

GROUPE WEBHOSTING

1. INTRODUCTION

Une offre Web complémentaire de celle de la DSI et plus en adéquation avec les besoins exprimés par le département 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.

Les temps forts du groupe WebHosting pour l'année 2023 est l'arrivée dans le groupe de Emmanuelle Vietti de l'unité MIA Paris Saclay à Saclay et le départ du groupe de Maurice Baudry de l'unité LAMME à Evry.

2. ETAT DES LIEUX

Présentation de l'organisation actuelle

Le groupe est composé de trois agents de différentes unités du département MathNum.

Mission principale :

Valérie Vidal, pour le Département MathNum (0,5 ETP) est l'animatrice du groupe WebHosting et occupe les fonctions de Webmaster du Portail MathInfo (site institutionnel du département), d'administratrice CMS Drupal et Webmaster, de la communication auprès des utilisateurs, aide, à la conception de site, du développement et d'accompagner les utilisateurs (conseils, support etc.)

Mission transversale :

Mikael Grialou, pour le Département Mathnum (0,2 ETP) occupe les fonctions d'administrateur LAMP (Linux, Apache, MySQL, PHP), administrateur CMS Drupal, développements spécifiques éventuels et l'accompagnement des utilisateurs (conseils, support etc.)

Sylvain Lanoe, pour le Département Mathnum (0,2 ETP) occupe les fonctions d'administrateur LAMP (Linux, Apache, MySQL, PHP), administrateur CMS Drupal, l'accompagnement des utilisateurs (conseils, support etc.)

Emmanuelle Vietti, pour le Département Mathnum (0,2 ETP) occupe les fonctions Webmaster, l'accompagnement des utilisateurs (conseils, support etc.)

Description de l'hébergement

En 2022, 47 sites étaient hébergés. En 2023, 49 sites sont hébergés par la plateforme sans compter les sites statiques. Le nombre de site est quasi constant sur les six dernières années.

Différentes catégories de sites sont hébergées

- Les réseaux métiers (ex RESSTE ...)
- Les évènements (ex colloque StatMathAppli)
- Des Unités internet/intranet (ex MaIAGE)
- Des outils de gestion (ex ColDem MaIAGE)
- Les CATI & PEPI (ex IMOTEP, IBIS)
- Les projets ANR & 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)

Tâches de fond 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.
- L'envoi d'un mail d'informations aux gestionnaires de sites.
- Réalisation de la tâche (script Rundeck).
- Vérification et validation de la tâche.
- L'envoi d'un mail d'informations aux gestionnaires de sites.

Création d'un site

- Réception d'une demande de site par mail via un formulaire ou par voie de connaissance.
- 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.
- L'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) dans un but d'archivage avec visibilité.*

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 de certificats de sécurité générique pour faciliter l'exploitation.*
 - *Type : gen-SSL-wildcard pour les sites en *.mathnum.inrae.fr*
 - *Type : gen-SSL-AlterName pour les sites avec un DNS dédié*

Infrastructure

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

Machines virtuelles :

- Serveur - vm-pkg-32 - 4 VCPU - 8go de RAM – permet de tester .
- 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

Taux d'occupation des sites de l'infrastructure des VMs			
vm-pkg-61			
49 sites sous Drupal 9 (taille – nom du site)			
247M : 2neurones 829M : applibugs 269M : biblio-maiage 383M :biobaye 289M : bioinfomics 1,2G : biosp 235M : biosp-cluster 576M: ingenum 418M : coldem-maiage 381M : coldem-mia-ps 579M : imotep 1,5G : maiage	238 M : metasimfood 1,2G : mexico 367 M : mia-ps 426M : migale 479M : mistea 660M : modstatsap 269M : mse 491M : ofe2021 258M : omer 314M : peakforest 1,2G : pepi-ibis 243M : pherosensor	302M : phis 259M : pole_sci 296M : portail-mathinfo 341M : prodmia 240M : pse 260M : reseauA2 523M : reseau-cisstats 538M : reseau-payote 1,2G : reseau-resste 247M : reseau-2neurones 282M : resiste 346M : rh-maiage	295M : sesame 265M : smpgd 289M : stat4plant 387M : statmathappli 650M : sysmics 370M : trombi-maiage 305M : usemetabo 272M : vitae 326M : webh 586M : workflow4metabolomics
10 sites statiques en HTML (taille – nom du site)			
203M : ag2019.mia.inrae.fr 289M : AG-MathNum2022 61M : cascidi.inrae.fr 556K : cbme.ssbgroup.fr	4,6M : ciam 290 M : dynstoch 255M : epicasa	696K : nemo.ssbgroup.fr 481M : reseau-m3d.fr 260M : ssb8	74M : swg2014 2,3M : projet-peerless 48M : workshop-ssiab9

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.

Matomo est un 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.

Rundeck est un 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.

KeePass est un 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 gestionnaire de mot de passe s'ouvre avec un mot de passe maître chiffré.

Uptime Robot est un 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.

NextCloud est un 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.

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. PERSPECTIVES

- **Architecture système** : Réfléchir aux évolutions de l'architecture du matériel pour suivre les besoins et collaboration avec les acteurs de la DSI.
- **Le groupe WebHosting** : Élargir le groupe dans la consolidation et la perspective de d'évolution du nombre de site hébergé.

4. LA FEUILLE DE ROUTE

Travaux réalisés en 2021

Nous nous étions donnés comme objectifs de réaliser ces deux actions étroitement liées ce qui a été fait.

La migration vers la version 9 du CMS Drupal avant le 2 novembre 2021, date de fin du support de la version 8 et 7 pour cela :

- Le provisionnement et la configuration d'une nouvelle VM a été nécessaire, pour accueillir Drupal 9.
- Une récolte des souhaits pour chaque site a été faite, auprès des gestionnaires (les choix étaient de rendre le site statique, l'archiver, le supprimer, le migrer vers la nouvelle version de Drupal ou from scratch).
- Une action a ensuite été effectuée, selon les vœux exprimés et le travail exécuté a été validé par les gestionnaires concernés.

L'évolution du département MIA vers MathNum dans le nouveau contexte de INRAE, a nécessité **le changement du nom de domaine** de `informatique-mia.inrae.fr/NomSite` pour `NomSite/mathnum.inrae.fr` puis, une règle de redirection apache a été mise en place pour les sites concernés.

Travaux commencés en 2022

Nous avons été contactés courant août, par Karine Daubas (Responsable de l'équipe Access Management), dans le cadre du projet IAM, un travail est en cours sur les accès LDAP à l'annuaire INRA et de nombreux accès à la `vm-pkg-61` via le LDAP INRA ont été constatés.

C'est la raison qui nous pousse à réviser notre mode de connexion au LDAP INRA, car il est voué à disparaître en 2023 et les `cnx` (connexions) directes pour l'authentification Web seront proscrites sur le nouveau LDAP.

Il est donc nécessaire d'étudier la possibilité de migrer vers la solution nationale `authentification.inrae.fr` avec une intégration dans la fédération Renater.

Concernant les connexions via le CAS ce protocole étant compatible avec la nouvelle solution, il suffit d'établir un ticket Ariane afin qu'on déclare le site sur la nouvelle solution, ainsi la migration pourra se faire à la date souhaitée.

Nous allons mettre en place le SAML (Security Assertion Markup Language) pour cela des up grade et mises à jour seront nécessaires :

- Up grade de PHP 7 vers PHP 8,
- up grade de composer,
- tests de mise en œuvre.

Travaux réalisés ou en cours de réalisation en 2023

Nous nous étions donnés comme **objectifs**, de réaliser la demande de la DSI dans le cadre du projet IAM. Nous n'avons pas pu mettre en place une authentification SAML, comme nous le souhaitions, en remplacement d'une connexion via le LDAP INRAE, le test avec des modules différents gérant le SAML testés sur Drupal n'étant pas triviaux ou compatibles.

Sachant que les sites utilisant une connexion CAS pouvaient migrer facilement vers cette nouvelle authentification SSO « `authentification.inrae.fr` », une demande pour chaque site a été faite auprès de la DSI.

De plus pour les sites ayant un internet et un intranet, un attribut complémentaire a été mis en place, il est basé sur l'utilisation de l'attribut « `supannEntiteAffectationPrincipale` » qui ne donne accès qu'aux agents de l'unité concernée via son numéro codique.

Nous nous étions donnés comme **perspective**, de réaliser un audit de sécurité, ce qui a été fait.

Nous l'avons réalisé en testant chacun des sites via ce site <https://www.ssllabs.com/ssltest>. Ce qui nous a permis de corriger les failles que nous avons, c'est-à-dire désactivé l'utilisation de TLS1.1 et TLS1.0

dans le fichier ssl.conf de Apache, et répondre ainsi aux consignes et contrôles aléatoires opérés par la CNIL.

Migration vers la version 10 du CMS Drupal : la communauté Drupal a annoncé le 17 juillet 2020 que la version de Drupal 9 ne sera plus supportée du fait de la sortie de la version de Drupal 10 le 1^e novembre 2023.

En **prévision** de cette migration courant 2023 :

- Récolte des souhaits de chaque gestionnaire de site puis évolution vers l'état souhaité a été réalisée en début d'année. (action : Valérie)
- Vérification via un module de la comptabilité des modules installés sur chaque site avant migration et en cours, près de 70% des sites sont prêts à être migrés (action : Valérie)
- Test de migration des sites et migration des sites après validation en cours de test. (action : Mikael et Valérie)

Nous hébergeons de nombreux sites tous différents, et il m'a semblé important de m'assurer que nous respectons les règles RGPD.

J'ai pris contact dans le courant de l'année 2023 avec Nathalie Gandon, qui a confié la demande Clémence Pascal (*Adjointe Déléguée Informatique et Libertés Juriste au droit numérique, informatique et libertés - Pôle Droit de la Recherche - **Direction des affaires juridiques***)

J'ai déjà eu de nombreux échanges avec Clémence, aujourd'hui les sites respectent les règles du RGPD.

Prévision pour l'année 2024

- Finaliser le changement de version de l'ensemble des sites.
- Réfléchir à la version du thème bootstrap 3 que nous utilisons largement pour passer à la version bootstrap 5.

5. FORMATION

Nous réfléchissons à ce style de formation autour des outils que nous utilisons quotidiennement, afin d'améliorer notre prestation, auprès des utilisateurs comme :

- **Drupal** administrateur et développeur
- Initiation à **Matomo** et **Symfony**

Nous ne formulons aucune demande particulière auprès du département pour l'année à venir.

6. BILAN

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.

Les opportunités : chacune des onze Unités du département représente un potentiel important de création de site pour des réseaux, évènements, Unités internet/intranet, CATI & PEPI, projets & les groupes de travail, plateformes, portails, écoles-chercheurs, etc.

La menace : 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 faille du département MathNum reste pour nous la meilleure des motivations.