

Réunion du groupe de travail ST2I "systèmes d'information" du 20/10/2010

Lieu : LSIIT (Illkirch)

Horaires : début 9h00 - fin 11h45

Présents :

InESS : Daniel Mathiot, Caroline Eckert, Nicolas Colin, Sébastien Schmitt

IMFS : Michael Essa

LSIIT : Fabrice Heitz, Michel de Mathelin, Christine Stutz, Aline Gérard, Stéphane Prunière, Jean-Marc Muller

LSP : Joel Fontaine

IPB : Jean-Paul Armspach

Ordre du jour :

Point d'avancement sur les deux projets :

- gestion des publications ;
- Système d'information du laboratoire.

1. Application de gestion des publications

Lors de la précédente réunion (12/07/10) une liste de tâche à effectuer a été définie :

- Mener une enquête d'utilisation/d'attentes concernant Basilic auprès des utilisateurs ;
- Mettre à disposition une version de test de Basilic pour tous les laboratoires du projet ;
- Produire une nouvelle version de Basilic en se basant sur les résultats de l'enquête ;
- Modifier Basilic afin de pouvoir mettre à jour automatiquement HAL lors de l'adjonction d'une publication dans Basilic ;
- Migrer les bases de publications existantes dans les laboratoires du projet vers la base de données sur laquelle s'appuie Basilic.

Bilan :

- L'enquête a été effectuée ;
- L'outil de test a été mis en place ;
- La nouvelle version de Basilic a été mise en production ;
- La fonctionnalité de mise à jour de HAL lors de l'ajout d'une publication a été développée. Reste à compléter celle-ci afin de pouvoir mettre à jour une publication existante que l'on désire modifier ;
- La phase de migration de bases n'est pas encore d'actualité.

Réflexions et remarques formulées lors de la réunion suite à la démonstration de la dernière version de Basilic par Aline :

- Pb de copyright :

- ➔ La mise en ligne de documents liés aux publications peut poser des problèmes de copyright . Certaines publications ne peuvent être mises à disposition en accès libre.
- ➔ Solution proposées : émission d'un avertissement pour l'utilisateur lors de l'ajout d'une publication ? Désignation d'un superviseur (responsable d'équipe) qui serait averti automatiquement par mail lors de l'ajout d'un document et qui aurait un rôle de modérateur ? Ne pas permettre l'ajout des documents mais plutôt faire figurer le DOI (Digital Object Identifier) de la publication qui pointe sur celle-ci.

- Import vers HAL :

- ➔ Actuellement l'adjonction d'une publication à HAL à partir de Basilic est soumis au choix de l'utilisateur (case à cocher). Ceci s'explique par la nécessité d'un contrôle par l'utilisateur avant ajout à HAL ;
- ➔ Ceci présente le risque de ne pas avoir toutes les publications dans HAL ;
- ➔ Nécessité de développer un contrôle de complétude de la référence d'une publication lors de son ajout à Basilic et si le contrôle est positif d'ajouter automatiquement celle-ci à HAL ;
- ➔ Se pose alors le problème de définir ce qu'est une référence complète. Ceci est actuellement le sujet d'un débat au sein de la communauté des publiants. Une définition exacte sera nécessaire pour mettre en place ce contrôle.

- "Impact factor" et "Acceptance rate"

- ➔ Séparation des deux critères "Impact Factor" et "Acceptance rate" ;
- ➔ Le champ Impact Factor nécessite une remise à jour annuelle et peut être déterminé automatiquement sur Internet. Pas de réelle nécessité de conserver cette information dans la base de données ;
- ➔ Le taux d'acceptation d'une publication doit être renseigné par le publiant qui lui seul dispose de cette information ;
- ➔ Le taux d'acceptation ne concerne que les conférences ;
- ➔ Le publiant a actuellement la possibilité de choisir s'il rend publiques ou non les informations liées à ces deux critères ("Impact Factor" et "Acceptance Rate") grâce à l'utilisation d'une case à cocher. Cette possibilité ne semble pas légitime pour tous. Elle devra donner lieu à une discussion ultérieure ;

- Année de publication

- ➔ Lors de la définition de l'année de publication, proposer le choix "To appear" ou "Accepted for publication".
- ➔ Ceci permettrait de mettre en place un contrôle régulier de l'état d'avancement des publications et ainsi effectuer les mises à jour de ce paramètre lors de la parution de la publication ;

- Détection des doublons

- ➔ Nécessité de mettre en place un système de détection des doublons dans la base de publication ;
- ➔ Par souci de performances il n'est pas envisageable d'effectuer un tel test directement lors de l'ajout d'une nouvelle publication ;
- ➔ Possibilité d'effectuer un traitement asynchrone de la base et d'envoyer un mail au publiant afin de l'avertir de la présence d'un doublon ;
- ➔ Ceci pose deux problèmes :
 - Sur quels critères se baser pour identifier les doublons ;
 - Comment gérer l'ajout de la publication à HAL (il faudrait attendre le résultat du contrôle de doublon avant d'ajouter la publication à HAL) ;
- ➔ Un nettoyage des doublons sera nécessaire lors de l'importation des bases des différents laboratoires dans Basilic ;

- Nomenclature fixant le type des publications renseignées dans Basilic

- ➔ Les responsables d'équipes sont souvent amenés à devoir corriger le type des publications renseignées par les membres de leur équipe ;
- ➔ Nécessité de fixer une liste officielle de type de publication; Besoin d'une nomenclature ;
- ➔ Besoin d'une version longue (texte) de la nomenclature et une version compacte (acronyme).
- ➔ Ce choix permettra ainsi de personnaliser le formulaire de saisie au type de publication choisi et de maintenir une classification cohérente des publications ;
- ➔ Cette liste doit permettre une classification "gros grain" des publications ;
- ➔ Une première liste a été proposée lors de la réunion :
 - Article,
 - Conference,
 - Thesis,
 - Book,
 - Report,
 - Brevet,
 - Misc
- ➔ Le choix initial ne pouvant être modifié ultérieurement est extrêmement important ;
- ➔ Une réunion est prévue le 28/10/2010 de 10h à 12h afin d'en discuter ;
- ➔ Prévoir pour celle-ci la nomenclature AERES, la nomenclature SIGAS, la liste des champs Bibtex ainsi que ceux de HAL.

- Import des bases des laboratoires dans Basilic

- ➔ Le portage des bases des laboratoires vers Basilic risque de poser quelques problèmes liés au contenu : pas les mêmes informations renseignées, pas la même classification (nomenclature évoquée ci-dessus) ;
- ➔ De nombreux champs risquent d'être à renseigner par défaut ;
- ➔ Pour rappel, figurent ci-dessous les formats des bases actuelles des labos :
 - IMFS : Bibtex (import automatisable) ;
 - INESS : XML (import automatisable) ;
 - LSP : fichier word (nécessite une saisie manuelle) ;
 - IPB : fichier word (nécessite une saisie manuelle).

- Points mineurs :

- ➔ Lors du renseignement de l'URL de la publication, indiquer automatiquement au début de l'URL http:// ;
- ➔ Vérifier le format lors de la saisie du numéro de téléphone du publiant. Forcer le format, proposer une pré-saisie modifiable ;
- ➔ Le champ Team dans la version de test de Basilic propose une liste un peu confuse. Les choix possibles pour la version d'exploitation devront être une des 13 équipes du laboratoire ou sinon "externe" ;
- ➔ Rajouter le nom du site (localisation géographique) du publiant ;
- ➔ Supprimer le champ "Key" qui semble redondant avec le champ "Keywords" déjà présent ;

2. Système d'information du laboratoire

Lors de la précédente réunion (12/07/10) une liste de tâche à effectuer a été définie :

- Collecte des procédures d'accueil des nouveaux arrivants dans les différents labos ;
- Production d'une fiche type et d'un outil de saisie en ligne ;
- Mise en place d'une base de données commune sur un serveur fiable ;
- Alimentation automatique des sites Web des différents laboratoires à partir de cette base.

Bilan :

- Christine Stutz a collecté les différentes procédures d'accueil (en vue de produire un vade-mecum commun)
- Une étude des procédures d'enregistrement des nouveaux arrivants au secrétariat du LSIIT a été effectuée ;
- Divers entretiens téléphoniques (Xavier Duthilleul pour Labintel, Jocelyn Kerleau pour Harpgest, Béatrice Meier-Muller pour Grall) ont permis d'obtenir des renseignements concernant les principaux logiciels de gestion des personnels et de la recherche du CNRS et de l'UdS ;
- Une étude du Kit AERES de la vague B a été effectuée afin de définir précisément les données à collecter dans la base ; production suite à cette étude d'une série de diagrammes à affiner avec Fabrice Heitz qui a proposé son aide pour ce faire ;
- Deux serveurs ont été déployés. Le premier pour héberger la base, le deuxième pour permettre la gestion du projet ainsi que de projets annexes menés par le service informatique.

Conclusion :

La réunion n'a pas permis de traiter correctement le deuxième point de l'ordre du jour. Le système d'information du laboratoire n'est pour l'instant pas prioritaire comparé à l'application de gestion des publications. Les conclusions suivantes ont néanmoins été formulées :

- Les logiciels institutionnels tels Harpgest et Labintel ne semblent actuellement pas prévus pour s'interfacer avec un logiciel laboratoire permettant une saisie unique des informations. Harpgest dispose de Webservices à l'état de test. Jocelyn Kerleau nous a fait parvenir une documentation concernant ceux-ci ainsi qu'un accès à une base de test.
- Nous allons dans un premier temps produire une base de donnée à même de nous permettre d'automatiser la production du prochain rapport d'activité ;
- Une partie du rapport d'activité est à produire sous forme de fichiers excel comprenant de nombreuses macros. Ceci risque de compromettre sa production de façon

automatisée à partir d'une extraction de la base de données. Des compétences de développement sur plate-forme windows (VB, excel,..) semblent nécessaires, or personne dans le service informatique ne dispose de telles compétences. Jean-Marc va tenter d'avancer sur ce sujet.

- Dans le pire des cas la production du rapport d'activité sera semi-automatisée (saisie des informations dans la base, extraction des informations dans des fichiers textes et copie de ces informations dans le fichier fourni par l'AERES) ;
- Le projet n'étant pas prioritaire comparé à Basilic, Aline ne s'y consacrera qu'après avoir finalisé ses travaux sur Basilic. Il faudrait néanmoins avancer sur ce projet pour la production du rapport d'activité.
- L'étude plus approfondie de ce projet avec les membres du comité de pilotage été reportée à une date ultérieure non définie. D'ici là Jean-Marc et Fabrice Heitz tenteront d'avancer (définition des données à consigner, production de la base, outil de saisie) au mieux.