RAPPORT

Projet: Europeana Newspapers

Convention de subvention: 297380

Titre du projet: Accès à la presse historique européenne en ligne

---

D6.3.1 Rapport de la journée d'information Europeana Newspapers à la Bibliothèque nationale de France, Paris

---

Version: 1.0

Auteurs: **Joséphine Renard (Bibliothèque nationale de France)**

Contributions: **Jean-Baptiste Vaisman, Elisabeth Freyre, Marion Ansel (Bibliothèque nationale de France)**

Projet co-financé par la Commission européenne dans le programme de support politique ICT		
Niveau de diffusion		
P	Public	x
C	Confidentiel, seulement pour les membres du consortium et des services de la Commission	

Historique des révisions

Révision	Date	Auteur	Organisation	Description
0.1	03-12-2014	Joséphine Renard	BnF	Première ébauche
0.2	10 – 12- 2014	Jean-Baptiste Vaisman et Ioannis Anagnostopoulos	BnF	Corrections
0.3	12-12-2014	Marion Ansel	BnF	Relecture et élaboration des annexes
0.4	15-12-2014	Elisabeth Freyre	BnF	Relecture
0.5	16-12-2014	Friedel Grant	LIBER	Examen et édition ultérieure
1.0	17-12-2014	Clemens Neudecker	SBB	Examen interne et version finale

Traduit du rapport D6. 3. 1 soumis à la Commission européenne le 17 décembre 2014

Table des matières

1. Résumé 4
2. Vue d'ensemble de l'évènement 5
3. Présentations et débats 7
4. Public et impact 18
ANNEXE I : Programme 21
ANNEXE II : Liste des participants 23
ANNEXE III : Couverture médiatique 26
ANNEXE IV : Résultats du questionnaire 31

1. Résumé

Ce rapport revient sur la journée d'information du projet Europeana Newspapers, organisée à Paris par la Bibliothèque nationale de France (BnF). Les journées d'information sont l'occasion pour les partenaires de présenter le travail effectué au niveau national et de promouvoir le projet ainsi que de relayer les messages suivants :

1. L'importance de la collaboration à l'échelle nationale et européenne pour rendre accessible en ligne la presse historique (selon différentes thématiques : sociale, économique, culturelle, recherche, technique etc.)

2. Comment répondre aux problèmes techniques associés à la mise en ligne de contenus de presse numérisée.

3. L'importance des collections de presse numérisées en tant que sources (par la promotion de contenus désormais disponibles).

La France a été le dernier pays à organiser une journée d'information Europeana Newspapers, une journée marquée par une forte participation. 88 personnes ont assisté à cet événement qui a permis de faire connaître le projet Europeana Newspapers au niveau national, à un public composé de chercheurs, de bibliothécaires, de professionnels du numérique, du patrimoine, de l'informatique, de la conservation, des archives, de l'audiovisuel et des médias, de la presse ainsi que d'enseignants.

Les intervenants de la journée étaient des professionnels des bibliothèques (Bibliothèque d'Etat de Berlin, la Bibliothèque universitaire d'Innsbruck) des chercheurs et des professionnels du numérique et de la sémantique représentés par CCS (Content Computer Specialists) et la société Syllabs. Ce panel diversifié a permis de définir au mieux le projet, son utilité et les attentes des chercheurs.

La diversité des interventions a permis au public de découvrir les divers aspects et possibles applications du projet Europeana Newspapers. Ce fut aussi l'occasion pour les acteurs du projet de mieux comprendre les attentes des chercheurs concernés par la numérisation de la presse historique.

Cet événement a été couronné de succès. Le questionnaire rempli par les participants révèle que 90% des personnes présentes ont qualifié les présentations d'intéressantes voire de très intéressantes et qu'elles consulteront le contenu numérisé désormais disponible grâce au projet (Cf. Annexe IV pour les résultats du questionnaire).

2. Vue d'ensemble de l'évènement

La journée d'information Europeana Newspapers s'est déroulée le 27 novembre 2014 au Grand Auditorium de la BnF. Cette journée avait pour but de présenter les enjeux du projet, auquel participe la BnF depuis trois ans, et de concevoir comment chercheurs et entreprises pouvaient utiliser cette nouvelle mine d'information.

Cette journée s'est déroulée en deux temps (Cf. Annexe I pour l'agenda complet) :

- Après une introduction générale, la matinée fut consacrée aux résultats techniques produits depuis la mise en place du projet à la BnF. Ces résultats ont permis un accès amélioré au contenu grâce au travail effectué en matière d'OCR (Optical Character Recognition), d'OLR (Optical Layout Recognition) mais aussi l'identification des entités nommées. De nouveaux modes de navigation ont donc vu le jour grâce à ce projet. Les sujets de l'indexation des articles de presse, des titres et sous-titres ainsi que des entités nommées (noms de lieux, de personnes) ont également été abordés.

- L'après-midi visait à mettre en avant l'utilité d'un tel projet tant du point de vue de la mémoire collective que de la recherche scientifique. Une table ronde a permis de s'attarder sur ces deux points et de définir les attentes, en constante évolution, des utilisateurs. Les chercheurs utilisant déjà les collections de presse numérisée sur Gallica et d'autres bibliothèques numériques ont fait part de leurs expériences en tant qu'utilisateurs et exprimé leurs attentes quant à l'accès au contenu et les nouvelles méthodes de recherche que ce nouveau portail offre. Enfin, la société Syllabs a expliqué comment les résultats d'Europeana Newspapers pouvaient être exploités dans son activité.

Ces deux temps forts ont permis à un large public composé de chercheurs, bibliothécaires, archivistes, historiens, professionnels de l'édition et de laboratoires d'informatique et de linguistique, de connaître ce projet européen de numérisation de la presse historique via le moteur de recherches de la bibliothèque européenne (TEL, The European Library).

88 des 123 inscrits ont participé à l'évènement au Grand Auditorium de la BnF (Cf. Annexe II la liste des participants). Le public était composé notamment de bibliothécaires de la bibliothèque interuniversitaire de la Sorbonne, de la bibliothèque Sainte Geneviève, de la bibliothèque du Sénat, de la bibliothèque du Patrimoine de Clermont-Ferrand, mais aussi d'archivistes des Archives départementales de la Gironde, des enseignants du CELSA (Ecole des hautes études en sciences de l'information et de la communication), de Telecom Paris Tech, de différentes universités et laboratoires de recherche en informatique et de linguistiques de Paris, de Rennes et de Rouen. Enfin, la présence de plusieurs sociétés de numérisation comme Syllabs, I2S, CEA, Arkhênum ou Immanens témoigne de la diversité du public concerné.

Un dossier contenant, notamment une brochure sur le projet et une carte postale réalisée en français à partir d'un article de presse français, a été remis à chaque participant.

L'évènement s'adressant à un public essentiellement français bénéficiait d'une traduction français/anglais et d'un enregistrement vidéo qui après post-production sera diffusé en ligne sur le site internet de la BnF et celui d'Europeana Newspapers début 2015.

Cette journée d'information s'inscrivait dans une série d'événements dont le thème principal était l'accès à la presse historique numérisée. En plus de la journée d'information Europeana Newspapers, deux autres évènements ont eu lieu à la BnF, le jour suivant (28 novembre 2014) :

- Une journée d'étude sur *La presse de la résistance à l'ère de la numérisation : nouvelles lectures et perspectives*[1]
- La conférence *SUCCEED in Digitisation, Spreading Excellence*[2]

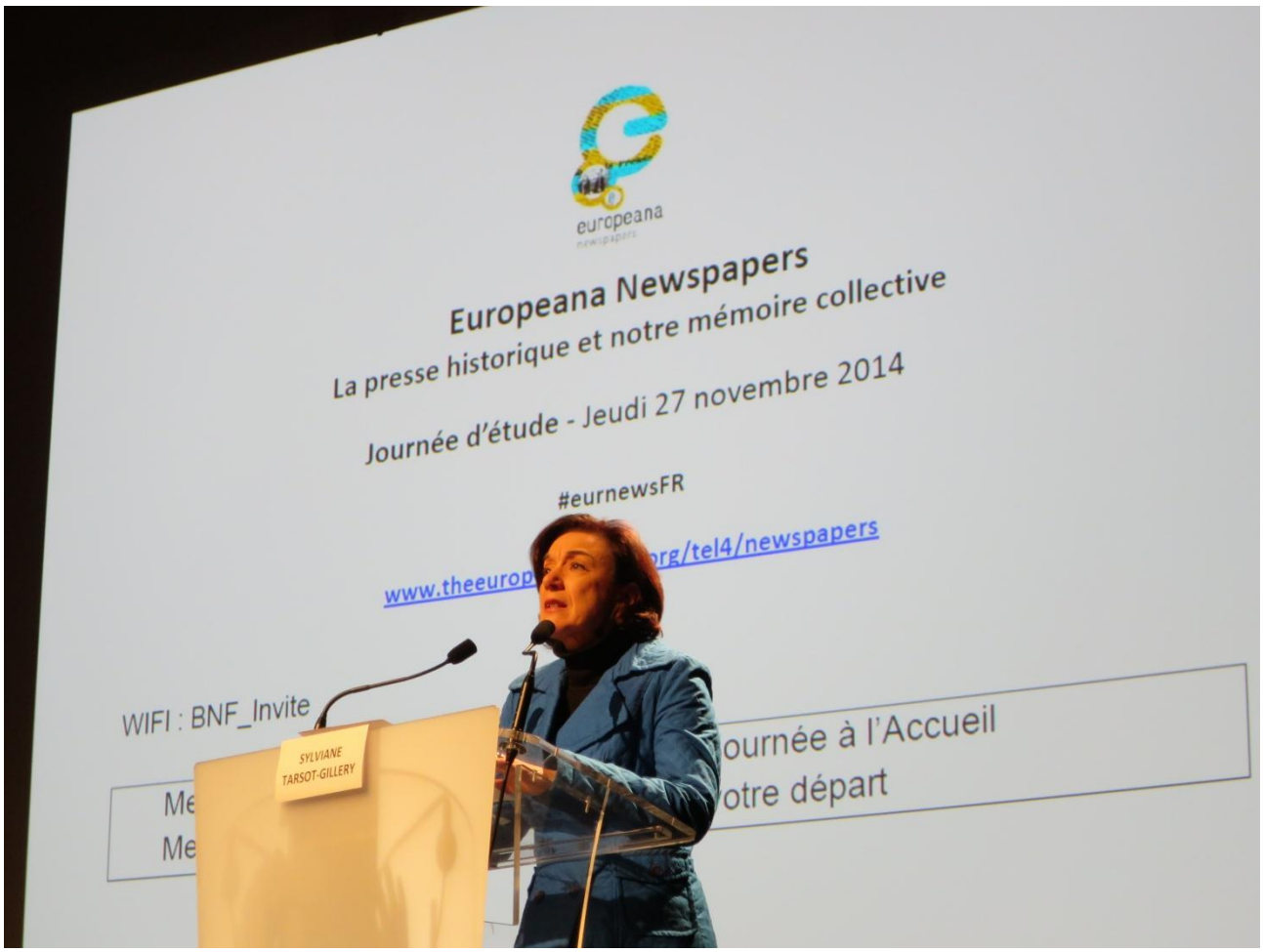
Une partie des participants a assisté à la fois à la journée d'information Europeana Newspapers et l'une ou l'autre des journées d'étude le 28 novembre 2014. Ces trois événèments ciblaient des publics relativement semblables.

[1] http://www.bnf.fr/fr/evenements_et_culture/auditoriums/f.journaux_resistance.html?seance=1223917800203

[2] http://www.bnf.fr/fr/professionnels/anx_journees_pro_2014/a.jp_141128_succeed.html

3. Présentations et débats

La journée a débuté par quelques mots d'introduction de Sylviane Tarsot-Gillery, directrice générale de la BnF qui a tenu à rappeler les raisons pour lesquelles la BnF s'est associée à cette initiative majeure pour la recherche et l'accès aux sources primaires que sont les journaux. Grâce aux technologies de l'OCR et de l'OLR, les chercheurs auront désormais un accès amélioré à la presse numérisée et une navigation facilitée à l'intérieur des contenus textuels. Le travail effectué avec la reconnaissance des entités nommées en français permet de promouvoir le français sur le web ce dont la BnF se réjouit. Alors que le projet Europeana Newspapers s'achève, il incombe désormais à la BnF de faire bénéficier les utilisateurs des collections de presse de ces avancées.



Sylviane Tarsot-Gillery à la journéee d'information à Paris

Le programme s'est poursuivi par trois tables rondes, chacune avec une problématique spécifique.

Table ronde # 1 : Pourquoi Europeana Newspapers?

La première table ronde a été modérée par **Elisabeth Freyre** de la BnF afin de présenter le projet et ses objectifs.

De gauche à droite : Ioannis Anagnostopoulos, Hans-Jörg Lieder, Elisabeth Freyre et Philippe Mezzasalma

La première présentation de **Philippe Mezzasalma**, Chef du service Presse au département Droit Économie Politique à la BnF, était intitulée ***Les collections de presse numérisées à la BnF : succès et limites des premières numérisations***[3]***.*** Philippe Mezzasalma est revenu sur le contexte de l'engagement de la BnF dans ce projet européen. La bibliothèque numérique Gallica disposait jusqu'alors d'entre 5,5 et 6 millions de pages océrisées. Cependant, ce traitement était surtout réservé aux grands quotidiens avec des résultats de qualité inégale. De plus, la navigation sur Gallica était succinte dans le sens où elle ne permettait pas de rubriquage ni de navigation dans le contenu. Les résultats d'une recherche dans Gallica étaient le plus souvent instables et leur pertinence aléatoire.

Lorsque la BnF s'est engagée dans ce projet de numérisation que représente Europeana Newspapers, elle a décidé de sélectionner des modèles éditoriaux cohérents et définitifs. Se sont

[3] http://www.slideshare.net/Europeana_Newspapers/pmezzalsalma-enp-paris-27112014

donc imposés en priorité, des titres morts et vivants publiés entre les XIX^{e} et XX^{e} siècles, des quotidiens et des hebdomadaires d'informations nationales et régionales. La BnF s'est efforcée de définir un panel regroupant l'ensemble des sensibilités politiques et culturelles nationales. Ce type de presse offrait, de plus, une rubrique internationale en adéquation avec l'ensemble des pays européens participant au projet.

Le projet Europeana Newspapers a permis de travailler à l'amélioration de l'OCR des journaux numérisés et à la segmentation à l'article de périodiques. En tout, ce sont pas moins de 24 titres de journaux qui ont pu soit être océrisés soit olrisés. Le but de ce projet est de répondre aux attentes majeures du public à savoir une meilleure identification des images et des légendes, une fiabilité accrue des résultats de recherche en mode plein texte, une plus grande pertinence des résultats d'une recherche par article ou type d'articles, un accès direct à l'article recherché, une lecture du journal reprenant le rubriquage d'origine et enfin une navigation plus facile à l'intérieur des contenus.

Grâce au moteur de recherche développé par TEL/The European Library, les chercheurs bénéficient d'une visibilité plus aisée de la presse ainsi qu'une meilleure intégration et pertinence des résultats.

Hans-Jörg Lieder[4], Chef du département des services bibliographiques de la Bibliothèque d'Etat de Berlin a ensuite rappelé l'envergure de ce projet qui a commencé en février 2012 et se terminera en janvier 2015. Doté d'un budget d'environ 5,16 millions d'euros, il regroupe 18 partenaires dont une majorité de bibliothèques nationales d'Europe, 11 partenaires associés et 22 réseaux partenaires de 28 pays différents. Le projet regroupe donc quasiment l'Europe dans sa totalité. 8 millions de pages ont été océrisées et 2 millions de pages olrisées. Ce décalage s'explique par le fait que l'OLR est une technique plus longue car elle nécessite un retraitement des résultats automatiques. Ce travail a été effectué par la société allemande CCS (Content Conversion Specialists). La Bibliothèque nationale des Pays-Bas (KB) s'est chargée de coordonner la reconnaissance d'entités nommées en néerlandais et en allemand. 950 titres ont été numérisés en 20 langues différentes. Le journal le plus ancien date de 1618 et le plus récent de 1955.

[4] http://www.slideshare.net/Europeana_Newspapers/presentation-42715138

Carte situant les partenaires du projet (en rouge), les partenaires associés (en bleu) et les partenaires réseaux (en vert)

L'intérêt du moteur de recherche TEL/The European Library [5] est de fournir aux chercheurs l'accès au texte dans son intégralité. Début 2015, l'objectif sera d'augmenter le nombre d'occurences lors d'une recherche par date. Aujourd'hui plusieurs types de recherches sont possibles sur TEL : une recherche par zone géographique ou une navigation par titre. Un travail en étroite collaboration avec la Bibliothèque du Congrès a par ailleurs été effectué sur la structure de ces métadonnées.

Néanmoins, le besoin de travailler sur le contrôle de l'OCR fourni automatiquement pour atteindre un taux qualité de 100% est réel.

Ioannis Anagnostopoulos, Coordinateur- gestionnaire du projet européen Europeana Newspapers pour la BnF a ensuite défini l'apport de la BnF dans le projet[6]. Ce projet permet l'agrégation de la presse historique européenne, une plus grande visibilité sur le portail Europeana et sur les portails de bibliothèques partenaires mais surtout une amélioration des fonctionnalités de recherche, et ce grâce à l'enrichissement sémantique des données et métadonnées relatives au corpus sélectionné.

Ioannis Anagnostopoulos a mis en avant la contribution des partenaires du projet, notamment de la BnF, ayant procédé à l'océrisation et l'olrisation de plus de 2 millions de pages. Il a également décrit comment la BnF a collaboré avec d'autres scientifiques et partenaires techniques du projet

[5] www.theeuropeanlibrary.org/tel4/newspapers

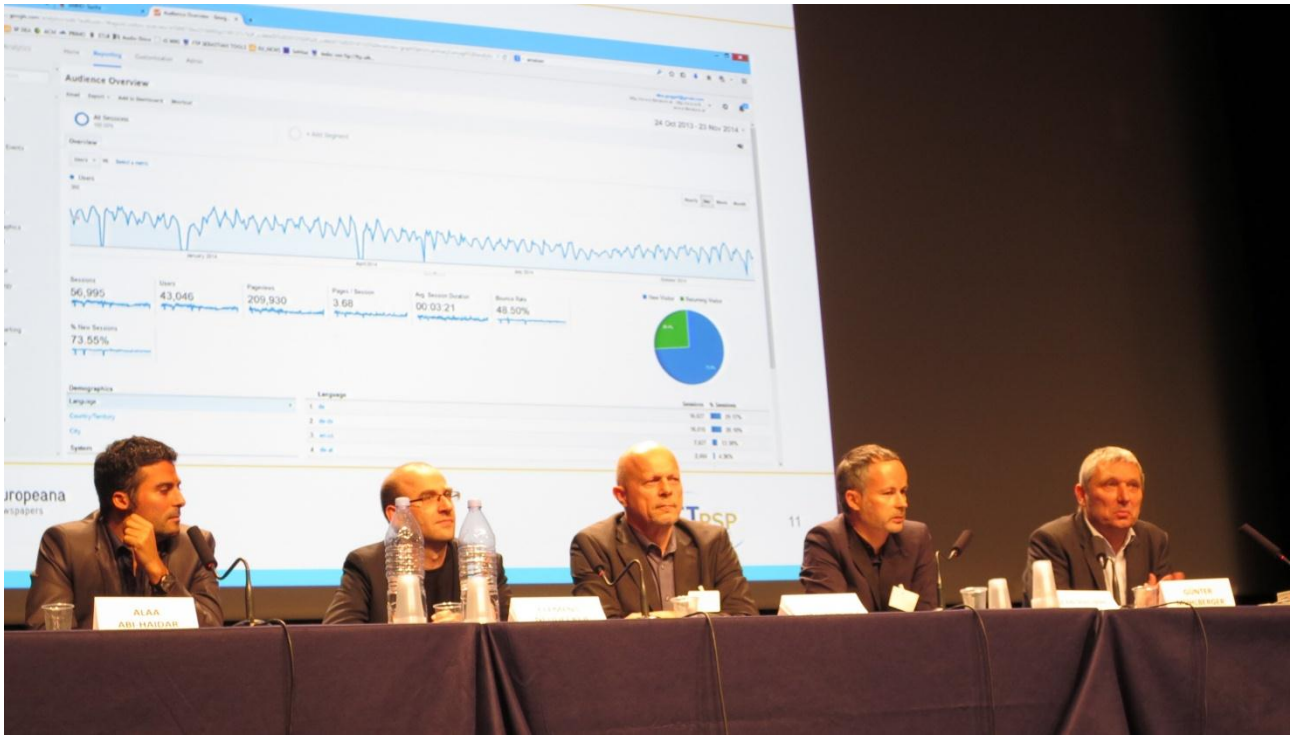
[6] http://www.slideshare.net/Europeana_Newspapers/presentation-of-the-bnf-at-their-information-day

(la Bibliothèque nationale des Pays-Bas (KB) et le laboratoire Lip6 à l'université Pierre et Marie Curie, UPMC) sur la reconnaissance des entités nommées (NER, Named-Entity Recognition).

Enfin, Ioannis Anagnostopoulos a démontré que des études récentes révèlent que le projet représente une amélioration des fonctionnalités de recherche et de consultation de la presse numérisée et comment le projet permettra d'augmenter la consultation du site Gallica.

Table ronde # 2 : Des innovations techniques au service des utilisateurs

La deuxième partie de la matinée modérée par **Jean-Philippe Moreux** de la BnF avait pour but de détailler les innovations techniques abordées précédemment.



De gauche à droite : Alaa Abi-Haidar, Clemens Neudecker, Claus Gravenhorst, Jean-Philippe Moreux et Günter Mühlberger

Günter Mühlberger est professeur et directeur de projets à l'Université d'Innsbruck, l'université qui a traité les 8 millions de pages en OCR du projet. Dans sa présentation intitulée ***Comment l'OCR aide les utilisateurs dans leurs recherches?***[7], il rappelle que la collaboration de l'Université d'Innsbruck avec la BnF remonte à 15 ans en arrière, le projet portait déjà sur les métadonnées, l'OCR et le format ALTO. A l'époque, on craignait que l'OCR ne soit pas suffisamment fiable mais l'investissement de Google en la matière en 2006 a prouvé qu'il était possible de lancer de vastes projets de numérisation avec océrisation. Aujourd'hui, le taux de fiabilité pour le français s'élève à 83,4 %.

Günter Mülhberger a ensuite évoqué les utilisateurs. Si certaines recherches ont un caractère ponctuel relevant surtout de la curiosité d'autres sont liées à un véritable travail de recherche.Dans ce cas, l'utilisateur peut alors passer des centaines d'heures à lire, à télécharger le contenu, à

[7] http://www.slideshare.net/Europeana_Newspapers/presentation-gunter

l'analyser , à échanger et à établir des liens avec les autres sources. Cela nous amène à nous interroger sur l'accès et la nécessité de le recontextualiser dans les différentes sources disponibles. Le chercheur doit rester critique envers ces sources et donc adopter la même démarche avec l'OCR. L'Université d'Innbruck s'est inspirée des travaux effectués par l'Université de Stanford pour justement tenir compte des marges d'erreurs. Ainsi, le chercheur a la possibilité de choisir parmi une liste de mots pour s'assurer que le portail recherche bien le mot demandé. Enfin, l'Université d'Innsbruck permet aujourd'hui aux utilisateurs de cliquer sur un lien automatique vers Wikipedia après une recherche d'entités nommées.

Claus Gravenhorst, Directeur de la stratégie de CCS, société responsable de l'olrisation des 2 millions de pages du projet, est revenu sur les différentes étapes de numérisation des journaux dans sa présentation : ***L'OLR, les différentes étapes vers des données de journaux structurées***[8]. Après l'étape de numérisation, une analyse de mise en page et de structure est effectuée afin de définir le contenu de la page numérisée. Puis un examen des données et une conversion des métadonnées permettent de reconnaître les éléments structurels du document. Les données ainsi produites sont ensuite retournées aux partenaires et à TEL. Toutes ces étapes sont nécessaires pour permettre à l'utilisateur d'obtenir un maximum d'informations sur le contenu des articles de presse recherchés.

Clemens Neudecker Coordinateur du projet Europeana Newspapers à la Bibliothèque d'Etat de Berlin nous a présenté ***Le développement des entités nommées en français dans la presse historique***[9]. Le principal intérêt pour l'utilisateur est de lier ces entités nommées à d'autres ressources en ligne comme les catalogues des autres bibliothèques par exemple. Des défis restent cependant à relever comme le fait que les outils de reconnaissance des entités nommées doivent être adaptés à chaque langue afin d'interpréter chaque phrase, chaque règle de grammaire, chaque contexte et structure de phrase. Il doit aussi pouvoir interpréter des tournures de phrases anciennes, ce qui nécessite un module supplémentaire.

Alaa Abi Haidar est un chercheur en fouille de données et systèmes complexes à la BnF et au laboratoire LiP6. Il a rappelé que l'objectif de la reconnaissance des entités nommées est d'indexer, résumer et classifier en vue d'enrichir la lecture numérique[10]. La désambiguïsation de certains termes reste encore à maîtriser. En effet, le mot Paris peut à la fois signifier la ville ou le prénom d'une personne; certains mots ne figurent pas dans tous les dictionnaires ou bien encore certains groupes de mots n'ont pas de limites prédéfinies et cela consiste autant de défis .

C'est pourquoi Alaa Abi Haidar a alors utilisé la méthode UNERD qui consiste une fois la reconnaissance des entités nommées effectuée, à classer les mots dans une classe grammaticale. On procède alors à une recherche dans les différents dictionnaires pour retrouver les correspondances sûres c'est à dire les entités nommées que l'on peut considérer comme fiables. Des données annotées manuellement permettent de comparer les résultats obtenus. La méthode UNERD, comparée à d'autres comme celle de Stanford, est la plus performante. La possibilité de surligner les résultats au sein de l'article sélectionné en facilite l'utilisation.

[8] http://www.slideshare.net/Europeana_Newspapers/enp-infoday2014-pptgravenhorst

[9] http://www.slideshare.net/Europeana_Newspapers/presentation-42715189

[10] http://www.slideshare.net/Europeana_Newspapers/presentation-alaa-abi-haidar

Table ronde # 3 : Quels usages, quelles pratiques, quelles attentes pour les usagers ?

L'après-midi de cette journée d'information a été marquée par une table ronde portant sur l'utilisation des corpus numérisés dans le cadre du projet Europeana Newspapers et les attentes des utilisateurs auxquelles le projet doit encore répondre.

Cette table ronde a été modérée par **Marc Minon**, directeur du portail Cairn. info, portail de revues francophones. On constate que les jeunes chercheurs ne conçoivent pas de travailler sans le numérique tandis que les chercheurs avertis ont vu leurs méthodes de travail évoluer. L'objet de cette table ronde était de montrer en quoi le numérique modifie les habitudes de travail des historiens et en quoi le projet Europeana Newspapers s'inscrit dans cette évolution.

De gauche à droite : Olivier Hamon , Sophie Kurkdjian, Philippe Tétart, Marc Minon et Laurent Martin

Laurent Martin est professeur d'histoire à l'Université de Paris III Sorbonne-Nouvelle, chercheur au laboratoire Intégration et Coopération des espaces européens (ICEE), chercheur associé au Centre d'histoire de Sciences Po et au CERLIS, enseignant à Paris III, Sciences Po et Ina Sup et membre du Comité d'histoire du ministère de la Culture. Laurent Martin est revenu sur les méthodes de travail qu'il avait employées pour rédiger sa thèse sur l'histoire de l'hebdomadaire *Le Canard enchaîné* en réfléchissant sur ce que les nouveaux outils numériques auraient pu lui apporter et comment ils auraient modifié son approche du sujet. La rédaction d'un *Dictionnaire de la presse en Europe au XX^{e} siècle* sera l'occasion pour lui de s'interroger sur l'utilité de ces outils.

Philippe Tétart est maître de conférences en histoire (Université du Maine), chercheur au VIP&S (université de Rennes 2 – EA3646) et expert auprès du Musée National du Sport. Philippe Tétart

travaille sur les rapports entre sport et médias, le journalisme sportif et les représentations sociales du sport. Il nous a fait part de l'avancée scientifique, tant en matière de méthode que sur le plan historiographique, que représente l'accès aux archives de presse numérisées. D'un point de vue de la méthode, la numérisation de la presse permet une liberté de recherches, impossible auparavant du fait de la lourdeur de consultation papier. D'un point de vue historiographique, la numérisation de la presse permet de croiser une multitude d'informations et de points de vue sur un thème spécifique. La numérisation de la presse s'est traduit pour ce chercheur par une véritable revisitation de ses objets de recherche.

Sophie Kurkdjian, docteur en histoire de l'Université Paris I, est chercheur invitée à la BNF, et chercheur associée à l'IHTP-CNRS. Elle travaille sur l'histoire de la presse illustrée, et notamment sur la presse féminine, du début du XXe siècle à la BnF. Sophie Kurkdjian a présenté les possibilités nouvelles de recherche offertes par la numérisation des titres de presse illustrée. En effet, elle permet de reconstituer le parcours d'hommes de presse lorsque les autres sources d'archives ont disparu et de travailler sur plusieurs titres simultanément et donc de mieux comprendre l'évolution formelle des journaux illustrés féminins ainsi que les enjeux associés.

Olivier Hamon, ingénieur de formation et spécialiste de web sémantique travaille à la recherche développement de la société Syllabs. L'intervention d'Olivier Hamon a porté un autre regard que celui des chercheurs sur la numérisation, plus orienté vers les utilisations potentielles de la presse numérisée. Cette table ronde a été l'occasion pour lui de faire un premier état des lieux des travaux effectués par Syllabs dans le cadre d'un autre projet sur la presse ancienne porté par BnF-Partenariats et son partenaire Immanens. Syllabs analyse les pages des périodiques afin d'en extraire des thématiques, des entités nommées et des sujets en se basant sur des référentiels existants.

Cette table ronde a été aussi l'occasion d'entendre les attentes des utilisateurs qui souhaiteraient pouvoir effectuer des traitements statistiques, des recherches d'occurrence, le calcul de nombres de mots, d'indépendance lexicale et pouvoir visualiser ces recherches sous la forme d'un graphique. Leur demande concerne la possibilité de récupérer les textes isolés, constituer eux-même un hypertexte pour pouvoir faire ces statistiques. Enfin, l'outil reste très lié au texte et beaucoup moins à l'image ce que l'on peut regretter pour la presse satirique, où le dessin compte autant que le texte, ou encore la presse féminine illustrée. Certains journaux numérisés présentent encore l'inconvénient de présenter la couverture ou la publicité à la suite du numéro et non pas à sa place originelle. Cependant, l'association du texte à l'image est un processus long et complexe.

Des participants ont souhaité revenir sur la notion d'erreur que pouvait engendrer la numérisation de la presse. En effet, la qualité de l'OCR a un impact majeur sur les recherches. Un utilisateur peut avoir accès à un texte contenant des erreurs, gênant ainsi la compréhension du texte, mais cette dimension est prise en compte. Cette notion d'erreur peut être contrée par une adaptation de l'interprétation de l'OCR. Il peut y avoir aussi des impacts sur la qualité de la numérisation, des noirceurs sur le journal entraînant des difficultés pour l'OCR à repérer les caractères. Il incombe aux acteurs du projet de tenir compte de ces problèmes.

Pascal Sanz, directeur du département Droit, Economie, Politique de la BnF, a profité de cette session de questions réponses pour rappeler que certains journaux français célèbres comme la *Dépêche du Midi* ou *le Canard enchaîné* ont dû renoncer à participer aux programmes de numérisation de la BnF pour des problèmes de droits. Ainsi, les illustrateurs du *Canard enchaîné*

se sont opposés à la numérisation de leurs dessins faisant perdre tout intérêt à la numérisation du journal. Des négociations sont en cours pour trouver des solutions.

Conclusions

Olivier Piffault, directeur du département de la Conservation à la BnF et **Pascal Sanz** sont revenus sur les points importants du projet évoqués tout au long de la journée.

Pascal Sanz et Olivier Piffault

Ce projet s'est inscrit dans le sillage des programmes de numérisation de masse pour lesquels on peut citer l'exemple de la BnF qui depuis 2005 a numérisé plusieurs millions de pages, une évolution constatée dans toutes les bibliothèques nationales et mises en avant par ce projet. Se pose alors la question de l'exhaustivité et l'on peut considérer que le lancement du projet en 2012 constitue une étape pour inciter les autres bibliothèques à poursuivre leurs efforts. C'est un tournant qui porte sur la qualité, l'utilité et le contenu de cette numérisation. L'une des premières conclusions du projet est la place fondamentale de l'OCR. Au terme des ces trois années, la question de l'OCR apparaît cruciale et sa qualité est déterminante pour les chercheurs. La segmentation à l'article, l'OLR, est un des points phares en terme de recherche et de confirmation des hypothèses de départ de ce projet. A côté de cet OLR, un autre apport technique important se profile. Il s'agit de la reconnaissance des entités nommées. L'alignement de ces entités nommées aux grands référentiels constitue avec la désambiguïsation une véritable nouveauté pour la BnF. Pour l'ensemble des techniques que regroupe la reconnaissance des entités nommées, il s'agit aujourd'hui de passer du stade expérimental au stade industriel.

Si les développements menés par Europeana Newspapers permettent d'approfondir la qualité de l'OCR, l'OLR et de la reconnaissance des entités nommées, la question de l'utilisation ultérieure se pose.

Une qualité de 100% de l'OCR n'est atteignable que sur quelques types de ce corpus. Il n'y a pas une démarche de qualité mais plusieurs pour se rapprocher de ce taux de réussite. Ce projet a confirmé la pertinence de l'utilisation d'outils déjà bien connus des acteurs de la numérisation comme TEI, METS, ALTO et qui ont été employés dans le cadre d'Europeana Newspapers. Le projet a mis au point un profil dans le cadre de la numérisation de la presse : European Newspapers METS/ ALTO profile (ENMAP) qui sera publié en janvier 2015. Si le projet peut paraitre s'être concentré sur la numérisation, son but reste avant tout l'usager et ses usages, la spécificité ou les besoins exprimés par les chercheurs, des besoins très spécifiques en terme d'analyse.

Bruno Racine

Le président de la BnF, **Bruno Racine** est revenu sur le défi que constitue la numérisation des collections de presse, source capitale pour les historiens. Malgré tout ce qui a été accompli, la tâche reste immense alors que les bibliothèques nationales sont soumises à des contraintes financières. Ainsi, le budget de numérisation en France est essentiellement dédié aux livres. La présence de la BnF dans le projet Europeana Newspapers, au-delà des indubitables avancées techniques, illustre l'engagement de l'établissement dans Europeana, un engagement qui

s’exprime à travers la participation à de nombreux projets européens dont l’objectif est d’enrichir les contenus et l’offre d’Europeana. Ce projet politique, initié par la Commission européenne en 2008, demeure à ce jour l’unique initiative culturelle à l’échelle du continent européen, une initiative qui rassemble non seulement des milliers d’institutions culturelles mais aussi des centres de recherche, des laboratoires, des industries créatives, des chercheurs et des développeurs.

4. Public et impact

L'évènement à rencontré un franc succès. 88 des 123 inscrits ont assisté à cette journée. 52 se sont abonnés à la lettre d'information du projet. Au vu de la liste des participants l'évènement, qui s'adressait à des professionnels et des chercheurs, a rencontré son public.

Participants à la journée d'information Europeana Newspapers

Comme la journée d'information Europeana Newspapers en Allemagne, les principaux objectifs de la journée d'information française étaient :

1. Informer les utilisateurs travaillant sur la presse historique de la disponibilité des contenus dans le cadre du projet ainsi que des fonctionnalités de recherche améliorées via le moteur de recherche TEL, notamment grâce aux développements techniques appliqués aux contenus numérisés.
2. Faire connaître le projet Europeana Newspapers.
3. Souligner l'importance de la presse historique en tant que source pour les chercheurs dans les humanités numériques.
4. Aborder les changements engendrés par ces nouvelles méthodes de consultation et par conséquent les attentes des utilisateurs.

Si l'ensemble de ces points ont été abordés, le principal objectif de cette journée d'information nationale consistait surtout à mieux faire connaître le moteur de recherché TEL, à permettre aux utilisateurs de l'utiliser et à débattre de l'importance de la presse historique numérisée pour les chercheurs. L'auditoire a fait part de son point de vue à travers des questions et des commentaires et a également partagé ses réactions sur les réseaux sociaux en publiant des photos de la journée, des idées sur les questions abordées et des commentaires sur Twitter. Il y a eu plus de 30 tweets mentionnant le hashtag #eurnewsFR comme le montrent les exemples ci-dessous :

Enfin, les réponses au questionnaire (Cf. Annexe IV) confirment que l'événement a été très apprécié. La plupart des participants ont particulièrement été intéressés par les présentations consacrées aux innovations technologiques et par la table ronde consacrée aux usages de la numérisation.

ANNEXE I: Programme

9h00 – 09h30 Accueil et enregistrement des participants

9h30 – 09h40 Quelques mots d'introduction par **Sylviane Tarsot-Gillery** | directrice générale | BnF

09h40 – 11h00 Pourquoi Europeana Newspapers ?

Modérateur : **Elisabeth Freyre** | BnF

- Les collections de journaux numérisés de la BnF : les réussites et les limites des premières numérisations par **Philippe Mezzasalma** | BnF
- Présentation du projet, ses objectifs, ses résultats par **Hans-Jörg Lieder**, Bibliothèque d'Etat de Berlin
- L'apport de la BnF dans le projet par **Ioannis Anagnostopoulos** | BnF
- Questions / réponses

11h00 – 11h30 Pause

11h30 – 13h00 Des innovations techniques au service des utilisateurs.

Modérateur : **Jean-Philippe Moreux** | BnF

- Comment les techniques de reconnaissance de caractères utilisées aident la recherche des utilisateurs ? par **Günter Mühlberger** | Bibliothèque universitaire d'Innsbruck
- Comment les techniques basées sur la reconnaissance des articles peuvent apporter de nouveaux outils aux utilisateurs ? par **Claus Gravenhorst** | CCS
- Le développement des entités nommées en français par **Clemens Neudecker** | Bibliothèque d'État de Berlin et **Alaa Abi-Haidar** | BnF
- Questions / réponses

13h00 – 14h30 Déjeuner libre

14h30 – 16h00 Quels usages, quelles pratiques, quelles attentes pour les usagers ?

Modérateur : **Marc Minon** | Cairn

Table ronde avec :

- **Philippe Tétart** | Université du Maine
- **Sophie Kurkdjian** | Université Paris I
- **Laurent Martin** | Université Paris III Sorbonne-Nouvelle
- **Olivier Hamon** | Syllabs

Les intervenants témoigneront de l'évolution de leurs usages (navigation à l'intérieur du texte, nouveaux champs d'étude et mode de lecture, classification différente des contenus) quant à la presse numérisée et exprimeront leurs attentes vis-à-vis de ces nouveaux modes de consultation.

16h00 – 16h15 Pause

16h15 – 16h30 Vidéo sur le projet

16h30 – 17h
- Conclusions et recommandations par **Pascal Sanz** et **Olivier Piffault** | BnF

17h00 – 17h30
- Clôture de la journée d'étude par **Bruno Racine** | président | BnF

ANNEXE II: Liste des participants

Sur plus de 123 inscrits sur Eventbrite, 88 personnes sont venues assister à la journée d'information (intervenants inclus). 74 personnes ont demandé à recevoir la newsletter du projet.

#	Nom	Prénom	Organisme	Souscription à la newsletter
1	Abi Haidar	Alaa	BnF/Lip6/CNRS	oui
2	Anagnonstopoulos	Ioannis	Bibliothèque nationale de France	oui
3	Angjeli	Anila	Bibliothèque nationale de France	non
4	Ansel	Marion	Bibliothèque nationale de France	oui
5	Baryla	Christiane	Bibliothèque nationale de France	oui
6	Bellamy	Dorothée	Cairn.info	oui
7	Bergeret-Cassagne	Axelle	Bibliothèque nationale de France	oui
8	Bertrand	Claire	Bibliothèque nationale de France	oui
9	Blackburn	Charles		oui
10	Blum	Catherine	Bibliothèque nationale de France	oui
11	Bouchard	Isabelle		oui
12	Bourdon	Françoise	Bibliothèque nationale de France	oui
13	Brault	Rachel	Bibliothèque du Patrimoine de Clermont Communauté	oui
14	Camus	Alice		oui
15	Cantie	Philippe	Bibliothèque nationale de France	oui
16	Coffin	Isabelle		oui
17	Constantin	Michel		oui
18	Coste	Marie-Madeleine	Bibliothèque nationale de France	oui
19	Dechavanne	Sylvie	Archives départementales du Val-d'Oise	oui
20	Delauney	Else	Bibliothèque nationale de France	oui
21	Denou	Nathalie	Service historique de la Defense	oui
22	Desbouchages	Beatrice	Sénat	oui
23	Descharrières	Benonît		oui
24	Desnonyer	Lydie	BSG	oui
25	Doumerc	Gaëlle		oui
26	Dutertre	Julien	Archives départementales de la Gironde	oui
27	Essard-Busail	Brunon	Centre du livre et de la lecture en Poitou-Charentes	oui
28	Floc'hlay	Catherine	Bibliothèque nationale de France	oui
29	Floiras	Jacky	Education nationale	oui
30	Freyre	Elisabeth	Bibliothèque nationale de France	oui
31	Gilles	Arnaud	Bibliothèque nationale de France	oui
32	Gravenhost	Claus	CCS	non
33	Grisoni	Philippe	Sénat	oui

#	Name	Surname	Organisation	Subscription to the project newsletter
34	Hamon	Olivier	Syllabs	oui
35	Haon	Sandrine	Bibliothèque	oui
36	Hebert-Chaminade	Sophie	Cinémathèque Française	oui
37	Hondet	Elsa	Sénat	oui
38	Hopfner	Cindy		oui
39	Josse	Isabelle	Bibliothèque nationale de France	non
40	Jouve	Jean-Jacques	CEA	non
41	Kurkdjian	Sophie	IHTP	oui
42	Lallement	Nicole		oui
43	Lambert	Sylvaine	Ville de Paris Bibliothèque Historique	non
44	Lame-Bergis	Alix	Bibliothèque nationale de France	oui
45	Langlais	Pierre-Carl	CELSA	non
46	Launay	Patricia	BnF	oui
47	Lavergne	Céline	ARKHÊNUM	oui
48	Lecoeur	Laurence	INA	oui
49	Lemaitre	Aurelie	Irisa-université Rennes 2	oui
50	Lerouge	Julien	LITIS (Université de Rouen)	non
51	Lervese	Maud	SHD, Département Bibliothèque	oui
52	Lieder	Hans Jörg	Staatsbibliothek zu Berlin	non
53	Likforman	Laurence	Telecom ParisTech	oui
54	Malé	Séverine	Immanens	non
55	Malec	Angélique	Université Paris Sud	oui
56	Mannaz-Dénarié	Christine		non
57	Martin	Laurent	Université Paris III Sorbonne-Nouvelle	non
58	Mauléon	Agnès	i2S	oui
59	Mezzasalma	Philippe	Bibliothèque nationale de France	oui
60	Minonn	Marc	Cairn.info	oui
61	Moreux	Jean-Philippe	Bibliothèque nationale de France	oui
62	Muehlberger	Guenter	Innsbruck University	oui
63	Neudecker	Clemens	Staatsbibliothek zu Berlin	oui
64	Nonwicki	Frederic		oui
65	Nunes	Eric		oui
66	Ogerau	Anne-Stéphane	BnF-Partenariats	oui
67	Oury	Antoine	ActuaLitté	oui
68	Panayiotou	Hélène	Bibliothèque nationale de France	oui
69	Paquet	Thierry	Université de Rouen	oui
70	Piffault	Olivier	Bibliothèque nationale de France	oui
71	Pluvieux	Isabelle	Centre de recherche du château de Versailles	oui
72	Poirier	Annick	Bibliothèque nationale de France	oui

#	Name	Surname	Organisation	Subscription to the project newsletter
73	Pomer	Nicolas	NPC	oui
74	Rastoul	Hélène	BPI	oui
75	Renard	Joséphine	Bibliothèque nationale de France	oui
76	Rouillard	Maurice		oui
77	Sanz	Marie-Lyse		oui
78	Sanz	Pascal	Bibliothèque nationale de France	oui
79	Steffen	Florian	Bibliothèque Nationale Suisse	oui
80	Tétart	Philippe	Université du Maine	non
81	Tomasi	Anne	B.I.T.	oui
82	Tomasi	Gilbert	B.I.T.	oui
83	Tousni	Slimane	Bibliothèque nationale de France	non
84	Vaisman	Jean-Baptiste	Bibliothèque nationale de France	oui
85	Valérie	Louison-Oudot	Bibliothèque nationale de France	non
86	Valotteau	Hélène	Mediathèque Françoise Sagan, Paris	oui
87	Vayness	Shalev		oui
88	Verdese	Vincent	Archives nationales	oui

ANNEXE III: Couverture médiatique

Avant l'événement, le programme de la journée d'information a été envoyé au Ministère de la Culture et de la Communication, aux directeurs des plus grandes institutions culturelles françaises comme les Archives nationales, aux bibliothèques universitaires, aux personnes en charge des collections de périodiques des bibliothèques parisiennes, aux personnes en charge du numérique dans les grandes maisons d'éditions françaises et dans les journaux de presse quotidienne français.

Cet événement a fait l'objet d'une diffusion via :

- La lettre d'invitation officielle
- La liste de diffusion par mail
- La diffusion du programme aux événements organisés par le département de la coopération nationale de la BnF
- La lettre d'information de la BnF
- Les réseaux sociaux de la BnF
- Le site et les réseaux sociaux Europeana Newspapers
- L'événement a été annoncé au n°1018 de la revue hebdomadaire *Livres Hebdo* et publiée le 14 novembre 2014

COLLOQUE
La presse dans Europeana

Jeudi 27 novembre, la Bibliothèque nationale de France (BNF) accueille une journée d'étude consacrée à la presse historique dans les collections de la bibliothèque numérique Europeana. Depuis 3 ans, la BNF participe à un projet de recherche soutenu par la commission européenne visant à donner accès en ligne à 18 millions de pages via Europeana et The European Library. Cette journée ouverte à tous gratuitement sur inscription, permettra de faire le point sur les collections disponibles et les améliorations techniques à apporter. La deuxième partie du programme montrera l'impact d'un tel projet sur l'historiographie et la recherche scientifique à travers les témoignages de chercheurs et utilisateurs de longue date des collections de presse numérisées. V. H.

Site : http://bnf.fr

Extrait du *Livres Hebdo* n°1018

- L'événement a été annoncé dans un bref article paru dans le magazine *Chroniques* n°71 de septembre 2014

Europeana Newspapers

Journée d'étude
Le projet Europeana Newspapers

Jeu. 27 novembre 2014
9 h 30 – 17 h
Site François-Mitterrand
Grand auditorium

Lancé en février 2012, Europeana Newspapers vise à rendre accessible en ligne les collections historiques de la presse quotidienne européenne : un outil précieux pour les chercheurs, mais aussi pour les passionnés et les simples curieux.

Le projet, coordonné par la Staatsbibliothek zu Berlin, compte dix-sept partenaires, dont neuf bibliothèques nationales, trois bibliothèques universitaires, deux bibliothèques régionales et la Ligue européenne des bibliothèques de recherche (LIBER). Il vise à rendre accessibles au public, *via* Europeana, les articles des collections de presse quotidienne européenne, soit dix-huit millions de pages. Une attention particulière est portée à ceux publiés pendant la Première Guerre mondiale, en synergie avec le projet Europeana Collections 1914-1918. Europeana Newspapers, c'est aussi bien sûr l'occasion d'optimiser les pratiques en matière de numérisation des fascicules de presse (découpage à l'article, extraction des entités nommées, reconnaissance optique de caractères) et d'enrichir les modes de recherche de ces contenus numériques. Le projet arrivera à son terme début 2015 : consultable gratuitement par tout internaute, Europeana Newspapers jouera alors pleinement son rôle d'outil d'information et de démocratisation de la culture. ●

Corine Koch

Extrait du *Chroniques* n°71

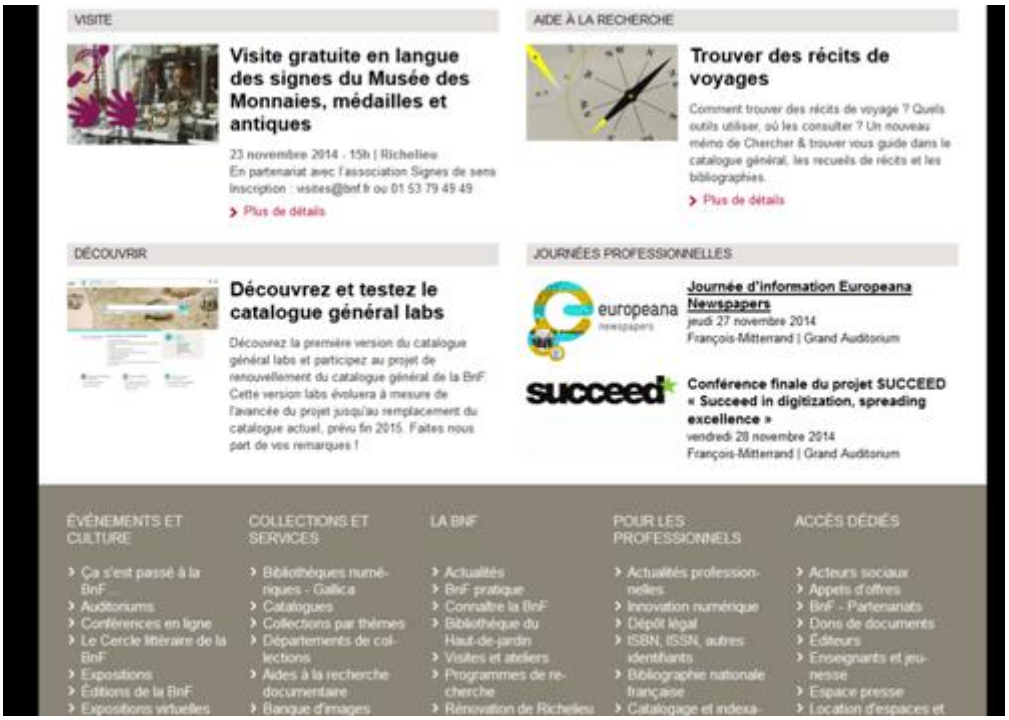
- Un événement Eventbrite a été spécialement créé pour promouvoir la journée et recueillir les inscriptions.

http://www.eventbrite.fr/e/billets-europeana-newspapers-journee-detude-13158504441

- Une page dédiée, présentant également le moteur de recherche TEL, a été créée sur le site BnF.fr

http://www.bnf.fr/fr/professionnels/anx_journees_pro_2014/a.jp_141127_europ_newspapers.html

- La semaine précédant l'événement a vu la parution d'une annonce sur la page d'accueil du site BnF.fr.



Annonce de la journée d'information sur la page d'accueil du site www.bnf.fr

Enfin, chaque participant a reçu un dossier contenant :

- Le programme de la journée et les biographies de chaque intervenant
- La liste des participants
- Une brochure et une carte postale sur le projet (en version française)
- Un dépliant sur l'action internationale de la BnF
- Un questionnaire de satisfaction à remplir

Dossiers remis aux participants

ANNEXE IV: Résultats du questionnaire

Le questionnaire nous a été retourné par une quarantaine de participants. Il témoigne de retours très positifs quant au programme et à l'organisation de cette journée.

1.Par quels moyens avez-vous connu le projet Europeana Newspapers :

- Site web : 11
- Réseaux sociaux : 1
- BnF : 22
- Autres : 6

2.Comment qualifieriez-vous les points abordés lors de cette journée :

- Insatisfaisant: 0
- Moyen: 1
- Bon : 28
- Très bon : 11

3.Pour quelle intervention en particulier êtes-vous venu ?

- Pourquoi Europeana Newspapers : 12
- Des innonvations techniques au service des utilisateurs : 23
- Quels usages, quelles pratiques, quelles attentes pour les usagers : 5

4.Quels thèmes abordés vous ont été particulièrement utiles :

- OCR, OLR et ENR : 12
- Les innonvations techniques : 7
- Les attentes des usagers : 5
- Tous : 12

5.Dans le cas où nous n'aurions pas abordé certains points, lesquels vous auraient intéressés ?

Certains participants auraient aimé avoir la démonstration d'une recherche dans le moteur de recherche, d'autres un aperçu comparatif par rapport à d'autres plateformes de numérisation. Enfin, certains auraient aimé élargir le projet à la numérisation de la presse des autres siècles.

6.Le contenu des tables rondes vous a semblé :

- Insatisfaisant : 0
- Moyen 4
- Bon 26
- Très bon 10

7.Utiliserez-vous montre moteur de recherche dans les prochains mois ?

http://www.theeuropeanlibrary.org/tel4/newspapers

- Oui : 36
- Non, pourquoi ? 4

Utilisation occasionnelle

8.Votre venue (lieu de la conférence, accueil, aménagements, traduction…) :

- Insatisfaisant : 0
- Moyen : 0
- Bon : 22
- Très bon : 14

9.L'organisation de la journée d'information (programme, support, café…) :

- Insatisfaisant : 0
- Moyen : 2
- Bon : 20
- Très bon : 15

10.Dans quel domaine travaillez-vous ?

La plupart des participants travaillaient dans les bibliothèques, les archives, l'informatique, l'enseignement et la recherche

11.Merci de nous préciser si vous avez des commentaires et des suggestions :

Concernant le souhait exprimé de voir mises en ligne les présentations, celles-ci sont d'ores et déjà disponibles sur Slideshare et la BnF diffusera la vidéo de la journée sur son site en 2015.