

**Comité de définition et d'orientation
du portail des thèses (CDO)
Réunion du 28 septembre 2012 (Paris, MESR)
Relevé de décisions**

Présents :

Véronique DE KOK, MISTRD
Sandra JOSEPH, DGESIP – MESR
Martine BENOIT, CPU
Nadia HILAL, CGE
Serge BAUIN, CNRS
Christiane STOCK, INIST/CNRS
Olivier ROUCHON, CINES
Joachim SCHOPFEL, ANRT
Mélanie RIBAS, Intelli'Agence Association Bernard Gregory
Vincent MIGNOTTE, Intelli'Agence Association Bernard Gregory
Grégory MIURA, ADBU
Sophie MAZENS, AURA
Raymond BÉRARD, ABES
Isabelle MAUGER PEREZ, ABES

Absents :

Caroline WIEGANDT, Radio France
Francine PIERRE, THALES GROUP
Sylvain COLLONGE, Institut de formation doctorale Paris 6
Laurent REGNIER, DGESIP – MESR
Françoise PROFIT, DGESIP – MESR

Excusés :

Michel MARIAN, MISTRD

1. Les prochains services proposés par theses.fr

• **theses.fr & l'ABG-Intelli'agence**

Mélanie RIBAS expose des perspectives de collaboration entre l'ABG-Intelli'agence et theses.fr

Theses.fr pourrait proposer dans chaque page de thèse (soutenue ou en préparation) un résumé destiné au monde économique, notamment les recruteurs. Ce résumé faciliterait le recrutement de jeunes docteurs dans les entreprises.

Une discussion suit sur les modalités de rédaction d'un tel résumé. Raymond BÉRARD indique que l'ABES peut facilement mettre en place une nouvelle zone destinée à cette information. Mais est-ce que les doctorants ou les jeunes docteurs la renseigneront ? Est-ce que le contenu de ce résumé sera réellement

une vulgarisation de leur travail doctoral ? Est-ce qu'il sera compris par les entrepreneurs ?

Mélanie RIBAS distingue les deux catégories de public : les doctorants, captifs pendant la durée de leur recherche doctorale, vs les jeunes docteurs pour lesquels il sera plus difficile de mettre en place un accompagnement.

Mélanie RIBAS cite comme mesure d'accompagnement possible la création de tutoriel en ligne, de webinaire et comme méthode un démarrage en douceur auprès de telle ou telle école doctorale.

Le CDO conclut ce point en définissant les domaines de compétences de chacun :

- à charge pour l'ABES de rendre possible la saisie de ce résumé ;
- à charge pour l'ABG de sensibiliser les doctorants, docteurs et personnels des établissements de l'intérêt de saisir cette zone, et de mettre en place les mesures d'accompagnement nécessaires.

Par ailleurs, les données de theses.fr pourraient venir alimenter la plateforme de l'ABG. Aujourd'hui c'est via un système déclaratif qu'un jeune docteur met à disposition son CV dans la CVthèque de l'ABG. Il renseigne lui-même le titre de son doctorat. La question de la fiabilité des informations est donc posée. Récupérer les données depuis theses.fr donnerait force de preuve puisque les données de theses.fr sont validées par les établissements de soutenance.

De manière générale, il faudrait que theses.fr et la plateforme de l'ABG incitent chaque doctorant/docteur utilisateur d'un des deux services à devenir utilisateur du second. Par exemple, si un jeune docteur dépose son CV dans la base de la CVthèque de l'ABG, une alerte pourrait lui suggérer de créer dans son espace personnel dans theses.fr un lien vers le CV déposé dans la CVthèque de l'ABG. Un rebond depuis theses.fr vers la consultation d'offres d'emplois serait également intéressant.

Mélanie RIBAS expose ensuite la possibilité de mettre en place un matching entre les mots clés présents dans theses.fr et ceux présents dans les offres d'emplois de l'ABG. Un tel alignement permettrait de suggérer à des utilisateurs de theses.fr une liste de jeunes docteurs compétents dans ces domaines. Inversement, un recruteur déposant une annonce sur la plateforme de l'ABG pourrait être orienter vers les pages des doctorants/docteurs de theses.fr correspondant à ce profil.

Une discussion s'engage sur ce point qui relève des techniques de data mining. Isabelle MAUGER PEREZ précise que l'exploitation des données de theses.fr, métadonnées mais aussi texte intégral des thèses, a été envisagé dans le cahier des charges initial du projet. Mais l'ABES n'est pas armée pour se lancer seule dans ce chantier. Sa pertinence n'est pas remise en doute mais les forces, les compétences et le temps manquent pour engager à court ou moyen terme ce dossier.

Mélanie RIBAS expose ensuite les contraintes à prendre compte. La plateforme actuelle de l'ABG-Intelli'agence fait l'objet de développements : début 2013 doit être livrée une nouvelle version contenant notamment des web services. Au cours de l'année 2013 sortira la version 2 de la plateforme, basée sur d'autres technologies qui permettront plus facilement l'interconnexion avec theses.fr.

Le CDO décide :

- de mettre en place dès à présent une communication conjointe ABES-ABG pour valoriser theses.fr. L'ABG propose à l'ABES de participer à la réunion des conseillers ABG prévue en janvier 2013.
- d'attendre la stabilisation technique de la plateforme de l'ABG pour amorcer les développements nécessaires aux échanges de données entre theses.fr et l'ABG.

• **theses.fr et le catalogue « Thèses à la carte » de l'ANRT de Lille**

Joachim SCHOPFEL présente le service « Thèses à la carte », service d'impression à la demande qui repose sur un contrat d'auteur non exclusif entre le docteur et l'ANRT de Lille, sans volonté de générer des recettes autres que celles nécessaires à la couverture de frais de fabrication. 7 000 titres sont disponibles.

Le CA de l'ABES du 14 juin 2012 a validé le principe du lien entre les catalogues de l'ABES et les sites publics de vente en ligne.

Le CDO valide le rebond de theses.fr vers Thèses à la carte. Pour les 7 000 thèses concernées, theses.fr proposera un bouton « acheter le fac-similé ». L'ANRT doit fournir à l'ABES la liste des numéros nationaux de thèses des titres concernés.

2. État d'avancement du projet depuis la réunion du 30 septembre 2011

Isabelle MAUGER PEREZ présente les réalisations.

Le cap des 12 000 thèses soutenues déposées dans STAR a été dépassé au mois d'août 2012. La base a doublé en un an.

La prise en main de STEP par les 80 établissements anciens utilisateurs du Fichier central des thèses s'est faite sans difficultés majeures. Les thèses en préparation en sciences dures et sciences appliquées commencent à être signalées.

Les statistiques de consultation de theses.fr (fournies par Google analytics) sont les suivantes :

	Visiteurs uniques	Visites	Pages vues	Pages par visite	taux de rebond (en %)	nouvelles visites (en %)
juil-11	12 711	15 815	86 810	5,49	39,70	77,10
août-11	11 868	15 288	75 383	4,93	44,53	72,27
sept-11	15 961	21 043	116 255	5,52	44,77	70,34
oct-11	61 815	80 467	402 331	5,00	53,97	73,55
nov-11	73 306	93 069	399 129	4,29	57,03	70,69

déc-11	66 891	83 918	349 976	4,17	57,13	70,07
janv-12	91 055	112 971	427 499	3,78	58,81	71,51
févr-12	95 259	119 530	440 975	3,69	60,08	69,38
mars-12	99 652	126 295	467 914	3,7	59,96	67,79
avr-12	93 250	117 746	426 059	3,62	60,35	67,36
mai-12	97 161	123 563	427 365	3,46	62,13	66,59
juin-12	94 124	118 702	409 503	3,45	62,74	66,75
juil-12	79 648	100 498	364 660	3,63	61,39	66,53
août-12	70 448	87 358	306 319	3,51	63,77	69,09

Le pic constaté à l'automne 2011 correspond à la mise en ligne des thèses en préparation.

Les nouvelles fonctionnalités concernant theses.fr développées au cours de l'année 2011-2012 sont :

- la recherche des personnes (ouverture le 17 janvier 2012)
- l'assistance déportée¹ (ouverture le 2 avril 2012). Il s'agit d'un service de mise en relation directe entre l'utilisateur de theses.fr ayant découvert une erreur dans les données et l'établissement de soutenance producteur de la donnée.
- la version anglaise de l'interface (ouverture le 7 septembre 2012).

Lors de la réunion du 25 mars 2011, le CDO avait choisi de retenir comme troisième langue de theses.fr celle du pays avec lequel la France signe le plus de co-tutelles. Les chiffres fournis hors réunion ont orienté le choix vers l'italien. En sa séance du 28 septembre 2012, le CDO s'interroge sur la pertinence de l'italien et décide que cette question sera mise à l'ordre du jour de la prochaine réunion.

Parallèlement aux réalisations, l'équipe projet a travaillé sur deux dossiers de longue haleine :

Le premier concerne l'interconnexion entre STEP et STAR ainsi que l'alimentation des deux applications par imports de données, notamment en provenance d'APOGEE ou d'ADUM.

Le second concerne le chargement dans theses.fr des notices Sudoc relatives aux thèses, en priorité les notices des thèses soutenues à partir de 1985.

Pour passer de n notices du Sudoc décrivant la même œuvre intellectuelle thèse à une seule page dans theses.fr, un algorithme de regroupement a été développé. Il permet la constitution d'une « famille » de notices, la notice mère décrivant la thèse faisant foi au titre du dépôt national des thèses et les notices enfants décrivant les éventuelles éditions sur d'autres supports ou versions

¹ cf. « Le cercle vertueux de theses.fr : mise en relation de l'utilisateur et du producteur de données », [Punktokomo](#), 10 avril 2012

remaniées. Les difficultés rencontrées tiennent à la qualité des notices présentes dans le Sudoc :

- numéro national de thèse absent ou mal constitué, ce qui empêche l'attribution d'une url pérenne dans theses.fr
- libellés de diplôme hétéroclites, ce qui complique l'identification des thèses de doctorat
- variété de formes possibles pour décrire la discipline de la thèse (plus de 27 000), ce qui ralentit l'attribution d'un code domaine basé sur la classification Dewey et nécessaire au fonctionnement de la facette domaine dans theses.fr
- liens erronés entre notices
- pauvreté initiale dans le Sudoc (pas de directeur de thèse, pas de résumés, pas de mots-clés), ce qui aboutira inéluctablement à une pauvreté dans theses.fr

Le travail de chargement des données du Sudoc dans theses.fr est ainsi l'occasion de plusieurs chantiers de corrections de masse touchant parfois à des centaines de milliers de notices.

Le CDO valide la démarche pragmatique de l'ABES : le chargement des données Sudoc dans theses.fr sera fait au fil de l'eau, au fur et à mesure des corrections de masse. Il est décidé de ne pas rechercher l'exhaustivité immédiate.

Ces deux dossiers aboutiront avant la fin de l'année civile.

3. Discussions autour du service d'achat en ligne

Le CDO est sollicité pour apporter au Conseil d'administration de l'ABES des éléments lui permettant de prendre une décision.

Isabelle MAUGER PEREZ résume le contenu de la note sur le service d'achat en ligne. Il s'agit de savoir à quel(s) opérateur(s) pourrait être confié le bouton « acheter » permettant à un utilisateur de theses.fr de commander une thèse remaniée chez un libraire en ligne.

Il n'est pas proposé d'exclusivité vers un seul fournisseur. En effet, s'il n'y a pas d'exclusivité, il n'est pas nécessaire de lancer un appel à concurrence ; il faudra signer avec le(s) partenaire(s) intéressé(s) de simples contrats d'affiliation, permettant à l'ABES de percevoir une rémunération pour chaque visite, chaque vente, ou chaque contact. En revanche, si l'ABES souhaitait travailler uniquement avec un site, et qu'il y ait par conséquent exclusivité, il faudrait faire un appel à concurrence.

Le troisième scénario qui propose de recourir à un comparateur de prix du livre a fait l'objet d'une étude plus poussée. JustBooks et Chasse aux livres sont présentés tour à tour. Chasse aux Livres présente plusieurs avantages :

- liens vers des libraires en ligne bien implantés sur le marché français (Furet du Nord, Decitre..). A contrario, JustBooks fait la part belle aux fournisseurs internationaux.

- interface de recherche intégrant des outils 2.0 comme le fait theses.fr
- accord du propriétaire de Chasse aux livres pour développer des web services répondant aux besoins de theses.fr

Une discussion suit. Véronique DE KOK indique que la MISTRD est attentive à la cohérence de ce projet avec la politique de l'État en faveur de la librairie de proximité. Elle mentionne Place des Libraires, site de la société d'outils d'informations pour les produits culturels Tite-Live. Place des Libraires permet de chercher un livre parmi plus de 150 librairies en France.

Isabelle MAUGER PEREZ indique que Place des Libraires ne permet pas aujourd'hui de faire une requête par ISBN. Or ce critère est déterminant dans le choix d'un prestataire.

Joachim SCHOPFEL s'interroge sur l'identité du propriétaire de Chasse aux livres et sur la pérennité de ce site web ouvert récemment. Grégory MIURA intervient pour rappeler que les comparateurs représentent un marché économique du web. Sophie MAZENS se demande dans quelle mesure il est possible de confier l'exclusivité du service « acheter » de theses.fr à un seul opérateur, qu'il s'agisse d'un libraire en ligne ou d'un comparateur.

Le CDO émet l'avis suivant :

- favorable au principe d'établir un lien de theses.fr vers des sites commerciaux
- nécessité d'obtenir des informations sur les propriétaires du site « Chasse aux Livres »
- possibilité de s'ouvrir à d'autres comparateurs ou libraires en ligne (notamment Places des libraires) sous réserve :
 - * que leur catalogue de thèses soit suffisamment fourni
 - * que leur ergonomie soit satisfaisante

[Hors réunion, Chasse aux livres est une société (immatriculation SIRET 52785178600014 en date du 1^{er} octobre 2010) appartenant à Thibault LE PORT, basée dans la région nantaise. La marque a été déposée à l'INPI le 28 février 2011 sous le numéro 3810631]

4. Discussion autour de la consultation du cabinet Bensoussan « Diffusion des thèses sur Internet » (avril 2012)

L'ABES a constaté qu'un site de partage et de publication de documents, www.youscribe.com, a moissonné le serveur OAI-PMH de STAR pour récupérer les métadonnées des thèses diffusables sur Internet. Le compte thesee de Youscribe propose gratuitement la lecture en ligne, le téléchargement et l'impression du texte intégral de plus de 3500 thèses sans qu'aucune autorisation n'ait a priori été demandée ni aux auteurs de thèse, ni aux établissements de soutenance.

Raymond BÉRARD rappelle les questions posées par l'ABES, les réponses apportées par le cabinet Bensoussan et commente la consultation « Diffusion de thèses sur Internet ».

Une discussion suit.

Le CDO estime que les mesures de sécurité proposées par le cabinet Bensoussan s'appuyant sur le référentiel de sécurité RGS (dont le verrouillage de l'entrepôt OAI-PMH et la traçabilité des internautes consultant le texte intégral via un questionnaire demandant nom, profession etc...) iraient à l'encontre de la volonté de favoriser la diffusion des thèses. Elles ne sont donc pas retenues.

Raymond BÉRARD propose que l'ABES mette en œuvre le plan d'action suivant :

1) Inciter les établissements à uniformiser les contrats conclus avec les auteurs et leur fournir une trame.

2) Ajouter sur les bordereaux STAR et le formulaire d'enregistrement des thèses papier une mention rappelant l'existence au sein de l'établissement de soutenance d'un contrat de diffusion soumis à l'auteur de la thèse

3) Conditions générales d'utilisation (CGU, liste des usages autorisés et des portails ou services de référencement autorisés) :

* inciter les établissements à insérer des CGU sur leurs sites

* insérer des CGU sur les sites et plateformes gérées par l'ABES

Le CDO valide ce plan d'action.

5. Prochaine réunion du Comité de définition et d'opération

Raymond BÉRARD rappelle au CDO que le projet Portail des thèses s'arrête fin 2012. L'opportunité de maintenir les réunions du CDO pour suivre le fonctionnement des applications en production est discutée.

Le CDO décide de maintenir une réunion par an. La prochaine réunion est fixée au **vendredi 27 septembre 2013 de 10h à 13h** dans les locaux du ministère.

6. Questions diverses

Grégory MIURA rappelle la situation dans laquelle se trouvent les grandes écoles pour le signalement de leurs thèses en préparation. L'environnement de production des données est l'application Thésa, site géré par la CGE et dont l'INIST CNRS est l'opérateur technique. Il est prévu que les données produites dans Thésa soient versées dans theses.fr. Le flux de circulation de données serait simplifié si l'environnement de production de données était STEP, les sites Thésa et theses.fr étant les environnements de valorisation. Cela permettrait à la CGE de recentrer son activité sur l'incitation à signaler les thèses en préparation et sur la valorisation des données produites : Thésa pourrait proposer un contenu éditorialisé autour des données produites dans STEP.

Nadia HILAL indique que cette proposition mérite d'être étudiée et précise qu'il n'y a pas d'opposition de principe à ce que les grandes écoles produisent leurs données dans STEP. L'essentiel est de mettre en place le circuit le plus simple possible pour garantir un flux de données permanent et de qualité.

Faute de temps, le CDO propose de rediscuter de cette proposition lors de sa prochaine réunion.

Raymond BÉRARD remercie les participants et lève la séance à 13h10.