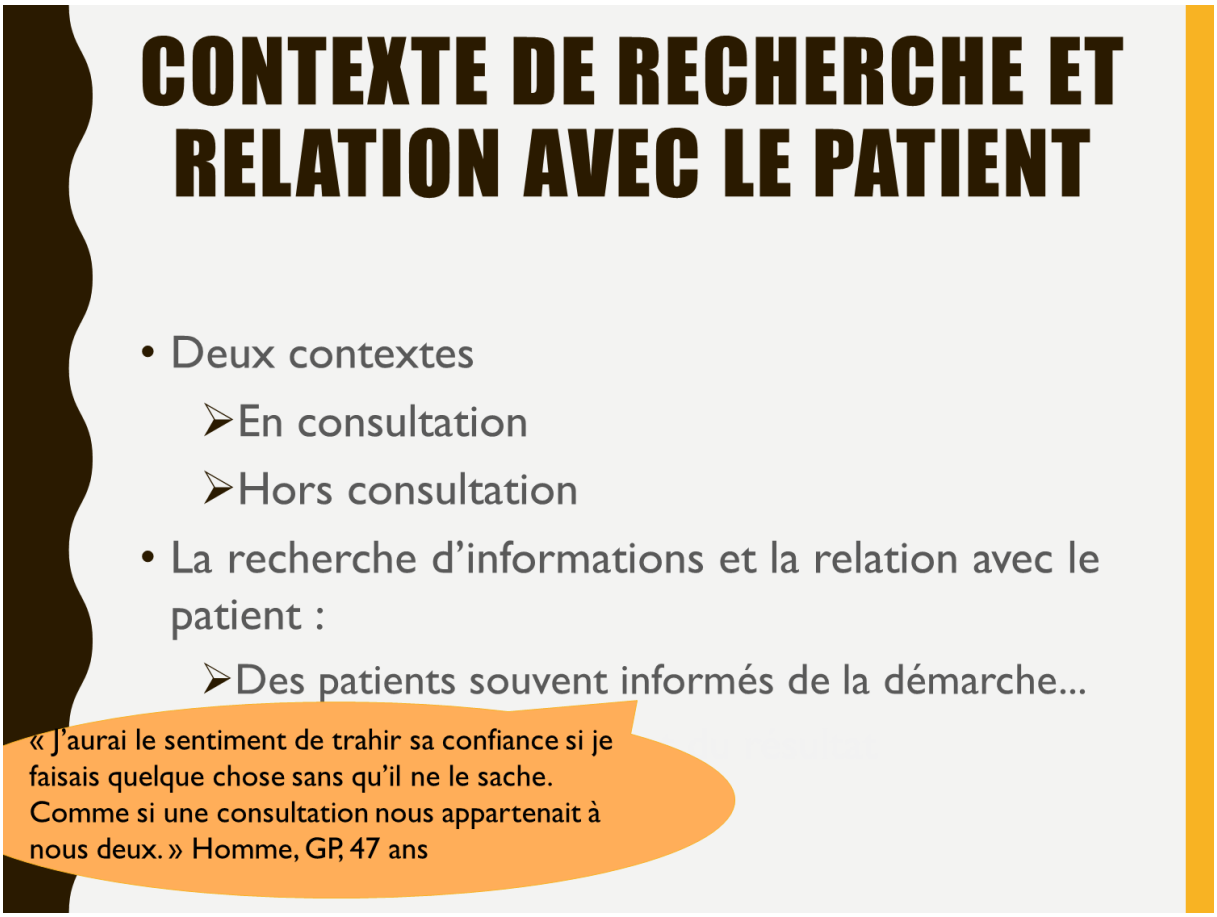
Globalement ils maitrisent très peu les bases de données médicales et cela systématiquement chez tous les participants, les internes ou les médecins généralistes. « Parce que c'est très compliqué, j'ai du mal à l'utiliser ». Tout le monde était d'accord sur ce point dans son groupe.



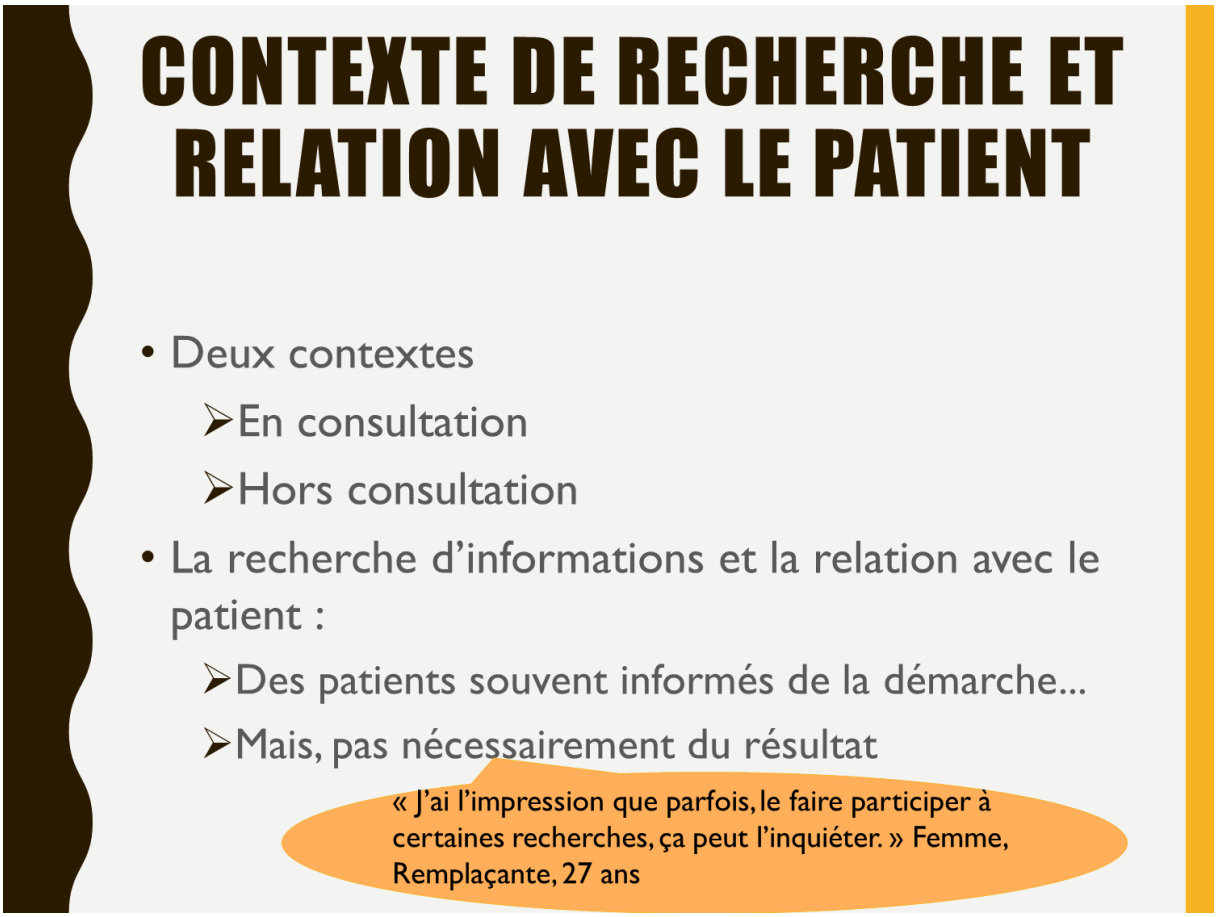
Cela engendrait une notion de culpabilité et d'insatisfaction. Ils sentaient bien qu'ils ne faisaient pas ce qu'il fallait. Ils étaient conscients des conséquences.



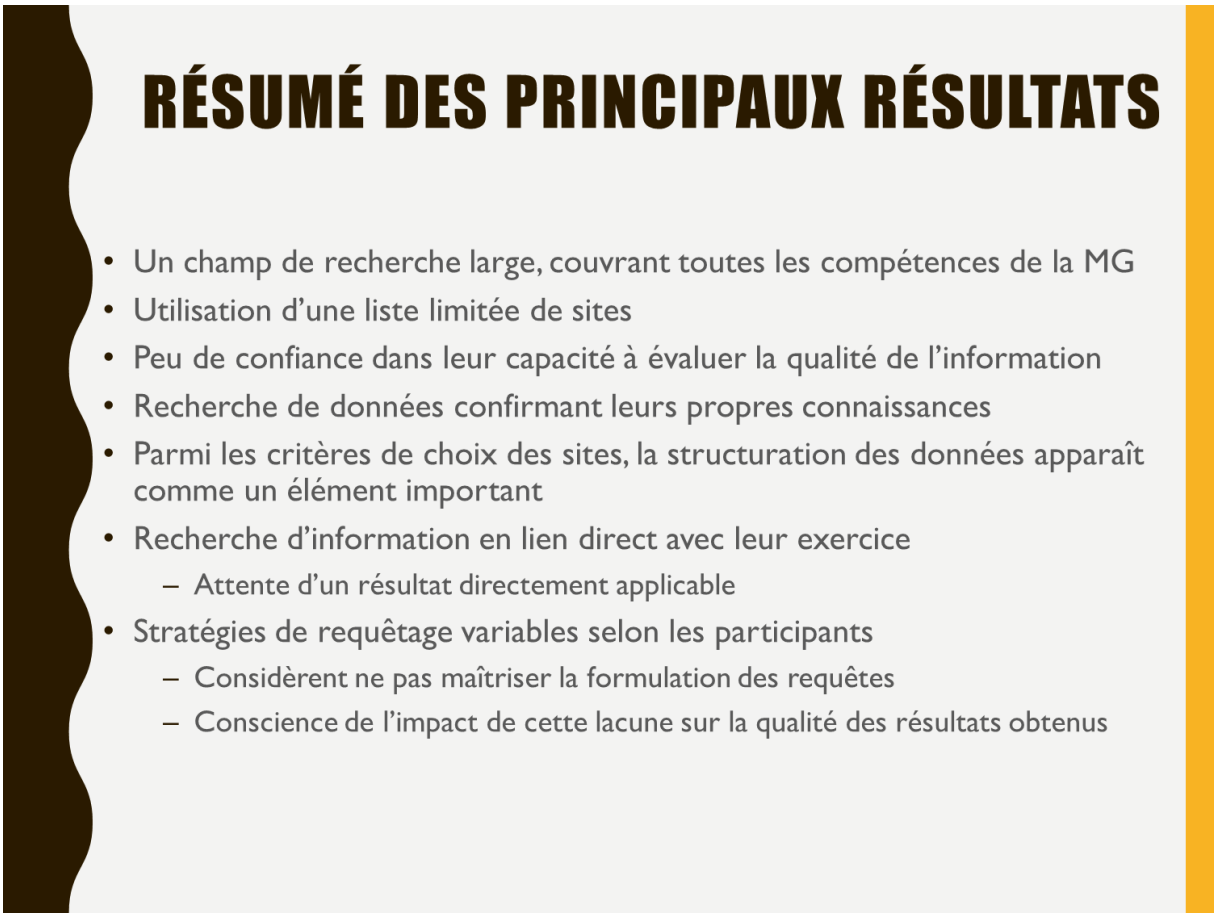
On les a aussi interrogé sur leur sur la manière dont ils cherchaient, à quel moment ils cherchaient. On s'est rendu donc qu'il y avait bien deux manières de rechercher, deux contextes différents ; en consultation et hors consultation. Le fait d'être en consultation ou non influençaient totalement la manière de rechercher.



Au-delà de cela il y avait aussi la question de la place du patient. Comment impliquer le patient dans la recherche d'information. Les patients étaient souvent informés de la démarche.



Après se posait la question de trouver une place équilibrée car les patients n'étaient pas toujours informés du résultat. C'était parfois source d'anxiété. « J'ai l'impression que parfois, le faire participer à certaines recherches peut inquiéter ». La place du patient est encore difficile à trouver. La manière de l'intégrer dans la recherche génère une anxiété supplémentaire.



Globalement on s'est rendu compte que la recherche d'informations des médecins généralistes et des internes ne concernait pas que le diagnostic et le traitement mais était beaucoup plus large et prenait en compte toutes les compétences des MG et notamment l'approche centrée patient qui était assez développée. Des problématiques préventives etc. qui étaient assez peu retrouvées dans les études qualitatives précédentes.

Ils utilisent une liste assez limitée de sites qu'ils connaissent bien et qu'ils utilisent quotidiennement mais ils ont peu confiance dans leur capacité à évaluer la qualité de l'information aussi bien la fiabilité d'un site que d'une donnée en elle-même. Ils ne savent pas le faire et ils en sont conscients. Parmi les critères de choix des sites, la structuration des données apparaissait comme un élément important. Ils cherchaient une information en lien direct avec leur exercice et attendaient un résultat directement applicable. Un site très souvent cité est le site Antibioclic développé par Paris 7 et qui permet sur une question précise sur une antibiothérapie chez un patient de fournir une réponse précise qui est formalisée structurée et référencée. C'était un site qui correspondait vraiment à leurs attentes en tout cas dans cette thématique-là.

Les stratégies de requêtage étaient vraiment variables suivant les participants qui considéraient ne pas maîtriser la formulation des requêtes et qui étaient conscients de l'impact de cette lacune sur la qualité des résultats.

Je n'ai pas fait de Discussion car j'ai pensé que ce serait intéressant d'en discuter ensemble. Merci.

Débat

[Question de Daniel Vasmant, ANR, sur la place de l'écran durant la consultation.]

>>Matthieu Schuers

Ça a été très variable. Cela a été souvent évoqué, à la fois la place de l'écran, dans quel sens est l'écran, la place du patient aussi, certains profitent du moment où le patient se déshabille ou se rhabille pour aller chercher l'information. On sentait que c'était très lié leur manière de maîtriser à la fois la problématique de santé du patient, à la fois leur lien avec le patient et à la fois leur technique de requêtage. Mais il y en a, moi en tout cas, qui ont l'écran tourné vers le patient. Beaucoup le faisaient mais c'était très variable.

[Catherine Tellaa, BIUS, question sur UpToDate]

>>Stéfan Darmoni

Très peu d'hôpitaux en France l'ont. Le marché UpToDate en France est inférieur à celui de la Belgique.

>>Matthieu Schuers

C'est cher, c'est en anglais. L'anglais finalement a été très peu cité mais je pense que c'était intuitif, l'anglais n'est pas un problème car je ne vais même pas chercher en anglais.

>>Matthieu Schuers

Pour un article on a ressortit une étude[3], assez récente, sur où en est le niveau d'anglais des étudiants en médecine versus ceux en business schools ou en école d'ingénieurs. Sur le TOEIC il y a une différence de 200 points. C'est énorme. Je parle d'aujourd'hui pas des vieux comme moi. Le niveau d'anglais de nos collègues est très faible, y compris dans les hôpitaux universitaires.

>> Gaétan Kerdelhué

Merci beaucoup Matthieu. Cela nous ramène à la réalité. J'espère avoir été assez enthousiaste que nous avons développé mais on mesure ici à quel point on est loin de pouvoir répondre à tous les besoins d'information des médecins, notamment les généralistes. Notre outil n'est pas là pour ça, il n'est pas adapté pour la consultation et je ne suis pas sûr que d'autres le soient vraiment. Même les outils de synthèse type UpToDate, Dynamed ou Clinical Key passent beaucoup par l'anglais et c'est un frein important. Cela met notre travail en perspective, cela nous donne l'objectif à atteindre en termes d'apport de connaissances. On espère que LiSSa va dans le bon sens.

[Question de la salle sur l'intégration de LiSSa dans le dossier médical]

>>Stéfan Darmoni

[3] Lebrun L, Fisbach M, Febvre J, Joly-Guillou M-L, Saint-André J-P, Richard I. Le niveau de compréhension de l'anglais des étudiants en médecine peut être amélioré. Résultats d'une stratégie d'évaluation systématique. La Presse Médicale. 2011 Oct;40(10):e471–6.

Ce n'est pas adapté. L'outil le plus adapté pour cela aujourd'hui serait VidalRecos qui est, en particulier pour la médecine générale, formalisé. Ce que la HAS ne fait toujours pas pour ces recommandations. Avec Antoine Vial, cela fait quinze qu'on dit que les recos de la HAS doivent être formalisées. Quelques-unes le sont mais, à ma connaissance, pas toutes. Si la HAS ne le fait pas, c'est une boîte privée qui le fait. Ce qui serait bien c'est que la HAS produise, au minimum, des arbres de décision sur les maladies chroniques. Pour moi c'est un service qui devrait être public.

[Question de la salle sur les utilisateurs rédigeant un mémoire.]

>>Matthieu Schuers

On a interrogé les internes qui cumulaient ces deux objectifs différents : à la fois répondre aux questions qu'ils se posent en stage et à la fois des recherches dans le cadre de leur thèse. On avait deux populations différentes dans l'étude en pensant avoir des réponses différentes mais finalement cela n'a pas été le cas. Les réponses se recoupaient de manière importante. Les difficultés rencontrées par les généralistes étaient globalement, a minima, celles rencontrées par les internes.

>> Catherine Tellaa

C'est le désespoir ! Ça veut dire qu'on a un gros travail. Il y a très peu d'étudiants qui savent que dans leur BU, ils peuvent avoir de l'aide et qu'il y a des formations à PubMed dans toutes les facs de médecine.

>>Stéfan Darmoni

Là où ça va impacter, c'est pour les étudiants. Pour les étudiants qui font leur thèse cela sert déjà. LiSSa rend ce service-là. Est-ce que cela pourra rendre plus de services ? J'en suis moins sûr. Ce n'est fait que pour répondre aux situations où on a besoin de répondre en utilisant la littérature scientifique. Pour reparler d'Antibioclic, si on cherche quelle est la meilleure antibiothérapie pour tel truc. Là ce n'est pas notre outil qui peut répondre. Ce n'est pas un système d'aide à la décision.

>>Matthieu Schuers

Il y a aussi des choses qu'on ne maîtrise pas et je pense que la manière dont sont organisées actuellement les études médicales ne favorise pas ça notamment les premiers et deuxièmes cycles et les épreuves classantes nationales où l'objectif est d'ingérer un certain nombre de connaissances. On n'est pas du tout dans une logique qui favorise le questionnement et la recherche d'informations. C'est seulement en troisième cycle qu'ils apprennent ces compétences, plus ou moins difficilement. Jusqu'en sixième année ils n'ont pas de recherches à faire car tout leur est déjà fourni

Ce qu'on a vu c'est que pendant la consultation les médecins cherchaient uniquement à confirmer leur hypothèse. Ils ne sont pas dans un raisonnement plus complexe. Ils ont une hypothèse, ils ont une connaissance, l'objectif de la recherche est de la confirmer. C'est après seulement, potentiellement hors consultation, qu'on peut aboutir à un raisonnement plus complexe mais cela nécessite une culture de la recherche que tout le monde n'a pas forcément. La notion de balance bénéfices/risques beaucoup de médecins ne la maîtrisent pas forcément et elle est pourtant importante pour choisir et prioriser les différentes prises en charge pour un problème de santé. En consultation la priorité c'est de confirmer l'hypothèse, confirmer ce qu'on souhaitait faire.