

FUTURS FANTASTIQUES

Vendredi 10 décembre 2021

Grand Auditorium

Gautier Poupeau, *En quoi l'intelligence artificielle constitue pour l'INA une opportunité pour renforcer ses missions de valorisation et de conservation du patrimoine audiovisuel ?*

Merci d'avoir accepté que je puisse présenter cette intervention qui va être l'occasion pour moi de vous présenter un nouveau projet qui est en cours d'élaboration à l'INA qui prend place dans une transformation plus globale, et surtout de pouvoir vous présenter la réflexion qui a sous-tendu ce projet, les aspects techniques qui sont sous-tendus à ce projet et puis les aspects organisationnels : on vient juste d'en parler, donc c'est bien, ça va faire le lien avec les autres.

Désolé pour mes amis anglophones : mon anglais est mauvais (comme on dit en France « je parle anglais comme une vache espagnole »), et donc je vais me contenter de parler français en espérant que tout le monde pourra me suivre avec la traduction et les tweets différents.

D'abord quelques constats assez simples. En chiffres l'Institut national de l'audiovisuel c'est :

- 100 km linéaires d'archives conservées pour les supports originaux
- 30 pétaoctets d'archives numérisées ou nativement numériques
- ça représente 22 millions d'heures de programmes télé radio
- 2 millions de photos
- 3 pétaoctets d'archives du web
- on capte 24 heures sur 24 / 7 jours sur 7, 184 chaînes de télévision
- ça représente à peu près 40 millions de programmes pour environ 59 millions de diffusions référencées.

Alors oui ça impressionne beaucoup ces chiffres mais enfin ça impressionne tellement qu'en fait, ils n'ont plus aucune signification : à chaque fois que je dois expliquer mon job, je vois bien les gens qui me disent "ah oui oui ça fait beaucoup quand même hein 20 millions d'heures" et le problème c'est que ça représente quoi ? Je sais pas, mon papa et ma maman ne comprennent toujours pas le travail que je fais à l'INA, j'en suis très triste, mais c'est ainsi.

Alors évidemment, vous allez me dire, *comment exploiter cette collection ? Vous avez certainement des moyens de le faire, il y a des bibliothèques numériques etc.* Oui effectivement l'INA, qui est un établissement public à caractère industriel et commercial, donc on est vous savez, entre les deux parties : on doit faire du business d'un côté et avoir des missions régaliennes de l'autre. On essaye de gérer ces deux parties comme on peut avec une offre qui est à destination du grand public avec notre site ina.fr où vous pouvez visionner gratuitement, librement, un certain nombre d'extraits issus de notre fonds, et notre Netflix à nous : Madelen, qui est une offre SVoD à destination du grand public. **ÇA C'EST POUR LE GRAND PUBLIC.** On a aussi des offres professionnelles B2B : c'est ce que vous voyez à la télévision française à chaque fois que vous voyez une archive, globalement, vous

pouvez vous dire qu'elle vient de chez nous et puis, pour les chercheurs, une offre de sites de centres de consultation de nos archives, ici même à la Bibliothèque nationale de France où l'Inathèque possède une salle où vous pouvez consulter nos archives, et puis un petit peu partout en France.

Le problème de tout ça, c'est que malgré tous nos efforts, ce n'est qu'une goutte d'eau dans l'océan. Si je mets ensuite les chiffres, les vrais chiffres, une fois que je vous ai fait la retape de tout ce qu'on fait chez nous, c'est que ina.fr c'est 400 000 contenus 80 000 heures de vidéo, Madelen c'est 13 000 programmes, bien sûr inamediapro c'est deux millions d'heures, mais je vous rappelle que tout à l'heure je vous ai parlé de 22 millions d'heures, donc c'est finalement une goutte d'eau.

La question du coup qui se pose à nous, c'est que à côté de nous il y a une société qui se pose un certain nombre de questions autour de la manière dont les médias rendent compte de cette société, comment on met à disposition, comment la parole des hommes et des femmes est perçue à la télévision, quel est le temps d'écran de tel ou tel homme politique (ne donnons pas de nom, on va faire encore gonfler le nombre de temps où on parle de lui). Donc toutes ces questions-là sont au cœur d'interrogations sociales très fortes, et comme vous le voyez ici j'ai pris quelques copies d'écran. Evidemment la problématique des fake news.... On voit bien que finalement la question qui sous-tend la société, c'est comment notre société, avec ses biais, avec sa manière de fonctionner, est répercutée dans les médias. Et donc toute la question qui se pose à nous à partir de ça, c'est finalement comment exploiter au mieux cette masse de données. Effectivement notre mission principale c'est de conserver, donner accès aux personnes accréditées (en particulier les chercheurs à travers nos différents centres de consultation), mais est-ce qu'on ne peut pas aller au-delà de ça, est-ce que je ne peux pas rendre un petit peu plus tangible notre collection auprès de mon papa et ma maman qui évidemment se posent des questions sur ce que fait leur fiston tous les jours dans cette institution ?

Au-delà de ça, comment on peut répondre aux interrogations politiques et sociales, est-ce que c'est notre rôle en tant qu'institution de conservation ? Et si oui, si c'est notre rôle, comment le faire évidemment avec nos impératifs juridiques un peu complexes ? Je viens de recevoir l'analyse juridique de la direction juridique de l'INA : ça ne va pas être simple, mais on va finir par y arriver. Et puis finalement, comment rendre tangible cette masse ? C'est toutes ces questions-là qu'on s'est posées. Dire que l'INA est un média patrimonial, c'est intéressant, mais il faut qu'on aille plus loin et rendre toute cette masse accessible à tout un chacun.

Alors notre proposition, c'est une rencontre entre trois éléments :

- des corpus qui sont issus de notre rôle, de notre mission de dépôt légal de la radio et de la télévision, en particulier ici deux corpus qu'on a pré-identifiés pour le début de notre projet, qui sont donc les chaînes d'information continue – en France il y a cinq chaînes d'information continue publiques et privées – et puis les journaux télévisés, qui sont une bonne vision de la manière dont les médias, perçoivent l'actualité et la société. Ça, c'est les corpus.
- Ensuite les technologies : on en a parlé. Des technologies à deux niveaux : d'abord tout ce qui va concerner le stockage et le traitement des données en masse (tout à l'heure notre collègue de la bibliothèque nationale Tchèque a beaucoup insisté et

beaucoup montré l'importance – et j'y reviendrai – d'avoir une infrastructure qui puisse permettre de monter en charge et de gérer toutes ces données qui vont être calculées), et évidemment des systèmes d'analyse automatique et du son qui nous permettent de prendre en charge tous les éléments d'un document audiovisuel, aussi bien l'image, aussi bien le son, aussi bien le contenu, et c'est bien ça qui fait toute la particularité de ces documents, qui sont protéiformes.

- Et puis évidemment des questions autour de trois ensembles : d'abord la diversité dans les médias, comment se répartit la parole des hommes et des femmes à l'intérieur des médias, comment les médias traduisent l'actualité, quelles sont les personnes, quels sont les lieux, quels sont les thèmes qui sont traités, cités, vus dans les médias, et puis enfin quelle est l'offre de programmes qui sont proposés par les médias et comment elle évolue dans le temps.

Et en fait, cette rencontre pour aboutir à quoi ? Eh bien pour aboutir à un site qui sera disponible, accessible à tout le monde, un site qui proposera des tableaux de bord interactifs pour montrer des chiffres, des indicateurs sur la durée – toujours les mêmes indicateurs dans la durée – autour de ces différents thèmes qu'on a répartis pour l'instant en quatre parties, qui sont les personnalités, la diversité, les lieux, les mots et thèmes qui sont dans l'actualité. Et pour l'instant avec 3 corpus : le premier corpus que sont les journaux télévisés, le second corpus qui sont les chaînes d'information continue, et puis le troisième corpus qui sont tous nos programmes et toutes les métadonnées de notre programme.

L'idée de ce site, c'est donc de pouvoir proposer en ligne, à partir du deuxième trimestre 2022, des tableaux de bord qui vont nous permettre, dans le temps, d'avoir une vision réelle des faits objectivables. Il ne s'agit pas pour nous de décrire ou d'analyser les faits et les chiffres qu'on va sortir, mais juste de les mettre à disposition, de constituer finalement ce que Jean-Philippe Genet, médiéviste français, appelait la métasource, c'est-à-dire ce premier niveau d'analyse qui va nous permettre ensuite d'élaborer, aux journalistes, aux chercheurs, aux curieux, d'élaborer finalement l'analyse qu'ils veulent faire sur ces différents éléments. Cette infrastructure fait l'objet d'une subvention dans le cadre du plan de relance français.

Pour rentrer un petit peu plus maintenant dans les détails, parce que, évidemment, le site, c'est la partie immergée de l'iceberg : en réalité pour pouvoir faire fonctionner ce site, il y a toute une infrastructure qui est nécessaire et on ne peut pas mettre en place des algorithmes de génération automatique de données sur autant de médias sans avoir une infrastructure sous-jacente. Cette infrastructure est en trois parties :

- une première partie qui concerne la création des données (je vais revenir dessus) avec des algorithmes qui ont été mis en place soit à l'INA soit avec des solutions externes
- une deuxième partie qui est tout ce qui va concerner le stockage et le traitement de la donnée en masse
- et puis, enfin, une troisième partie qui va être la brique d'exposition de visualisation autour de ce site web que je viens de vous présenter, avec la possibilité de reprendre ces différents éléments de ce site web au sein de nos différentes offres éditoriales ou sur data.gouv pour les données mises à disposition librement et gratuitement.

Je vais essayer de décrire assez rapidement ces différents éléments. Au niveau de la création des données, déjà au niveau des algorithmes qu'on va utiliser, il y a toute une série

de traitements qui vont être effectués : la transcription de la parole en texte (effectuée avec une solution du marché qui est Vocapia), extraction des entités nommées (c'est-à-dire qu'on va repérer les noms de personnes, de lieux, d'organisations au sein de ces textes et les relier à nos propres référentiels par l'intermédiaire de Wikidata, ce qui va nous permettre ici de pouvoir travailler autour du thème, autour de la notion de contenu à l'intérieur de nos différents corpus). L'idée, à terme, (j'y reviendrai) étant aussi de prolonger ce travail avec d'autres algorithmes de traitement automatique du langage. ÇA, C'EST LE PREMIER NIVEAU.

Deuxième type d'algorithmes : l'identification des visages. Un chercheur de l'INA, Pierre Letessier, a mis au point un logiciel qui s'appelle Trombinos qui nous permet aujourd'hui d'identifier 70 000 visages automatiquement et qui, en plus, permet au fur à mesure de rajouter des nouveaux visages, ce qui nous permettra d'alimenter – et on reviendra sur le rôle organisationnel des différents professionnels de l'information.

Troisième algorithme, le plus important : InaSpeechSegmenter. Là aussi c'est donc un logiciel (open source pour le coup, celui-ci) mis au point par un chercheur de l'INA, David Doukhan, qui permet d'analyser automatiquement le son et donc les paroles homme, de pouvoir bien identifier la parole des hommes, la parole des femmes, le bruit, et la musique pour qu'on puisse avoir des indicateurs sur cette répartition.

Il y a d'autres algorithmes qui entrent en jeu : on a aussi des systèmes d'OCR, des systèmes de classification automatique d'images pour réussir à segmenter les différentes images -(je pense que les gens qui étaient là hier et qui ont vu la présentation d'Olivio Segura sur les journaux télévisés ont pu voir le travail qui a été fait autour de ce système de classification automatique). On a aussi des systèmes de reconnaissance d'image, de détection des jingles aussi – on va s'attaquer maintenant à la radio, l'image est une chose, mais on a aussi le son ! – et aussi, et évidemment, sur l'ensemble des métadonnées, des référentiels qui sont mis au point depuis la naissance de l'INA par nos professionnels.

Donc ça c'est les algos, les algos, c'est une chose, mais il faut évidemment avoir une plateforme derrière sur laquelle on va faire tourner ces algos et en plus une plateforme qui est capable de ce qu'on appelle « monter en charge », *scaler*, gérer la scalabilité, avec deux niveaux : un premier niveau qui est ce qu'on appelle un « moteur de workflow ». On utilise un moteur de workflow qui s'appelle Media Cloud AI qui est une plateforme de micro service open source qui a été mise au point et qui est maintenue par France Télévisions et puis une société qui s'appelle Media-IO, et qui va nous permettre justement de rajouter au fur à mesure où on a des nouveaux corpus, de rajouter des capacités de calcul au sein de notre système, et puis bien sûr une plateforme, une infrastructure technologique : à l'INA tous ses algorithmes sont déployés sur notre propre datacenter, sur nos propres datacenters, sur nos propres machines, on n'utilise pas du tout le cloud, tous ces algorithmes sont répartis sur deux sites distants pour permettre évidemment la bascule en cas de problème sur un des deux sites, avec un cluster Kubernetes, donc c'est un système qui permet automatiquement de monter en charge au fur à mesure des besoins de calcul. Cette plateforme nous a pris énormément de temps à mettre en place, et c'est ça qui est important et à avoir en tête : c'est que mettre en place des algorithmes, mettre en place des tests, faire des POC [proof of concept] autour de l'intelligence artificielle c'est une chose, par contre l'industrialisation est beaucoup plus complexe. Donc le conseil que je donnerais c'est d'abord penser à l'industrialisation avant d'aller voir les algorithmes et c'est bien dans ce sens-là qu'on a

essayé de le faire, pour être capable ensuite, eh bien d'ajouter des choses au fur et à mesure sans avoir de problématiques complexes.

Sur le stockage et le traitement des données, on déploie à l'INA depuis maintenant plusieurs années (ça fait maintenant cinq ans qu'on travaille sur ce projet), un projet qui s'appelle le lac de données qui vise à centraliser l'ensemble des données de l'INA au sein d'une seule infrastructure technique, qui est composée elle-même de plusieurs types de bases de données et qui vont nous permettre de stocker les données de manière brute, et ensuite de séparer les données de la manière d'utiliser ces données, et donc d'avoir une plus grande souplesse à l'intérieur de notre système d'information pour exploiter toute cette masse de données. Donc tous les calculs qui sont faits par nos différents algorithmes arrivent dans ce système de stockage et ensuite sont répartis en fonction des besoins dans différents autres systèmes de stockage pour en permettre l'exploitation. C'est ce qu'il y a ici sur ce petit diagramme où j'explique comment, à partir des différentes données primaires qui sont calculées et stockées en brut, on va pouvoir ensuite exploiter ces données et les réexposer pour leur exploitation à l'intérieur du site web que je vous ai montré tout à l'heure.

Enfin tout ça ne peut pas fonctionner si on n'a pas une organisation, et donc différents rôles qui se répartissent. Ici j'ai fait une analogie entre finalement les différents rôles nécessaires par rapport à ce genre de projet, avec un ballet. On a, comme dans un ballet, différents intervenants qui vont jouer ce rôle :

- l'architecte de données – ici le compositeur du ballet – qui va organiser le système d'information, qui va aider à la modélisation des données, déterminer comment rendre cohérent l'ensemble de ce système et comment on va pouvoir jouer ensuite avec
- le chorégraphe, les data scientists qui vont pouvoir créer les algorithmes, les mettre en place, les concevoir, et donc c'est eux qui vont modéliser la donnée
- Mais les data scientists ne sont rien sans les autres. On a beaucoup parlé des data scientists ces dernières années, mais en fait ils ne sont rien sans les autres acteurs, et en particulier les ingénieurs de la donnée. L'orchestre, finalement, c'est eux, qui vont créer toutes ces chaînes de traitement qui permettent de faire fonctionner les algorithmes entre eux, qui vont permettre de stocker les données, qui vont permettre de les exploiter, qui vont permettre de les mettre à disposition.
- Et évidemment aussi ici mes petits danseurs, qui sont donc tous les professionnels de l'information de l'INA. Hier je pense qu'Olivio a énormément parlé de l'importance du rôle des professionnels de l'information dans la mise en place de ce genre de système, la base de visages : évidemment, aujourd'hui, il y a soixante-dix mille visages mais s'il n'y a pas de professionnels de l'information capables de la maintenir, de la gérer, on ne pourra pas aller beaucoup plus loin. Les modèles de machine learning pour pouvoir découper automatiquement les différents programmes, les JT et les chaînes d'information continue, si on n'a pas des gens qui sont capables de comprendre le contenu et qui sont capables d'accompagner les data scientists (dont ce n'est pas le travail), pour pouvoir mettre en valeur les différents contenus, eh bien ça ne fonctionnera pas et donc il faut bien qu'on puisse avoir ce différent couple ensemble : les ingénieurs de la donnée et les data scientists (très tech), et puis les métiers qui travaillent main dans la main. Et l'exemple d'Olivio est intéressant puisqu'il vient du métier et il a travaillé même physiquement dans les mêmes bureaux et les mêmes endroits que les ingénieurs de la donnée et data scientists.

- Enfin un rôle prépondérant, je ne serais pas sur scène et je ne pourrais pas vous parler aujourd'hui sans eux : les ingénieurs du son. Dans le milieu technique c'est ce qu'on appelle les DevOps, c'est-à-dire les gens qui vont être capables de mettre en musique finalement tout ça, de permettre d'avoir de la lumière, d'avoir le bon son, d'avoir les bonnes machines au bon moment avec les bonnes performances, de déployer toutes ces technologies de la manière la plus simplifiée possible. Et puis aussi le DataOps (on appelait ça les administrateurs de bases de données avant, les « DBA ») : aujourd'hui ces DataOps – sont au cœur aussi du travail. Il est nécessaire qu'ils soient rapprochés des ingénieurs mais aussi du métier pour bien comprendre les enjeux qui sont sous-jacents et donc c'est bien tous ces rôles-là ensemble qui permettent d'aboutir à ce résultat.

Ce que je voulais finir pour conclure (je crois qu'il me reste encore un petit peu de temps mais je vais aller très vite), ce que je voulais dire, c'est que au-delà du machine learning, au-delà des algorithmes, au-delà des modèles et au-delà des technologies, c'est bien à la fois un socle technologique qu'il faut mettre en place pour réussir à déployer ses systèmes en production et aussi pour réussir à faire en sorte qu'ils puissent monter en charge et répondre à différents besoins. Ici je vous ai montré le besoin auquel on répondait et c'est bien en ça, grâce à ces différentes technologies, grâce à ces différentes organisations que du coup l'intelligence artificielle va pouvoir nous permettre, nous en tant qu'institution de conservation de prendre un rôle central dans les débats de la société en permettant une analyse exhaustive ou du moins qui tend vers l'exhaustivité de l'ensemble des collections qui rentrent tous les jours à l'intérieur de nos institutions.

Voilà, merci.

Questions :

Thank you Gautier for your very interesting presentation and indeed you're very ambitious and I will be very interesting to follow your road. Other any questions for Gautier?

Bonsoir, la plate-forme que vous avez présentée, le site web, il ressemble beaucoup à l'architecture qui a été mise place par l'AFP : j'aimerais savoir si vous vous en êtes inspiré. Et deuxième question est-ce que vous pouvez nous parler un peu de la monétisation, comment vous allez procéder à ça, puisque vous à vous avez abordé le sujet brièvement.

Gautier Poupeau :

Alors premier niveau sur les questions : non pas du tout, on s'est pas du tout inspiré de l'AFP. En l'occurrence on s'est plutôt inspiré de notre site ina.fr pour mettre au point, d'autant plus que là c'est du work in progress donc il y a encore beaucoup de choses à faire. Ce qui est important, et là où on va être différenciant par rapport au travail (vous êtes peut-être de l'AFP), en tout cas par rapport au travail de l'AFP, c'est que me semble-t-il le travail de l'AFP est essentiellement tourné autour de la fake news et de l'infox, là où nous, notre travail va être essentiellement tourné vers l'établissement d'indicateurs et de faits vraiment objectivables. C'est-à-dire qu'il n'y aura pas du tout d'analyse, en tout cas pas sur ce site-là (éventuellement sur ina.fr et la revue des médias qui sont d'autres sites qui sont liés à l'INA, on pourra reprendre ce genre d'indicateurs), et puis les indicateurs seront proposés de manière constante, toujours les mêmes, pour réussir à avoir une constance à ce niveau-là.

La monétisation. Alors, l'ensemble des tableaux de bord, l'ensemble du site et les données elles-mêmes seront gratuites, accessibles librement et gratuitement, les données qui permettent de faire les graphes seront en open data, pas les données brutes, pas les transcriptions, pas les données de « tel visage apparaît à tel moment », par contre les données d'indicateurs, elles, seront en open data. Sur la monétisation on va voir au fur et à mesure, autour d'API, s'il y a un marché autour de nos différents partenaires, aussi bien partenaires broadcast, chaîne de télé, médias ou autres qui seraient intéressés par une offre payante autour d'API pour avoir directement ces données à l'intérieur de leur système d'information mais pour l'instant c'est pas l'objectif premier.

Emmanuelle a une question je crois.

Questions :

Merci pour cette présentation, ça a été assez rapide sur la partie algorithmes. Je pense notamment parce que effectivement il y a un autre collègue de l'INA qui l'a présentée assez longuement hier dans la track française, je ne suis pas sûre que tout le monde l'ait vu et du coup j'avais envie de demander si tu pouvais résumer quels étaient les principaux challenges en termes d'applications de l'IA sur le type de contenu que vous avez.

Gautier Poupeau :

Vaste question..., et j'ai deux minutes, ok super.

Le travail qu'on a fait, il est à deux niveaux et systématiquement on fait ce travail à deux niveaux : d'abord la structure du média et ensuite le contenu du média. Là où on travaille habituellement, quand on parle d'intelligence artificielle, immédiatement ce qu'on pense c'est aide à la documentation, voire même, puisque c'est la question que me pose mon comex [comité exécutif] systématiquement, c'est combien de documentalistes je vais avoir en moins si vous mettez en place ce genre de choses. Evidemment je refuse de répondre à cette question systématiquement, mais c'est vrai que c'est le 1er niveau qu'on a en tête.

Au niveau du contenu, ce qui est complexe c'est l'hétérogénéité (alors que ça soit le contenu ou la structure), au niveau du contenu ce qui est complexe c'est le monde ouvert, c'est le fait qu'on a de tout à l'intérieur. C'est-à-dire que quand je parle d'une base de visages j'ai pas que les hommes politiques, j'ai pas que les sportifs, j'ai pas que les artistes, j'ai toutes les personnalités qui peuvent apparaître à un moment ou un autre à l'intérieur de notre système. Donc ce qui est assez compliqué, sur la partie contenu, c'est ce côté « monde ouvert », qui est quelque chose qui fait très peur aux data scientists, puisqu'on peut avoir tout et n'importe quoi, alors que tout le machine learning repose sur le principe d'avoir des modèles d'entraînement, donc d'avoir entraîné sur des corpus tout à fait précis, et donc là à ce moment-là ça pose un certain nombre de difficultés.

Sur la partie structure finalement c'est aussi un peu la même problématique. La problématique, c'est l'hétérogénéité des structures : un JT c'est pas la même chose qu'un jeu, c'est pas la même chose qu'un magazine, c'est pas la même chose qu'une retransmission sportive, etc. Et en plus, toutes ces choses-là évoluent dans le temps : si vous regardez le plateau de télé de France télé, de France 2 aujourd'hui c'est pas le même qu'il y a cinq ans et c'est encore moins le même que dans les années 90, et je ne vous parle même pas dans les années 70 où c'étaient des fonds à peu près grisâtres et où on ne peut à

peu près rien faire en termes de machine learning parce que c'est une horreur : il n'y a aucun aucune façon de réussir à avoir des éléments différenciants par rapport au reste.

Donc on est obligé de faire un travail, et donc de partitionner les choses, et du coup de travailler par corpus parce que l'idée que l'IA universelle qui va fonctionner sur l'ensemble de nos corpus, l'ensemble de notre fonds, en fait elle n'existe pas malheureusement pour moi, même si on aurait bien voulu, et du coup on a été obligé de segmenter par cas métier, cas d'usages métier et grands ensembles de corpus. Donc là je vous en ai présenté deux puisque c'est là-dessus qu'on a beaucoup travaillé, les chaînes d'information continue d'une part les journaux télévisés d'autre part et puis au fur et à mesure de notre projet (*je vais changer un peu l'image qu'on voit à l'écran*), au fur et à mesure de notre projet on va augmenter ces différents corpus et du coup spécifier des différents traitements par rapport à chacun de ces corpus. On s'attaque normalement l'an prochain aux matinales radio qu'on transcrit déjà mais qu'on a du mal à exploiter parce qu'une matinale radio entre l'amuseur de 9h, le journal et puis l'interview, vous avez différentes parties et si vous voulez faire une analyse de traitement automatique de la langue sur l'ensemble de la matinale ça n'a pas de sens : en fait vous êtes obligé de restructurer les différents éléments et donc il va falloir qu'on travaille pour qu'automatiquement on puisse mettre à disposition ces différents éléments.

J'ai oublié de dire quelque chose dans le site pardon c'est que l'ensemble de ces indicateurs qu'on va mettre à disposition vont être mis à disposition entre deux et quatre jours après diffusion, donc c'est ça l'enjeu : c'est évidemment d'être en prise avec l'actualité. Actuellement pour documenter le journal c'est à peu près un à deux mois de travail, enfin de décalage entre la diffusion et le résultat de l'analyse par nos documentalistes. Là, l'objectif c'est qu'entre deux et quatre jours après diffusion vous ayez l'ensemble de ces indicateurs qui soient disponibles sur la plate-forme.