

Proposition d'architecture pour le portail math V2

Version V1 9/9/2013

Historique des versions

- *La première version de ce document V0.2 avait été rédigée par Jean-Luc Archimbaud avec l'aide de Nicolas Houillon et Franck Longin et soumis à la PLM Team.*
- *Suite à des propositions de corrections de la PLM team, une version V0.3 a suivi, rédigée par Jean-Luc qui a inclus ces demandes et ajouté quelques précisions.*
- *Cette version V1, rédigée par Jean-Luc fait suite à une dernière relecture avec quelques correctifs de la PLM team et donc approuvée par la PLM Team.*

Pour les concepteurs du portail, les différentes versions sont stockées sur le serveur portail math dans la page privée accessible via le menu Concepteurs portail / Réunions-Docs internes.

Ce document est une proposition d'architecture pour le portail math V2, validé par la PLM Team, pour soumission ensuite au comité exécutif du portail math.

Dans ce document, nous appelons :

- 'portail math' (<http://maquette-portail-math.mathdoc.fr/> actuellement, très prochainement <http://www.portail-math.fr>) l'objectif du projet portail math (http://www.mathdoc.fr/projet_portail_math)
- 'PLM 2.0' ce que l'équipe Mathrice met en place <https://plm.math.cnrs.fr/portail/>

1. Le portail math V1

L'organisation du projet portail math, les phases, les choix techniques... sont décrits là : http://www.mathdoc.fr/projet_portail_math

Cette version V1 du portail math devait ouvrir fin septembre 2013. Actuellement, seul reste à faire la mise en place d'un nouveau look . Ceci est en cours, confié à une société externe. Mais les choix prennent plus de temps que prévu d'où le retard qui s'annonce : l'ouverture se fera plutôt en octobre. Cette version V1 est basée sur Drupal avec identification-authentification local Drupal et contrôle d'accès sur le contenu... géré par Drupal.

Elle contient des descriptions de ressources documentaires (ensembles de documents, points d'accès, licences, bases biblio, documents utiles), de services informatiques (c'est là que se retrouve les services de la PLM) et de sites utiles.

Ces descriptions pointent les unes vers les autres et évidemment vers les sites où les informations/services se trouvent.

Dans les descriptions d'ensemble de documents on a déjà les listes des revues par exemples, on pourra aller jusqu'à la liste des articles (dans des Vn) avec des imports automatiques.

Ces descriptifs sont actuellement remplis manuellement progressivement par des documentalistes, informaticiens, chercheurs, avec des débuts d'essai d'intégration d'informations plus automatiquement (documents utiles par exemple).

Il y a des menus, des mots-clés et une recherche plein texte (qu'on essaie de rendre performante) pour l'utilisateur.

Le nombre et le contenu des descriptifs augmente de jour en jour et pourra continuer lorsque le portail math V1 sera ouvert. Donc il se peut que lors de l'ouverture on ait peu d'informations sur le portail math mais que ça augmente exponentiellement ensuite. Voici une photo des descriptifs sur le portail math au 9 sept :

- Ensembles de documents : 47 descriptifs : 18 publiés, 29 en cours de création.
- Bases de données bibliographiques: 7 descriptifs : 1 publié, 6 en cours de création.
- Points d'accès: 42 descriptifs : 15 publiés, 27 en cours de création.
- Licences/contrats: 20 descriptifs : 11 publiés, 9 en cours de création.
- Documents utiles : 23 descriptifs : 20 publiés, 3 en cours de création.
- Services informatiques : 6 descriptifs : 5 publiés, 1 en cours de création.
- Sites utiles : 21 descriptifs : 13 publiés, 8 en cours de création.

Pour les ressources documentaires, un descriptif d'ensemble de documents est lié à un ou plusieurs descriptifs de licences (s'il n'est pas en accès libre), à un ou plusieurs descriptifs de points d'accès (qui sont les différents sites-portails pour y accéder) et symétriquement. Ceci permet à l'utilisateur de savoir comment accéder à son information, d'abord manuelle en V1 puis progressivement automatiquement en V2 et suivantes. Les points d'accès indiquent par exemple la méthode d'accès, qui a le droit d'y accéder...

Les services informatiques sont les services assurés par Mathrice mais aussi ceux assurés par d'autres (RENATER, labos de recherches, universités... voire privés).

Pour le descriptif des services de la PLM, une option est de fondre l'actuel plm.math.cnrs.fr (le user guide de la PLM) dans le portail. Mais cela étant couteux en temps, si l'on choisit cette option, il est raisonnable d'attendre la V3 pour que le choix de l'architecture cible soit définitif.

L'architecture fonctionnelle du portail math V1 est en annexe 1.

2. Objectif du portail math V2

Le comité de pilotage a demandé en janvier que cette version V2 soit opérationnelle en janvier 2014. Si la V1 a du retard, celui-ci sera à prendre en compte. Elle devra intégrer : l'identification et l'authentification des utilisateurs (permettant par exemple, pour les ressources documentaires, d'indiquer à l'utilisateur ce à quoi il a accès (au niveau ensemble de documents) et le meilleur chemin (point d'accès) pour y accéder), et l'intégration de l'accès aux ressources documentaires actuellement accessibles via la PLM. Si on peut faire plus, ce sera bienvenu.

3. Proposition d'architecture pour le portail math V2

On garde Drupal pour la partie documentaire (au sens large) du portail : accès aux ressources documentaires et aux sites utiles.

Les utilisateurs arrivent sur la page d'accueil du portail. Ils ont alors la possibilité de s'authentifier et/ou de se diriger vers les composantes Ressources documentaires, Sites utiles et Services informatiques. Les 3 composantes du portail s'appuient sur un processus unique d'authentification / identification qui leur fournit les attributs de l'utilisateur issus de l'authentification. Ce processus s'appuie sur un module non Drupal fourni par Mathrice et sur un module Drupal (interface entre Drupal et le module Mathrice) qu'il faut développer.

L'option précédente d'installer les modules Drupal standards d'interfaçage avec LDAP et Shibboleth pour cette fonction d'authentification/identification est abandonnée au profit de l'option ci-dessus.

Donc Drupal, comme les autres composantes du portail (services PLM), s'appuie comme les autres sur un module d'interface développé par Mathrice qui lui permet de récupérer les attributs de l'utilisateur issus de la fédération d'identité Renater et de l'annuaire de la communauté mathématique française (<http://annuaire.emath.fr>), dont les informations nécessaires pour déterminer les droits d'accès (en particulier ses entités d'appartenance).

Pour l'accès aux descriptifs (en particulier des ressources documentaires) on assure un contrôle d'accès (sachant que beaucoup d'informations seront publiques). Ce contrôle est géré par Drupal.

Ce qui suit est au conditionnel car à valider.

Sur Drupal, on utiliserait les comptes, les rôles et les groupes (organic groups en jargon Drupal). Chaque compte (utilisateur) pourrait avoir un ou plusieurs rôles, appartenir à un ou plusieurs groupes. Un groupe pourrait être un ensemble de personnes, tous les membres d'un laboratoire, tous les membres d'une université, les membres d'un ensemble de labos ou d'universités,...

L'accès aux descriptifs (et à tout ce qui peut être visualisé et modifié dans Drupal) se ferait via les droits et les mécanismes de contrôle d'accès Drupal. Cette gestion des droits Drupal sur le portail math permettrait à l'utilisateur, suivant les rôles qu'il a et les groupes auxquels il appartient :

- d'afficher ou non des blocs, des menus, des parties de menus,
- d'afficher ou non des types de descriptifs,
- d'afficher ou non chaque descriptif séparément,
- d'afficher ou non chaque champ de descriptif,
- de pouvoir modifier ou non un type de descriptif,
- de pouvoir modifier ou non chaque descriptif séparément,
- d'avoir des droits différents sur les descriptifs dans le workflow de publication et d'autres workflow,
- d'avoir des listes de contenus (descriptifs...) formatées automatiquement suivant ses droits (les éléments auxquels il n'a pas accès ne sont pas affichés),
- d'avoir des résultats de recherche différents suivant ce à quoi il a accès,
- ...

Cette liste est classique pour un CMS mais elle permet de se rendre compte que le contrôle d'accès dans un portail n'est pas simplement un contrôle d'accès à des informations – des documents mais beaucoup d'autres choses imbriquées

On utilise déjà la plupart de ces possibilités dans le portail math V1, pour les menus, les workflow, l'accès aux descriptifs, les recherches... mais actuellement uniquement basé sur les rôles, on n'a pas mis en place de groupes car inutile en V1. Ceci veut dire qu'il faut vérifier et tester la faisabilité de cet ensemble de possibilités de contrôle d'accès avec les groupes.

Un compte Drupal serait créé dynamiquement pour chaque nouvel utilisateur et sur la base des informations fournies par le processus d'authentification/identification décrit avant l'utilisateur serait automatiquement affecté aux bons rôles et aux bons groupes.

Pour l'authentification, lorsqu'un utilisateur arrive sur le portail math, il peut se connecter en simple visiteur sans authentification. Quand il le souhaite, il active l'authentification (lien ou bouton login) ce qui le renvoie vers un Where Are You From (WAYF) (à priori le WAYF de Mathrice hébergé sur la PLM qui propose les établissements de la Fédération Education Recherche, plus le fournisseur d'identité spécifique Mathrice). Il s'authentifie ensuite sur l'établissement choisi ce qui transmet au final au portail un certain nombre d'informations :

- issues des éléments renvoyés par l'établissement (attributs Shibbolhet) (informations brutes) : login, [login@domaine](#), Nom Prenom, mail
- mais également issues de traitement effectués sur ces éléments et des éléments puisés dans

l'annuaire emath et Mathrice

- canonisation de l'adresse email – mécanisme de convergence d'identité pour les personnes disposant d'un compte PLM)
- information d'appartenance à des groupes en rapport avec la consultation des ressources documentaires

Sur la base de ces informations Drupal détermine les informations auxquelles l'utilisateur a accès.

4. Accès aux ressources documentaires accessibles actuellement via la PLM en V2 du portail math.

Un objectif fixé par le comité de pilotage du portail math pour la V2 est d'intégrer les ressources documentaires actuellement accessibles via la PLM (ensembles de documents : [Elsevier Backfiles Mathematics \(archives avant 1994\)](#), [EMS Journals](#), [NUMIR](#), [Springer Journals](#), [Springer Journal Archives](#), [Springer LNM Archives](#) et bases bibliographiques [Mathscinet](#) et [Zentralblatt](#)) dans le portail math. Se pose la question : comment ?

Deux orientations principales, sans avoir réfléchi :

- Sur le portail math, dans la description des ensembles de documents concernés on affiche comme point d'accès 'portail PLM', comme pour les services informatiques de la PLM et le portail math ne fait rien. Tout est redirigé vers le portail PLM. Cette solution n'est pas vraiment une intégration, simplement un renvoi vers le portail PLM. Mais elle ne demande aucun développement côté portail math, avantage certain,
- Sur le portail math, on fait quelque chose de plus intégré, à définir précisément. L'identification est faite, les droits contrôlés, donc on peut permettre à l'utilisateur d'accéder ou non aux revues concernées avec des mécanismes de proxy, tunnel... A noter que le portail math étant un portail institutionnel INSMI, des négociations avec les éditeurs sont possibles pour le faire accepter comme un « contrôleur d'accès de confiance » (l'INIST est considéré ainsi par certains éditeurs).

Le choix n'est pas uniquement technique mais aussi stratégique (INSMI, RNBM, Mathdoc...) et demande de tenir compte de l'évolution de la documentation : plate-forme ISTEEX, évolutions de l'INIST, des archives ouvertes...

Mais ce n'est pas la priorité de réflexion pour l'instant.

5. Versions V3 et suivantes du portail math

Pour la V3 du portail math un cycle de réflexion de fond sur l'architecture pourra être mené, si le comité de pilotage du portail math l'autorise.

Drupal a été choisi très rapidement car on n'avait pas le temps de mener une réflexion sur le fond : l'INSMI voulait un portail en 6 mois. La question du choix du CMS pourra être posée ultérieurement, si cela s'avère nécessaire.

D'autres ingénieurs à Mathdoc ou ailleurs imaginent d'autres systèmes d'informations complets : bases de données, briques logicielles, mécanismes... Les idées ne manquent pas.

Annexe : architecture portail math V1

