

Le 15 novembre 2021

GROUPE WEBHOSTING

1. INTRODUCTION

Une offre Web complémentaire de celle de la DSI et plus en adéquation avec les besoins exprimés par le département MIA devenu MathNum, a été proposée et mise en place en 2008, avec le Cati 1ère génération CIAM.

Le groupe de travail WebHosting propose et assure le suivi du CMS (content management system) Drupal, dans la création de sites Web au sein d'un hébergement sécurisé.

Une dizaine d'année plus tard, nous sommes toujours actifs et gérons trente à quarante sites par an. Les temps forts du groupe WebHosting depuis 2012 sont :

- La migration du 1er serveur physique hébergé à Jouy en Josas jusqu'en 2014, vers un 2nd serveur physique mutualisé entre la ForgeMIA et WebHosting hébergé au Data Center de Toulouse.
- Changement de version des sites développés sous Drupal en version 6 vers la version 7 en 2015.
- La migration de l'ensemble des sites sous un serveur virtuel (VM DSI packagés) hébergé par le Data Center de Toulouse 2019. Le stockage des données a été placé sur un espace disque lui aussi loué à la DSI. Cet espace payé par le département est partagé avec l'application Gitlab (ForgeMIA).
- Changement de version des sites développés sous Drupal en version 7 vers la version 8 en 2020.
- La création d'un nouveau serveur virtuel (VM packagé) pour l'hébergement des sites en Drupal 9 et hébergé au Data Center de Toulouse 2021.
- Changement de version des sites développés sous Drupal en version 8 vers la version 9 (en cours).

2. ETAT DES LIEUX

A. Présentation de l'organisation actuelle

Rôle : Le groupe de travail WebHosting, propose une offre en mode SaaS (Software as a Service) d'hébergement de site web avec le CMS Drupal aux unités du Département MathNum.

Le groupe assure le suivi du CMS Drupal, dans le cadre d'un hébergement sécurisé. L'offre peut être adaptée aux compétences techniques des utilisateurs.

Le point fort de cette offre réside dans la possibilité de disposer d'un outil et de possibilités de personnalisation différentes de l'offre institutionnelle.

Mission du groupe WebHosting :

Le groupe est composé de quatre agents du département MathNum.

Mission principale :

Valérie Vidal :

- Département MathNum (0,5 ETP)
- Animatrice du groupe WebHosting
- Webmaster du Portail MathInfo (site institutionnel du département)
- Administratrice CMS Drupal et Webmaster
- Communication auprès des utilisateurs, aide à la conception de site
- Développement
- Accompagner les utilisateurs (conseils, support etc.)

Mission transversale :

Mikael Grialou :

- Administrateur LAMP (Linux, Apache, MySQL, PHP)
- Administrateur CMS Drupal
- Développements spécifiques éventuels
- Accompagnement des utilisateurs (conseils, support etc.)

Christian Poirier (0,05 ETP) :

- Support à l'administration système (Rundeck)
- Conseil et veille technique

Maurice Baudry (0,05 ETP) :

- Support à l'administration système
- Conseil et veille technique

B. Description de l'hébergement

En 2020, 45 sites étaient hébergés. En 2021, [43 sites](#) sont hébergés par la plateforme sans compter les sites statiques. Le nombre de site est quasi constant sur les cinq dernières années.

Différentes catégories de sites sont hébergées :

- Les réseaux métiers (ex RESSTE ...)
- Les évènements (ex colloque StatMathAppli)
- Des Unités internet/intranet (ex MaIAGE)
- Les CATI & PEPI (ex IMOTEP, IBIS)
- Les projets & les groupes de travail (ex MSE, Stat4Plant)
- Des plateformes (ex Migale).
- Des portails (ex Workflow4Metabolomics, MathInfo).
- Des écoles chercheurs (ex EPICASA)
- Les clusters (ex BioSP-Cluster)

C. Taches réalisées au niveau du groupe

Mise à jour mensuelle du core (noyau), des modules et des thèmes de Drupal avec :

- Planification de la tâche à accomplir.
- Envoi d'un mail d'informations aux gestionnaires de sites.
- Réalisation de la tâche (script Rundeck).
- Vérification et validation de la tâche (Selenium).
- Envoi d'un mail d'informations aux gestionnaires de sites.

Création d'un site :

- Réception d'une demande de site par mail.
- Échange avec la personne des besoins exprimés :
 - Choix du nom du site avec ou sans DNS.
 - Établir le ticket auprès de la DSI (le demandeur ou nous) si DNS.
- Validation avec l'utilisateur des besoins :
 - Par exemple le module Webform pour les colloques.
 - Mail applicatif.
- Création, installation et paramétrage du site.
 - Création d'une base de données MySQL.
 - Création et configuration d'une instance Drupal.
 - Configuration du fichier setting.php de Drupal et Apache.

- Installation des modules que nous conseillons et d'un choix de thèmes stables.
- Rédaction et affichage des contenus obligatoires : Mentions légales, CGU et Gestion des cookies.
- Création d'un compte administrateur pour l'utilisateur.
- Envoi d'un mail le prévenant et lui rappelant les points importants :
 - Le mail du support.
 - Son login et mot de passe.
 - Ce qui a été fait selon les besoins exprimés.
- Demande de certificat de sécurité à la DSI.

Aide et accompagnement pour les sites déjà créés :

- Installation de module spécifique à la demande des webmasters de sites.
- Aide au paramétrage de ces modules
- Accompagnement à la migration de sites dans les différentes versions du CMS
- Transfert de site sans évolution dans un mode dit « site statique » (*page HTML et CSS*).

Rôle de Webmaster pour une dizaine de sites

- Design du site
- Installation modules pour l'ergonomie du site
- Mise en ligne du contenu

Veille et évolution des outils et services :

- Mise en place d'un serveur d'administration pour centraliser l'exploitation du service (Rundeck).
- Ouverture d'un espace de travail collaboratif pour la gestion des documentations (Nextcloud)
- Mise en place de certificats de sécurité générique pour faciliter l'exploitation.

D. Infrastructure

Espace NetAPP : Les données de webh représentent 16Go sur les 3 To loué à la DSI (mutualisé avec la ForgeMIA)

Machines virtuelles :

- Serveur Drupal 7 et 8 - vm-pkg-32 - 4 VCPU - 8go de RAM
- Serveur Dupal 9 - vm-pkg-61 - 4 VCPU - 8go de RAM - Drupal 9.
- Serveur d'administration - admin-webh - 1 VCPU - 2 Go de RAM - 20 Go disque
- Serveur Pré-Production MATHINFO - vm-pkg-27 - 1 VCPU - 2go de RAM - 5,4 Go disque
- Serveur Production MATHINFO - vm-pkg-28 - 1 VCPU - 2go de RAM - 5,4 Go disque

E. Taux d'occupation des sites de l'infrastructure des VMs

vm-pkg-32 occupation drupalWH_C8 - 13G	vm-pkg-61 occupation drupalWH_C9 - 2,7G occupation static-website - 872M
<p style="text-align: center;"><u>Tailles Sites actifs</u></p> 388M migale 632M modstatsap 288M phis 410M reseau-payote 262M smpgd 275M biosp-cluster 275M biblio-maiage 335M statmathappli 1,2G maiage 370M biobaye 366M ingenum 1,2G mexico 1,1G pepi-ibis 325M prodmia 604M sysmics 264M webh 547M imotep 936M reseau-resste 249M omer 304M peakforest 253M reseauA2 249M ssb8 430M mistea 267M mse	<p style="text-align: center;"><u>Tailles Sites actifs</u></p> 235M stat4plant 563M workflow4metabolomics 1,1G biosp 306M ofe2021 260M pse
	<p style="text-align: center;"><u>Tailles Sites statiques</u></p> 556K cbme.ssbgroup.fr 4,6M ciam 696K nemo.ssbgroup.fr 481M reseau-m3d.fr 2,3M projet-peerless 48M workshop-ssiab9 74M swg2014 61M cascisdi.inrae.fr 203M ag2019.mia.inrae.fr 255M epicasa

280M sesame 276M usemetabo 269M resiste 331M reseau-cisstats	
vm-pkg-28 Prod MathInfo	vm-pkg-27 Préprod MathInfo
288M mathinfo	281M mathinfo

F. Stack technique mise en œuvre

CMS Drupal : est un système de gestion de contenu libre et open-source, publié sous la licence publique générale et écrit en PHP. Drupal peut être utilisé à quatre niveaux différents :

- **Tel quel** : une fois celui-ci installé et paramétré, il est utilisable pour créer du contenu structuré et les annotations des utilisateurs peuvent s'enregistrer sur le site. Les menus du site ont alors un aspect standard.
- **Personnalisation simple** : il est ensuite possible de personnaliser l'emplacement d'affichage ou l'affichage lui-même, de composants visuels standards ainsi que le thème d'affichage du site.
- **Extension par ajouts externes** : ajout, paramétrage et personnalisation de modules optionnels n'appartenant pas au noyau.
- **Extension par développement interne** : écriture de nouveaux modules, qu'il est souvent bien (mais nullement obligatoire) de présenter ensuite à la communauté afin que celle-ci puisse participer à leur évolution.

Matomo : logiciel libre et open source de mesure de statistique web, conçu pour être une alternative libre à Google Analytics. Matomo fonctionne sur les serveurs web PHP/MySQL et s'appuie sur l'analyse des logs du serveur web pour fournir des indicateurs de fréquentation. Les statistiques détaillées sont aussi disponibles et sont réparties en quatre catégories : Visiteurs, Actions, Référénts et Objectifs.

- La partie **Visiteurs** comporte un onglet Récapitulatif qui renseigne entre autres sur le nombre de visites, de visiteurs uniques, d'actions effectuées et sur la durée moyenne des visites.

- La section Actions informe du nombre de pages vues, du temps moyen sur les pages, des pages d'entrées et de sortie du site, du nombre de clics sur les liens sortants et des téléchargements effectués.
- La catégorie Référents apporte des informations sur la provenance des visites : entrées directes, liens depuis d'autres sites internet ou moteurs de recherche avec des statistiques sur les mots clefs utilisés.
- La partie Objectifs donnera des statistiques de fréquentation du site de la plateforme Migale.

Les différentes statistiques chiffrées sont consultables sous forme de tableaux, de graphiques ou de nuages de tags. Le logiciel diffère de Google Analytics par la prise en compte de l'intégralité des données d'audience des sites web alors que l'outil Google fonctionne par échantillon.

Selenium : est un framework pour tester des applications Web. Il fournit un outil de lecture pour création de test fonctionnel. Il fournit également un langage spécifique au domaine de test pour écrire des tests, ceux-ci pourront ensuite s'exécuter sur les navigateurs. Selenium est composé de plusieurs composants, chacun jouant un rôle spécifique dans l'aide au développement de l'automatisation des tests d'applications Web.

Rundeck : logiciel libre permettant l'automatisation d'administration de serveurs via la création de jobs ou tâches. L'interface de Rundeck permet de créer des jobs ou tâches applicatives, qui peuvent s'exécuter de façon périodique ou manuelle, sur un ou plusieurs serveurs.

- Elle permet d'exécuter une simple commande sur plusieurs serveurs.
- L'exécution des jobs ou des commandes peut s'effectuer en parallèle sur plusieurs serveurs afin de gagner en temps d'exécution.
- Un onglet recensant toute l'activité de Rundeck permet à l'administrateur de pouvoir consulter tout l'historique des tâches exécutées par l'outil.

KeePass : gestionnaire de mot de passe utilisant deux algorithmes de chiffrement, permettant de sauvegarder un ensemble de mots de passe dans une base de données chiffrée sous la forme d'un fichier. Ce fichier de base de données s'ouvre avec un mot de passe maître chiffré.

Uptime Robot : service de surveillance en ligne de la disponibilité des sites, notre choix s'est porté sur le forfait gratuit qui propose la :

- Surveillance du HTTP(S) sans les erreurs SSL, du temps de réponse, des pings, des ports, des mots clés et une rétention de trois mois des journaux.

Quire : logiciel collaboratif d'espace de travail individuel ou partagé, en ligne pour gérer tous nos projets, basé sur le diagramme de Gantt.

NextCloud : portail d'accès, de partage et de coédition de document, il s'agit d'une offre de la DSI, auprès de qui nous avons ouvert un compte.

G. Identifications des risques encourus et moyens de contrôles mis en place

Le risque majeur est la sécurité des serveurs, des logiciels et des sites. Pour assurer cette sécurité, le collectif a mis en place une veille technologique afin d'être réactif en fonction des alertes nous pouvons intervenir sous 24 heures :

- Serveurs virtuels packagés (VM) :
 - Il s'agit d'une offre d'environnement logiciel de virtualisation hébergé au data center de Toulouse. La sécurité et les problèmes relatifs à ses différentes couches sont gérées et administrées par la DSI. Nous interagissons avec eux via des tickets Ariane.
- Une veille technologique du CMS Drupal :
 - Réception de mails quotidiennement, sur une adresse mail dédiée (groupe-webhmia@inrae.fr) via les instances Drupal, des mises à jour de sécurité du core, des modules et des thèmes qui doivent être faits.
 - Compte individuel sur le site de drupal.org pour recevoir la lettre d'information hebdomadaire de Drupal, le bulletin mensuel de l'Association Drupal et des mails ad hoc occasionnels avec des mises à jour importantes.
- Une veille technologique de Matomo :
 - Compte commun au groupe WebHosting donnant accès aux données des sites utilisant ces statistiques et au blog

de Matomo où les informations concernant les dernières mises à jour disponibles.

- Une veille technologique de Rundeck :
 - Compte individuel sur l'interface en ligne donnant accès aux scripts de gestion des projets et aux news de la communauté Rundeck.

3. BILAN

A. Analyse des forces et les faiblesses de notre organisation, ainsi que les opportunités et les menaces de l'environnement

Notre **force** :

- La cohésion du groupe qui est stable depuis 2016.
- La mise en place d'une Visio conférence hebdomadaire favorisant les échanges sur l'avancement d'un projet, collaborer pour répondre rapidement à une demande de service, etc.
- L'élaboration de tutoriels disponibles à tous les membres du groupe (partage de connaissance)
- L'amélioration des outils d'administration pour faciliter la gestion.
- Le respect du RGPD et l'affichage sur tous les sites des Mentions Légales, CGU et Gestion des cookies, ainsi que le suivi des nouveautés/modifications sur la protection des données personnelles à INRAE.

Notre principale **difficulté** :

- Réside dans la taille du groupe et en particulier sur les quatre personnes, il n'y a que deux personnes disponibles qui consacrent du temps sur les demandes de sites et les interventions.

Les **opportunités** :

- La fusion, INRA IRSTEA a impliqué la reconstruction du département MathNum (MIA) avec l'intégration de cinq unités parmi lesquels, l'unité ITAP a montré un intérêt pour l'offre (voir le site [OFE2021](#)).
- Chacune de onze Unités du département représente un potentiel important de création de site pour des réseaux,

événements, Unités internet/intranet, CATI & PEPI, projets & les groupes de travail, plateformes, portails, écoles chercheurs, etc.

Les **menaces** :

- Manque de visibilité sur l'offre DSI d'un hébergement dans l'état de l'art comme l'offre du Data Center de Toulouse (offre VM et stockage DSI).

En conclusion, le soutien sans discontinuité du département MIA puis MathNum reste pour nous la meilleure des motivations.

4. PERSPECTIVES

A. Audit sécurité

- Faire réaliser un audit de sécurité.

B. Architecture système

- Réfléchir aux évolutions de l'architecture matériel pour suivre les besoins et collaboration avec les acteurs DSI.

C. Le groupe WebHosting

- Élargir le groupe dans la consolidation et la perspective de d'évolution du nombre de site hébergé.

5. LA FEUILLE DE ROUTE

A. Focalisation du groupe sur deux actions étroitement liées

- Migration vers la version 9 du CMS Drupal : la communauté Drupal a annoncé courant février 2019 que la version de Drupal 8 ne serait plus supportée du fait de la sortie de la version de Drupal 9 le 2 novembre 2021, plus d'information sur le site officiel de [Drupal](https://www.drupal.org/).
- Prise en compte de l'évolution du département MIA vers le département MathNum dans le nouveau contexte INRAE
 - Le changement du nom du département.
 - La réservation du nom DNS auprès de la DSI.
 - L'amélioration de l'affichage des sites du département.
 - Le changement de version de Drupal et la fin des versions antérieures.

- La migration est nécessaire pour des raisons de sécurité.

B. Nouvelle organisation

- Tous nos sites auront un domaine en « *.mathnum.inrae.fr » à la fin de chaque url, cela donnera par exemple : au lieu de « **informatique-mia.inrae.fr/NomSite** » ceci « **NomSite.mathnum.inrae.fr.** ».
- Chaque site ayant son propre DNS, facilitera la mise à jour des certificats de sécurité des sites, les migrations, etc.
- Nous gérons la totalité des sites dont le domaine est en *.mathnum.inrae.fr devra passer par nous, même si nous ne gérons pas le site par la suite.
- Les utilisateurs pourront continuer d'obtenir, un nom de domaine propre (*cf Ariane*) s'il le souhaite.

C. Avancé du travail 2021

- Provisionnement et configuration d'une nouvelle VM pour accueillir Drupal 9 (*achevé*)
- Récolte des souhaits de chaque gestionnaire de site (*achevé*)
- Évolution des sites en accord avec leur Webmestre vers les états suivants : (*achevé*)
 - Rendu statique, deux sites
 - Archivé, un site
 - Supprimé, six sites
 - Mise en place d'une protection apache, un site
 - From scratch, deux sites
 - Migration des sites en Drupal 8 vers Drupal 9
 - Vérification via un module de la comptabilité des modules installés sur chaque site avant migration
 - Test de migration des sites
 - Migrations des sites après validation
 - Redirection dans Apache des sites sous « **informatique-mia.inrae.fr/NomSite** » en « **NomSite.mathnum.inrae.fr** »

Sites concernés par la redirection et le renommage		
Sites actifs		Sites statiques
<ul style="list-style-type: none"> • BioBayes • Biosp-Cluster • MSE • OMeR • PHIS • ProdMIA • réseauA2 	<ul style="list-style-type: none"> • CiSStats • ModStatSAP • RESSTE • RESISTE • SESAME • WebHosting 	<ul style="list-style-type: none"> • EpiCasa • SWG2014 • SSIAB

Le DNS « informatique-mia.inrae.fr » ne sera pas supprimé par la DSI afin que la redirection fonction toujours, comme c'est déjà le cas pour le DNS des premiers sites du cati CIAM.